

University of Warsaw,
Division of Biophysics, Institute of Experimental Physics, Faculty of Physics,
The Centre of New Technologies

Application of homology modeling to study the function of proteins and their interactions with ligands

Presented by
Krzysztof Kuchta

Supervisors:

Dr hab. Krzysztof Ginalski

Prof. Edward Darzynkiewicz

Warsaw, 2019

Introduction

Currently, the number of protein sequences (<https://www.ncbi.nlm.nih.gov/>) is 736 times larger than the number of protein 3D structures deposited in the Protein Data Bank (<https://www.rcsb.org/>) (1). Thus, fast and accurate computational structural determination methods seem to be crucial tools to fill the gap between the number of sequences available and protein 3D structures solved experimentally (1). Homology modeling is considered to be the most accurate of the computational structure prediction methods (1). It has many applications in: the drug discovery process (1-3), studying the effect of mutations (4), identifying active and binding sites on protein (5), searching for ligands of a given binding site (6), modeling substrate specificity (7), predicting antigenic epitopes (8), protein-protein docking simulations (9), molecular replacement in X-ray structure refinement (10), rationalizing known experimental observations (11) and planning new computational and biochemical experiments with the provided models.

In my studies I used homology modeling to investigate the function of proteins and their interactions with ligands. I focused on two classes of proteins: eukaryotic translation initiation factors 4E (eIF4Es) and non-canonical poly(A) polymerases. eIF4E, combined with other initiation factors, mediates the mRNA recruitment on ribosomes to start the translation via the cap-dependent mechanism (12-14). Non-canonical poly(A) polymerases, including poly(A) and poly(U) polymerases (PAPs/PUPs), and C/U-adding enzymes, modify 3'-ends of different transcripts affecting their functionality and stability (15,16). Here, I examined the relationship between 3D protein structures obtained with homology modeling and the experimentally showed ability of *Arabidopsis thaliana* and *Drosophila melanogaster* eIF4E isoforms to bind diverse cap analogs (13,14). I also compared obtained 3D model of *Thielavia terrestris* CutA (TtCutA) to selected structures of non-canonical nucleotidyltransferases (NTases) to investigate protein substrate specificity and design rational mutagenesis converting CutA, acting toward CTP, into PAP or PUP, and analogously Cid1 PUP into C/U-adding enzyme (16). Finally, I generated the reliable 3D model for one of the FAM46 family members (FAM46C) and confidently assigned uncharacterized FAM46 proteins to an active non-canonical poly(A) polymerases. Obtained structural model for human FAM46C shed some new light on the molecular role of mutations found in cancer patients in the FAM46 genes (15).

“Five eIF4E isoforms from *Arabidopsis thaliana* are characterized by distinct features of cap analogs binding” (13)

We built 3D homology models of *Arabidopsis* eIF4E isoforms (AtIF4E, AtIF4E1b, AtIF4E1c, AtIF(iso)4E and AtnCBP) to investigate differences in their structures and ability to bind the cap considering the sequence diversity and the variety of contacts between protein and ligand. All analyzed proteins showed the conserved eIF4E-like fold and the cap-binding pocket characteristic for eIF4E class I and class II proteins. Biochemical data revealed that binding affinities of plant eIF4Es for the cap can be arranged in the rank of diminishing affinity: AtIF4E $\approx$ AtIF4E1c > AtIF4E1b > AtIF(iso)4E $\approx$ AtnCBP. We observed the highest cap binding affinities for AtIF4E and for the non-canonical eIF4E isoform AtIF4E1c. Surprisingly, in spite of 95% amino acid sequence identity between AtIF4E1b and AtIF4E1c almost all cap analogs, except m⁷GMP, have association constants for AtIF4E1b ca. 2-fold lower than for AtIF4E1c. 3D models showed that the majority of residues that differ between these two proteins are located within or near the cap-binding region. The main possible reason for the difference in their binding affinity may be the replacement of large hydrophobic Phe147 and Thr145 in AtIF4E1c with Met147 and Ser145 in AtIF4E1b, which could result in weaker stabilization of the conserved arginine (Arg188 in AtIF4E1b/AtIF4E1c, equivalent to Lys162 in mammalian eIF4E) that binds cap's β - and γ -phosphate groups. Additionally, Ala184 in AtIF4E1b replaces Thr184 in AtIF4E1c that may form hydrogen bonding (i.e. through water molecules) with the γ -phosphate. Unexpectedly, the second canonical translation initiation factor, AtIF(iso)4E protein, binds all cap analogs from 4-fold to 10-fold weaker than AtIF4E. Probably, the poorer cap-binding efficiency of AtIF(iso)4E is mainly caused by the shortening of the long loop (that contains a short α -helix) between β -strands β 7 and β 8. This loop is longer by six amino acids in AtIF4E, AtIF4E1b and AtIF4E1c, and forms the edge of the cap binding cavity, taking part in tight binding of the sugar and β - and γ -phosphate moieties of cap analogs. Finally, the most prominent differences that may be responsible for the poor binding of cap by AtnCBP seems to be the lack of the conserved arginine (Arg183 in AtIF4E, replaced by Ile166 in AtnCBP), that may interact with the cap's phosphate moiety. The lack of arginine in that position can be partially compensated in AtnCBP by Arg125 and Lys202, which are replaced by uncharged amino acids in AtIF4E, AtIF4E1b and AtIF4E1c. However, AtIF4E, AtIF4E1b and AtIF4E1c possess also other positively charged residues that may take part in the binding of cap's phosphate groups and which are not conserved in AtnCBP. In

summary, the use of differentially chemically modified cap analogs in binding studies and a detailed analysis of the obtained homology models shed a new light on the molecular characteristic of varying cap-binding abilities of *Arabidopsis* eIF4E isoforms.

“Diverse cap-binding properties of *Drosophila* eIF4E isoforms” (14)

We build homology models of seven *Drosophila melanogaster* eIF4E structures (*Dm* eIF4E-1, *Dm* eIF4E-2, *Dm* eIF4E-3, *Dm* eIF4E-4, *Dm* eIF4E-5, *Dm* eIF4E-7 and *Dm* 4EHP) with various ligands and analyzed the sequence and structure differences that may influence interactions with the cap and the eIF4G peptide in relation to the results obtained from experimental binding studies. Biochemical data revealed that *Dm* eIF4E-1 binds cap analogues with significantly weaker affinity than its human counterpart (human eIF4E). The main difference, which might affect lower *Dm* eIF4E-1 cap binding affinity, seemed to be an additional α -helix present in human eIF4E and other vertebrates, but absent in *Drosophila*, yeast and plant eIF4Es. The extra α -helix lifts the position of the neighboring $\beta 5\beta 6$ loop towards the cap-binding pocket. As a consequence, in human eIF4E Lys159 located at the tip of $\beta 5\beta 6$ loop interacts tightly with the cap phosphate moiety, in contrast to its counterpart Lys201 in *Dm* eIF4E-1. Secondly, *Dm* eIF4E-3 showed the highest cap binding affinity among class I *Dm* eIF4Es. According to our model, His234 in *Dm* eIF4E-3, located in $\beta 7\beta 8$ loop, is likely to cause the observed high affinity of *Dm* eIF4E-3 for dinucleotide cap analogues. This aromatic amino acid probably takes part in the stacking interactions with the second base of the cap and is replaced by Ser207, Gln249, Gln238, Gln219 and Leu419 in human eIF4E, *Dm* eIF4E-1, *Dm* eIF4E-2, *Dm* eIF4E-4 and *Dm* eIF4E-7, respectively. Prediction of stacking interactions between His234 and the second base of the cap was also confirmed by weaker binding to dinucleotide cap analogues containing pyrimidines instead of purines. *Dm* eIF4E-5 has also an aromatic residue, Phe222, at the position corresponding to His234 in *Dm* eIF4E-3. In consequence, the binding affinity for complexes of *Dm* eIF4E-5 with cap analogues was ca. 3-fold higher than those observed for *Dm* eIF4E-1. Thirdly, a significant increase in cap binding affinity associated with the addition of an extra δ -phosphate group was observed for *Dm* eIF4E-5 and *Dm* eIF4E-7. Based on our modeling, no significant changes in *Dm* eIF4E-7 responsible for its higher affinity to the m⁷Gp₄ cap analogue could be found. However, we noticed that Asn223 in *Dm* eIF4E-5 (replaced by Gly in *Dm* eIF4E-1, -2, -3, -4, and -7) could extra interact with δ -phosphate group. Finally, we observed increased binding affinities of *Dm* eIF4E-4 and *Dm* eIF4E-5 to *Dm* eIF4G peptides,

which may be explained by the presence of Glu55 and Glu58 in *Dm* eIF4E-4 and *Dm* eIF4E-5, respectively. These glutamic acids (absent in other isoforms, for which biochemical data were available) may form additional interactions with two *Dm* eIF4G conserved lysines, namely Lys618 and Lys 619. To conclude, using fluorescence binding studies combined with sequence/structure analyses we demonstrated shared and distinct features of cap binding by *Drosophila* eIF4E family members. Our studies showed that some local sequence differences play an important role in the eIF4E-cap interactions that might have an impact on their biological role.

“Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity toward CTP and increased tendency for cytidine incorporation at the 3’-terminal positions of synthesized tails” (16)

To answer the question what could be the structural basis for the observed *in vitro* preference of CutA enzymes toward CTP, we generated a 3D homology model of *Thielavia terrestris* CutA (TtCutA) and compared it to the experimentally solved structures of closely related enzymes: wild-type and H336N mutant of poly(U) polymerase Cid1 from *S. pombe* (17) and non-canonical poly(A) polymerase Trf4 from *S. cerevisiae* (18). We noticed that the analyzed CutA variants might interact mainly with the last three nucleotides of the RNA substrate. The interaction of the third nucleotide with the protein is the weakest, while the second nucleotide is most probably interacting mainly with the C-terminal, unstructured extension of the enzyme. As both AnCutA (CutA from *Aspergillus nidulans*) and TtCutA modify mRNA *in vivo*, they are evolutionarily designed to bind to RNAs ending with adenosines. Consequently, substrates with two pyrimidines at the 3’-end probably could not be strongly bound in their active sites. This observation, combined with the ability of these proteins to incorporate their preferred nucleotide - cytidine, may fully explain the observed phenomenon that the final product was almost always terminated with two cytidines. We concluded also that the TtCutA nucleotide recognition motif (NRM) region located between the fourth and fifth core α -helices of the PAP/OAS1 SBD remains likely the most important NTP-specificity determinant (NSD) of the analyzed enzymes. Having identified the NSD of CutA enzymes, we focused on verifying whether particular amino acids within NRM indeed play a role in the CTP specificity, and rationally designed mutations in this region of TtCutA. Results of high-throughput 3’-RACE-seq analysis, combined with structural inspection, demonstrated the critical importance of arginine 557, glutamic acid 551 and asparagine 554 of

TtCutA, located within NRM, for CTP-specificity of the examined enzyme. Furthermore, proposed R557A substitution enabled partial conversion of TtCutA nucleotidyltransferase activity from PCP to PAP, whereas R557H, E551D and N554E substitutions changed the enzyme specificity from PCP to PUP. Notably, however, while we successfully engineered proteins with significantly reduced or increased specific nucleotide selectivity, we were not able to construct the enzymes, which would use only one preferred nucleotide for the polymerization process. This may be caused by the character of the CutA enzymes' NRM region, which allows binding of all NTPs in the active site without a stringent preference for a single nucleotide. To further confirm our conclusions about the importance of particular amino acids within the NRM of CutA enzymes, and to extend our findings to other fungal non-canonical NTases, we attempted to construct the mutant of *S. pombe* Cid1 PUP with increased preference for cytidine. Thus, we designed Cid1 variant - Cid1^{D330E/E333N/H336R}, in which aspartic acid, glutamic acid and histidine, at positions 330, 333 and 336, equivalent to critical TtCutA positions 551, 554 and 557, were changed into arginine, glutamate and asparagine present in the latter protein, respectively. Such amino acid changes, designed to convert Cid1 PUP into PCP were rationalized not only by 3D structure modeling, but also justified by data obtained for TtCutA, demonstrating that the reciprocal substitutions within its NRM partially alter NTP specificity of this enzyme from CTP toward UTP. The 3'-terminal tails synthesized by Cid1^{WT} were composed almost exclusively of uridines (specificity >97%), as expected. On the contrary, Cid1^{D330E/E333N/H336R} variant generated 3'-extensions containing only ~50% of uridines. In parallel, incorporation of remaining nucleotides increased significantly, in the following order: C > A > G. Importantly, fraction of cytidines built into the tails achieved 30%–40%, indicating that the mutations that we designed indeed significantly increased cytidine and decreased uridines incorporation by Cid1 enzyme.

“FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases” (15)

FAM46 proteins, encoded in all known animal genomes, belong to the NTase fold superfamily. All four human FAM46 paralogs (FAM46A, FAM46B, FAM46C, FAM46D) were thought to be involved in several diseases, with FAM46C reported as a causal driver of multiple myeloma; however, their exact functions remained unknown. By using a combination of various bioinformatics analyses (e.g. domain architecture, cellular

localization) and exhaustive literature and database searches (e.g. expression profiles, protein interactors), we classified FAM46 proteins as active non-canonical poly(A) polymerases, which modify cytosolic and/or nuclear RNA 3' ends. Our predictions were later confirmed experimentally by Mroczek et *al.* using several *in vitro* and *in vivo* approaches. They proved that FAM46C encodes an active non-canonical poly(A) polymerase which enhances mRNA stability and gene expression (19). A detailed analysis of sequence and structure diversity of known NTases possessing PAP/OAS1 SBD domain, combined with state-of-the-art comparative modeling, allowed us to identify potential active site residues responsible for catalysis and substrate binding. We showed that FAM46 proteins share all the key motifs in the NTase domain responsible for catalysis and substrate binding, including [DE]h[DE]h and h[DE]h patterns with three conserved aspartate/glutamate residues (Asp90, Asp92 and Glu166) and hG[GS] motif with Gly73 and Ser74 in *human* FAM46C. Although we were not able to generate a complete 3D model of *human* FAM46C, we were able to identify additional residues responsible for NTP binding. Comparison of active sites of experimentally solved structures with human FAM46C revealed that Ser248 may bind, directly or indirectly via a water molecule, 2'-OH or/and 3'-OH hydroxyl group of a ribose-base moiety as it is observed for Thr172 in poly(U) polymerase Cid1 (17) or it can participate in a nucleobase binding similarly to Thr190 in 2'-5'-oligoadenylate synthase OAS1 (20). FAM46C shares also all the conserved residues in PAP/OAS1 SBD responsible for substrate binding. The FAM46C Leu282 probably interacts with a nucleobase of the incoming NTP like Tyr212 in poly(U) polymerase Cid1 or make van der Waals contacts with the ribose-base moiety of NTP similar to Val234 in poly(A) polymerase Pap1 (21). Finally, NTP β - and γ -phosphates most likely interact with the conserved Arg268 in addition to Ser74 from hG[GS] motif. The results of our sequence and structure analyses suggested that FAM46 proteins are active NTases, which have both NTase fold and PAP/OAS1 substrate-binding domains. Our studies were consistent with the M. Tian studies, where he showed that mutated *mouse* FAM46C may modulate poly(A) tails of specific transcripts during erythroid differentiation (22). We proposed that FAM46 proteins may play a very important role in mRNA stability as active non-canonical poly(A) polymerases rather than some other factors, which prevent early mRNA degradation by disrupting interactions between ribonuclease docking complex and RNA suggested by M. Tian. Our functional assignment was also in line with Chapman et *al.* studies (23). They proposed that FAM46C is functionally related in some way to the regulation of translation (e.g. as a mRNA stability factor).

Additionally in this work, we mapped reported numerous somatic mutations in cancer patients, leading to single point mutations, onto a 3D model of FAM46C and found that they can be divided into two groups. The first includes mutations that are located in a highly conserved area of the active site and its close vicinity, and probably may increase/decrease FAM46 catalytic activity and/or affect substrate-binding affinity. Mutations belonging to the second group are located mostly on the protein surface, in evolutionary low conserved region, probably having no crucial role in the reported cancer or affecting protein-protein interactions. Moreover, identification of all mutations of highly conserved residues in FAM46 proteins allowed us to detect mutations found in a number of malignancies, which probably are highly related with diagnosed cancer. Consequently, we suggested that, in addition to multiple myeloma, FAM46 genes may be also involved in pathogenesis of various other cancers including liver hepatocellular carcinoma, bladder urothelial carcinoma, head and neck squamous cell carcinoma, uterine corpus endometrial carcinoma, kidney renal papillary cell carcinoma, lung adenocarcinoma, ductal adenocarcinoma, colorectal adenocarcinoma, primary plasma cell leukemia and skin cutaneous melanoma.

Conclusions

In many cases homology modeling may play important role in examining proteins functions and interactions with ligands as generated models may provide valuable clues about investigated proteins. In my studies I showed that homology modeling might be very useful in: (1) explaining biochemical data related to substrate binding and specificity, (2) planning new biochemical experiments to change the enzyme specificity and (3) annotating protein function. Firstly, the use of differentially chemically modified cap analogues in binding studies and a detailed analysis of the obtained homology models gave better insight into the molecular characteristic of varying cap-binding abilities of *Arabidopsis* and *Drosophila* eIF4Es isoforms. Using 3D models of examined proteins, I was able to indicate crucial structural and sequence differences responsible for significant changes in binding affinities. Secondly, comparison of the 3D model of CutA with solved crystal structures of Cid1 and Trf4, combined with site-directed mutagenesis allowed to define an unusual nucleotide-binding site, in which the unique NRM present in the non-canonical PAPs and PUPs, is mainly responsible for the observed preference of CutA toward cytidine. Proposed Cid1^{D330E/E333N/H336R} variant significantly increased cytidine incorporation at the expense of uridine inclusion into the 3'-terminal tails. On the other hand, designed

TtCutA^{E551D/N554E/R557H} mutant generated the tails containing less cytidines and more uridines, but did not lead to the complete conversion of CutA into PUP enzyme. Importantly, I was not able to propose mutations converting CutA into highly efficient PAP, probably due to a unique geometry and environment of the nucleotide-binding site, which makes it impossible to achieve specificity toward a single nucleotide by only a few mutations. Finally, I showed that homology modeling combined with other bioinformatics analyses and biochemical data may be also adopted in the assignment of specific function to uncharacterized proteins. A comprehensive analysis of various biological information available in literature and databases combined with numerous sequence and structure analyses (including a state-of-the-art distant homology detection, fold recognition and 3D modeling) allowed me to propose that FAM46 family members function as cytoplasmic and/or nuclear non-canonical poly(A) polymerases. This finding was later confirmed experimentally (19).

References

1. Muhammed, M.T. and Aki-Yalcin, E. (2019) Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des*, **93**, 12-20.
2. Vyas, V.K., Ukawala, R.D., Ghate, M. and Chintha, C. (2012) Homology modeling a fast tool for drug discovery: current perspectives. *Indian J Pharm Sci*, **74**, 1-17.
3. Cavasotto, C.N. and Phatak, S.S. (2009) Homology modeling in drug discovery: current trends and applications. *Drug Discov Today*, **14**, 676-683.
4. Boissel, J.P., Lee, W.R., Presnell, S.R., Cohen, F.E. and Bunn, H.F. (1993) Erythropoietin structure-function relationships. Mutant proteins that test a model of tertiary structure. *J Biol Chem*, **268**, 15983-15993.
5. Sheng, Y., Sali, A., Herzog, H., Lahnstein, J. and Krilis, S.A. (1996) Site-directed mutagenesis of recombinant human beta 2-glycoprotein I identifies a cluster of lysine residues that are critical for phospholipid binding and anti-cardiolipin antibody activity. *J Immunol*, **157**, 3744-3751.
6. Ring, C.S., Sun, E., McKerrow, J.H., Lee, G.K., Rosenthal, P.J., Kuntz, I.D. and Cohen, F.E. (1993) Structure-based inhibitor design by using protein models for the development of antiparasitic agents. *Proc Natl Acad Sci U S A*, **90**, 3583-3587.
7. Xu, L.Z., Sanchez, R., Sali, A. and Heintz, N. (1996) Ligand specificity of brain lipid-binding protein. *J Biol Chem*, **271**, 24711-24719.
8. Sali, A., Matsumoto, R., McNeil, H.P., Karplus, M. and Stevens, R.L. (1993) Three-dimensional models of four mouse mast cell chymases. Identification of proteoglycan binding regions and protease-specific antigenic epitopes. *J Biol Chem*, **268**, 9023-9034.
9. Vakser, I.A. (1997) Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins*, **Suppl 1**, 226-230.
10. Howell, P.L., Almo, S.C., Parsons, M.R., Hajdu, J. and Petsko, G.A. (1992) Structure determination of turkey egg-white lysozyme using Laue diffraction data. *Acta Crystallogr B*, **48 (Pt 2)**, 200-207.
11. Guenther, B., Onrust, R., Sali, A., O'Donnell, M. and Kuriyan, J. (1997) Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III. *Cell*, **91**, 335-345.

12. Pisera, A., Campo, A. and Campo, S. (2018) Structure and functions of the translation initiation factor eIF4E and its role in cancer development and treatment. *J Genet Genomics*, **45**, 13-24.
13. Kropiwnicka, A., Kuchta, K., Lukaszewicz, M., Kowalska, J., Jemielity, J., Ginalski, K., Darzynkiewicz, E. and Zuberek, J. (2015) Five eIF4E isoforms from *Arabidopsis thaliana* are characterized by distinct features of cap analogs binding. *Biochem Biophys Res Commun*, **456**, 47-52.
14. Zuberek, J., Kuchta, K., Hernandez, G., Sonenberg, N. and Ginalski, K. (2016) Diverse cap-binding properties of *Drosophila* eIF4E isoforms. *Biochim Biophys Acta*, **1864**, 1292-1303.
15. Kuchta, K., Muszewska, A., Knizewski, L., Steczkiewicz, K., Wyrwicz, L.S., Pawlowski, K., Rychlewski, L. and Ginalski, K. (2016) FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res*, **44**, 3534-3548.
16. Kobylecki, K., Kuchta, K., Dziembowski, A., Ginalski, K. and Tomecki, R. (2017) Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity toward CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails. *RNA*, **23**, 1902-1926.
17. Lunde, B.M., Magler, I. and Meinhart, A. (2012) Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res*, **40**, 9815-9824.
18. Hamill, S., Wolin, S.L. and Reinisch, K.M. (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc Natl Acad Sci U S A*, **107**, 15045-15050.
19. Mroczek, S., Chlebowska, J., Kulinski, T.M., Gewartowska, O., Gruchota, J., Cysewski, D., Liudkovska, V., Borsuk, E., Nowis, D. and Dziembowski, A. (2017) The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat Commun*, **8**, 619.
20. Lohofener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D.J. and Fedorov, R. (2015) The Activation Mechanism of 2'-5'-Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. *Structure*, **23**, 851-862.

21. Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L. and Bohm, A. (2000) Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science*, **289**, 1346-1349.
22. Tian, M. (2010) The molecular cloning and characterization of Fam46c RNA stability factor. *Harvard University; PhD dissertation*.
23. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467-472.



Five eIF4E isoforms from *Arabidopsis thaliana* are characterized by distinct features of cap analogs binding



Anna Kropiwnicka^{a,b}, Krzysztof Kuchta^{a,c}, Maciej Lukaszewicz^b, Joanna Kowalska^b, Jacek Jemielity^{b,c}, Krzysztof Ginalski^c, Edward Darzynkiewicz^{b,c}, Joanna Zuberek^{b,*}

^a College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland

^b Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland

^c Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

ARTICLE INFO

Article history:

Received 10 November 2014

Available online 18 November 2014

Keywords:

eIF4E

Cap analog

Binding affinity

Fluorescence titration

Homology modeling

ABSTRACT

The assembly of the ribosome on majority of eukaryotic mRNAs is initiated by the recruitment of eIF4E protein to the mRNA 5' end cap structure. Flowering plants use two eIF4E isoforms, named eIF4E and eIF(iso)4E, as canonical translation initiation factors and possess a homolog of mammalian 4EHP (or eIF4E-2) termed nCBP. Plants from *Brassicaceae* family additionally conserve a close paralog of eIF4E which in *Arabidopsis thaliana* has two copies named eIF4E1b and eIF4E1c. In order to assess the efficiency of plant non-canonical (eIF4E1b/1c and nCBP) and canonical (eIF4E and eIF(iso)4E) eIF4E proteins to bind mRNAs we utilized fluorescence titrations to determine accurate binding affinities of five *A. thaliana* eIF4E isoforms for a series of cap analogs. We found that eIF4E binds cap analogs from 4-fold to 10-fold stronger than eIF(iso)4E, while binding affinities of nCBP and eIF(iso)4E are comparable. Furthermore, eIF4E1c interacts similarly strongly with the cap as eIF4E, but eIF4E1b binds cap analogs ca. 2-fold weaker than eIF4E1c, regardless of the 95% sequence identity between these two proteins. The use of differentially chemically modified cap analogs in binding studies and a detailed analysis of the obtained homology models gave us insight into the molecular characteristic of varying cap-binding abilities of *Arabidopsis* eIF4E isoforms.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The recognition of the cap (m^7GpppN , where N is any nucleoside) present at the 5' end of eukaryotic mRNAs by eIF4F complex is a key point of gene expression regulation at the level of translation initiation. The eIF4F cap-binding complex consists of eIF4E protein, which recognizes the cap, and the scaffolding protein eIF4G that binds other factors to recruit the ribosome to the mRNA. In mammals eIF4F complex additionally contains eIF4A helicase which unwinds the mRNA 5' secondary structure [1].

In many organisms multiple eIF4E homologs have been identified, which are grouped into three classes [2]. Flowering plants conserve two canonical translation initiation factors: eIF4E and eIF(iso)4E, both belonging to class I, and one non-canonical class II isoform termed nCBP. eIF(iso)4E together with its counterpart eIF(iso)4G make up a plant-specific isozyme form of the eIF4F complex named eIF(iso)4F [3]. Available data (including differential expression patterns for eIF4E and eIF(iso)4E [4–7]) suggests some

mRNAs are preferentially translated using either eIF4F or eIF(iso)4F [8–11], which could contribute to selective translation of transcripts during plant growth and development.

In certain plant species additional paralogs of eIF4E or eIF(iso)4E are present, such as a close homolog of eIF4E: eIF4E1b, which is conserved in *Brassicaceae* and in *Arabidopsis* has diverged into two copies termed eIF4E1b and eIF4E1c [12]. *Arabidopsis thaliana* eIF4E1b, eIF4E1c and nCBP are able to bind m^7GTP -Sepharose, interact with plant 4G-type isoforms, and to promote translation initiation in yeast (eIF4E1b and eIF4E1c) and/or in wheat germ extracts (eIF4E1b, eIF4E1c and nCBP) [12,13]. However, based on their low levels of expression they do not seem to function as general translation initiation factors. Instead, *eIF4E1b* gene up-regulation in pollen and developing embryos [12] implies its possible role in processes of reproduction. Similarly, the increased expression of nCBP in early stages of cell growth [5] provides evidence for its yet unknown specialized function.

In this study we examine the ability of *A. thaliana* eIF4E isoforms (AtelF4E, AtelF4E1b, AtelF4E1c, AtelF(iso)4E and AtnCBP) to bind cap analogs and discuss them in relation to 3D protein structures obtained with homology modeling.

* Corresponding author. Fax: +48 225540771.

E-mail address: jzuberek@biogeo.uw.edu.pl (J. Zuberek).

2. Materials and methods

2.1. Protein structure homology modeling

Homologs of *A. thaliana* eIF4E proteins were identified with PSI-Blast [14] searches (*E*-value threshold of 0.005) performed against the NCBI nonredundant protein sequence database using sequences of AtelF4E, AtelF4E1b, AtelF4E1c, AtelF(iso)4E, AtnCBP and mouse eIF4E as queries. Multiple sequence alignment was derived using PCMA program [15] followed by some manual adjustments. Secondary structures were predicted with PSIPRED [16]. The 3D models of *Arabidopsis* eIF4Es were constructed with MODELLER [17] using the closest homologs of known structure as templates [18]: *Pisum sativum* eIF4E (PDB: 2WMC) and *Mus musculus* eIF4E (PDB: 1EJ1) for AtelF4E, AtelF4E1b, AtelF4E1c and AtelF(iso)4E; *Homo sapiens* 4EHP (PDB: 2JGB), *P. sativum* eIF4E (PDB: 2WMC) and *M. musculus* eIF4E (PDB: 1EJ1) for AtnCBP.

2.2. Protein expression and purification

A. thaliana eIF4E proteins (Fig. S1) were prepared as recombinant proteins without any affinity tags and purified without a contact with cap analogs as described previously [19]. In the case of purification of AtelF4E1b, AtelF4E1c and AtnCBP the protocol was modified in order to remove co-purifying nucleic acids using benzonase, silica and ion-exchange chromatography during purification. The presence of nucleic acids was detected by absorption (Fig. S2). Proteins concentration was determined by absorption using extinction coefficients calculated by ProtParam tool (Expasy server).

2.3. Fluorescence titrations

Fluorescence time-synchronized titrations [20] were carried out on a spectrofluorometer LS-55 (Perkin Elmer Co.) at 20 ± 0.3 °C in 50 mM HEPES/KOH pH 7.2, 134.5 mM KCl, 0.5 mM EDTA, 1 mM DTT. Equilibrium association constants (K_{AS}) for complexes of eIF4E with structurally modified cap analogs (Table S1) were determined as described previously [20]. The fitting procedure employing the non-linear least-squared regression analysis was performed using ORIGIN 6.0 (Microcal Software). The final K_{AS} was calculated as a weighted average of 3–8 independent titrations.

3. Results

3.1. Structures of *Arabidopsis* eIF4E proteins predicted by homology modeling

Several structures of eIF4E from human, mouse, fruit fly, pea, wheat and yeast have been solved, most often with various cap analogs. The eIF4E-like fold is characterized by the presence of an eight-stranded antiparallel β -sheet backed by α -helices (with $\beta 1\alpha\beta 2\alpha\beta 3\beta 4\alpha\beta 5\beta 6\alpha\beta 7\alpha\beta 8$ topology) [21,22]. The interatomic contacts between eIF4E and cap analogs can be divided into three classes: (i) sandwiching of the alkylated base between two tryptophans (in class I eIF4E isoforms) or a tryptophan and a tyrosine (in class II), (ii) hydrogen bonds and van der Waals contacts with N(7)-methylguanosine and with the second nucleoside, (iii) direct interactions and water-mediated contacts with the phosphate chain [20–22].

We built 3D homology models of *Arabidopsis* eIF4E isoforms to investigate differences in their structures (Fig. 1) and ability to bind the cap considering the sequence diversity (Fig. 1D) and the variety of contacts between protein and ligand. All analyzed proteins show the conserved eIF4E-like fold and the cap-binding pocket characteristic of eIF4E class I [21] and II [23] proteins. The sequence/structural differences that may influence the interaction

with the cap are discussed below in relation to obtained results from the binding studies with cap analogs.

3.2. eIF(iso)4E has the lowest binding affinity for the cap among class I *Arabidopsis* eIF4E isoforms

All *Arabidopsis* eIF4Es, similarly to the mouse protein, exhibit quenching of tryptophan fluorescence upon binding of a cap analog (Fig. 2) and their binding affinities for the cap are from one to two orders of magnitude lower than those of mouse eIF4E (Table 1). The highest cap binding affinities are for AtelF4E and surprisingly for the non-canonical eIF4E isoform AtelF4E1c, and are comparable to values obtained for the yeast protein (Zuberek and Stelmachowska, unpublished result). In spite of 95% amino acid sequence identity between AtelF4E1b and AtelF4E1c (Fig. 1D), almost all cap analogs, except m⁷GMP, have association constants (K_{AS}) for AtelF4E1b ca. 2-fold lower than for AtelF4E1c. The majority of residues that differ between these two proteins are located within or near the cap-binding region (Fig. 1A). The main possible reason for the difference in their binding affinity may be the replacement of large hydrophobic Phe147 and Thr145 in AtelF4E1c with Met147 and Ser145 in AtelF4E1b, which could result in weaker stabilization of the conserved arginine (Arg188 in AtelF4E1b/AtelF4E1c, equivalent to Lys162 in mammalian eIF4E) that binds cap's β - and γ -phosphate groups. In addition, Thr184 in AtelF4E1c, that may form hydrogen bonding (i.e. through water molecules) with the γ -phosphate, is replaced by Ala184 in AtelF4E1b.

Further unexpectedly, the second canonical translation initiation factor, AtelF(iso)4E protein, binds all cap analogs from 4-fold to 10-fold weaker than AtelF4E (Table 1). Probably, the poorer cap-binding efficiency of AtelF(iso)4E is mainly caused by the shortening of the long loop (that contains a short α -helix) between β -strands $\beta 7$ and $\beta 8$ (Fig. 1B). This loop is longer by six amino acids in AtelF4E, AtelF4E1b and AtelF4E1c, and forms the edge of the cap binding cavity, taking part in tight binding of the sugar and β - and γ -phosphate moieties of cap analogs. For instance, AtelF(iso)4E lacks a conserved aspartate (Asp226 in AtelF4E) from this region that interacts with the ribose of N(7)-methylguanosine.

All analyzed class I plant eIF4E proteins are specific to N(7)-monomethylated cap analogs because they interact poorly with either GTP or m³,^{2,7}GTP (Table 1) and this result contradicts the one previously reported [24] (see Section 4). Similarly to mammalian eIF4E K_{AS} increases with the extension of the cap's phosphate chain, while the addition of a second nucleoside to m⁷GTP has a negative effect on interaction with *Arabidopsis* eIF4E isoforms.

3.3. nCBP binds cap similarly weakly as eIF(iso)4E

Arabidopsis class II eIF4E isoform, nCBP protein, interacts weakly with the cap, with the K_{AS} for m⁷GTP 8.5-fold lower than that of AtelF4E (Table 1). In contrast to Ruud et al. paper [13] we obtained binding affinities of AtnCBP for N(7)-methylated cap analogs similar to those observed for AtelF(iso)4E, with the differences not exceeding 1.5-fold, and to values reported for human 4EHP protein [25]. The most prominent difference that may be responsible for the poor binding of the cap by AtnCBP seems to be the lack of the conserved arginine (Arg183 in AtelF4E, equivalent to Lys162 in mammalian eIF4E), that may interact with the cap's phosphate moiety, which is substituted in AtnCBP with isoleucine (Ile166, conserved in close homologs) (Fig. 1C). The lack of arginine in this position can be partially compensated in AtnCBP by Arg125 and Lys202, which are replaced by uncharged amino acids in AtelF4E, AtelF4E1b and AtelF4E1c (i.e. Thr142 and Ile217 in AtelF4E, respectively). However, AtelF4E, AtelF4E1b and AtelF4E1c possess also other positively charged residues that may take part in the binding

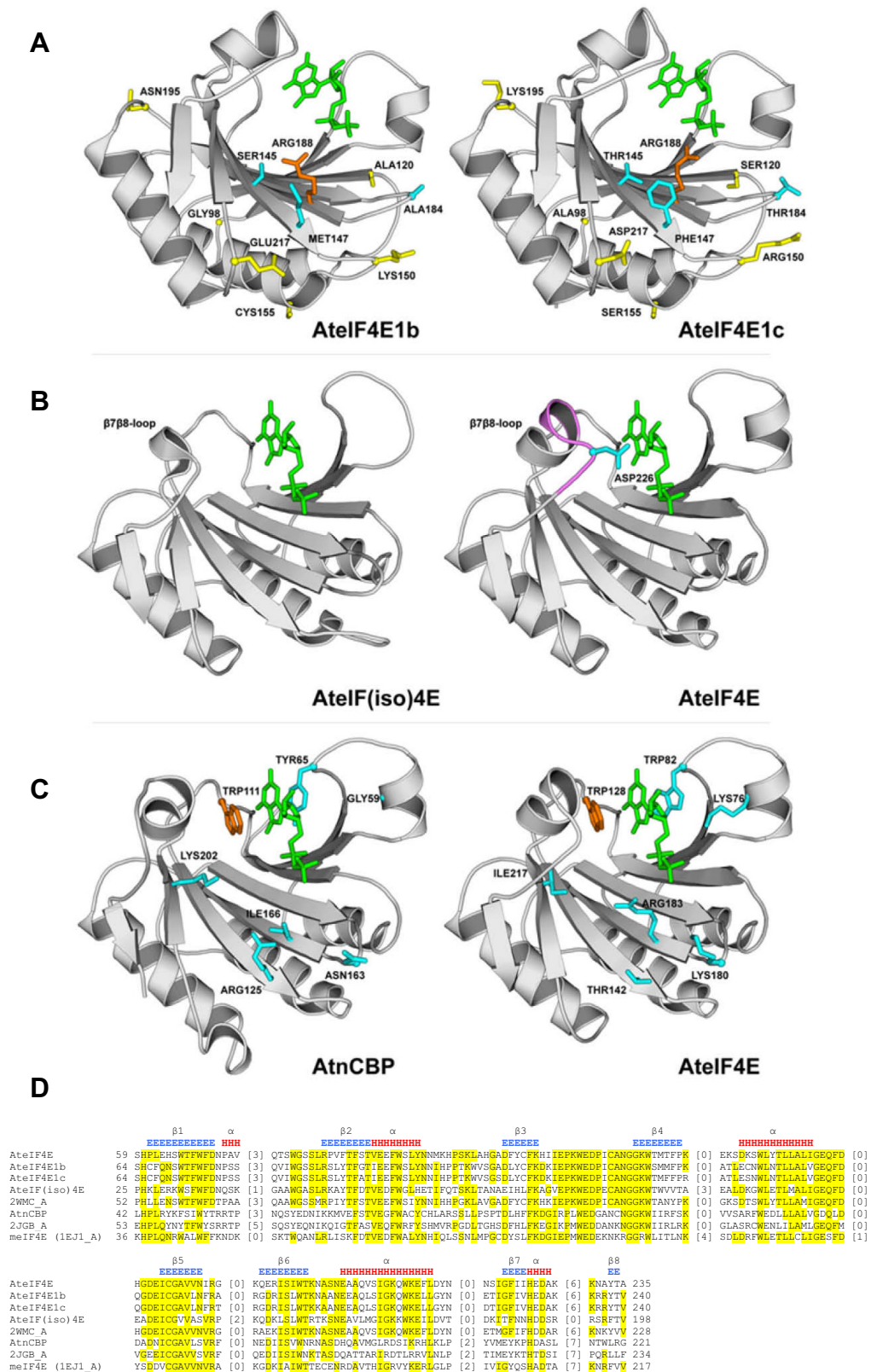


Fig. 1. Comparison of *A. thaliana* eIF4E proteins. (A–C) Comparison of modeled structures. Cap analog (m^7GDP) taken from PDB: 1EJ1 structure (*M. musculus* eIF4E) is shown in green. Amino acids critical for the difference in the binding affinity for cap analogs are denoted in cyan. (A) All remaining residues that differ between AtelF4E1b and AtelF4E1c are presented in yellow. Conserved Arg188 is denoted in orange. (B) Insertion of six amino acids in the loop between β -strands $\beta 7$ and $\beta 8$ is denoted in violet. (C) Conserved tryptophan is colored orange. (D) Multiple sequence alignment for *A. thaliana* eIF4E isoforms and their closest homologs of known structure (pea eIF4E, PDB: 2WMC; mouse eIF4E, PDB: 1EJ1; human 4EHP, PDB: 2JGB). The numbers of excluded residues are specified in square brackets. Conserved residues (50% or greater) are highlighted in yellow. Locations of secondary structure elements (β , β -strand; H, α -helix) in PDB: 2WMC are marked above sequences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

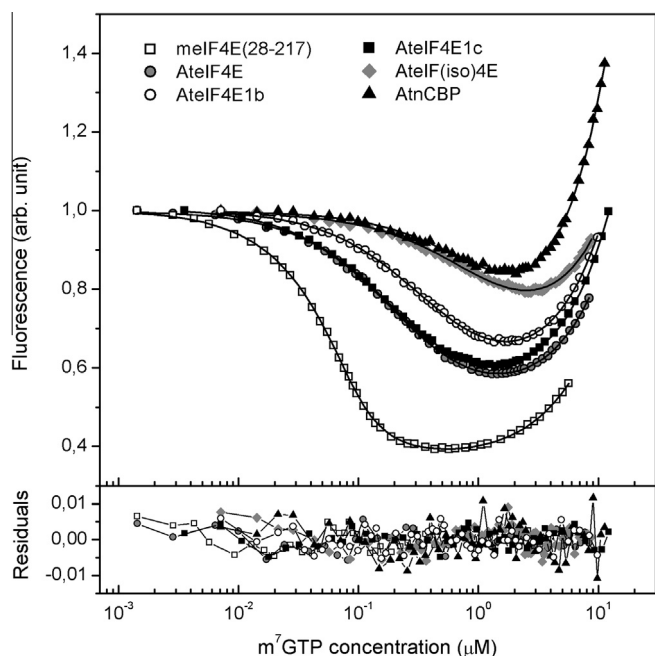


Fig. 2. Representative fluorescence titration curves for m^7 GTP binding to *A. thaliana* eIF4E isoforms. Fitted curves (black lines) and experimental data points (upper panel) with fitting residuals (lower panel) for *A. thaliana* eIF4E (●), eIF4E1b (○), eIF4E1c (■), eIF(iso)4E (◆) and nCBP (▲) and mouse eIF4E (□). The increasing fluorescence signal at higher concentrations of m^7 GTP originates from the fluorescence of a free cap analog in solution.

of cap's phosphate groups and which are not conserved in AtnCBP. This includes Lys76 and Lys180 in AtelF4E which are replaced in AtnCBP by Gly59 and Asn163.

Characteristic molecular feature of AtnCBP is its lower specificity toward monomethylation of guanosine at N(7), observed also for human 4EHP [25]. The difference between K_{AS} values for m^7 GTP and GTP is only 10-fold for AtnCBP, whereas other plant proteins bind m^7 GTP from 50-fold to 400-fold stronger than GTP (Table 1). This property is most likely not caused by the tryptophan-to-tyrosine substitution, that is specific to class II eIF4E proteins, at the residue involved in the sandwiching of the N(7)-methylguanine (Tyr65 in AtnCBP, equivalent to Trp82 in AtelF4E or Trp56 in mammalian eIF4E). Human eIF4E wild type and Trp56Tyr mutant proteins show similar cap-binding efficiencies [25].

3.4. Contribution of cap's phosphate groups into stabilizing the eIF4E-cap complex

Due to the location of basic amino acids at the entrance to the eIF4E cap-binding pocket, the more phosphate groups are linked

to N(7)-methylguanosine, the tighter eIF4E binds the cap. When in the cap analog the phosphate chain is extended from one (m^7 GMP) to four groups (m^7 Gp₄), its binding affinity for mouse eIF4E increases 600-fold (Table 1). The changes in K_{AS} accompanying the lengthening of the cap's phosphate chain are less prominent for *Arabidopsis* eIF4E proteins. Yet, similarly to mouse eIF4E, the highest rise in the binding affinity is observed for the addition of the β -phosphate (m^7 GMP $\rightarrow$ m^7 GDP). Only AtelF4E1c binds the β -phosphate moiety comparably strongly as mouse eIF4E (20-fold increase in K_{AS}) while K_{AS} of AtelF4E and AtelF4E1b for m^7 GDP are only 10-fold higher than for m^7 GMP. However, the ca. -1.4 kcal/mol energetic gain of introducing the β -phosphate into the cap, evaluated as the change in the Gibbs free energy of binding ($\Delta\Delta G^\circ$, Table 2), suggests AtelF4E and AtelF4E1b, as well as AtelF4E1c, probably create two hydrogen bonds with the β -phosphate, similarly to mammalian eIF4E [22]. On the contrary, for AtnCBP the corresponding $\Delta\Delta G^\circ$ of ca. -0.5 kcal/mol is less than the value typical for a hydrogen bond [26], which indicates AtnCBP's interaction with the β -phosphate is probably indirect. In comparison, for human 4EHP the binding energy of the β -phosphate moiety equals ca. -0.7 kcal/mol [25] and corresponds to one direct hydrogen bond and one mediated by a water molecule [23]. The changes in the binding energy accompanying the extension of the cap's phosphate chain to a triphosphate (m^7 GDP $\rightarrow$ m^7 GTP) imply all *Arabidopsis* eIF4E isoforms except AtelF4E probably form one hydrogen bond with the cap's γ -phosphate group, similarly to mammalian eIF4E [22] and 4EHP [23]. AtelF4E interacts weakly with either γ - or δ -phosphate, while only for AtnCBP and mouse eIF4E the change in the binding energy which might correspond to the energy of forming one hydrogen bond with the δ -phosphate is observed.

3.5. Binding affinities of phosphorothioate cap analogs for *Arabidopsis* eIF4E isoforms

Previous studies investigated the properties of cap analogs in which one of non-bridging oxygens in the phosphate chain was substituted with a sulfur atom. It was found out these phosphorothioate cap analogs have strong affinity for mouse eIF4E and some of them are resistant to enzymatic cleavage by decapping enzymes (DcpS or Dcp2) [27,28]. Therefore, they act as strong inhibitors of translation and are able to increase stability and translational efficiency of transcripts [28–30]. In this work we examined the interaction of *Arabidopsis* eIF4E proteins with m^7 GpppG analogs modified with the phosphorothioate moiety at the γ - or β -phosphate position, abbreviated m^7 Gp_sppG and m^7 Gpp_spG, respectively (Table 1). Due to a stereogenic P center in the phosphorothioate moiety each of the cap analogs exists as two diastereomers, termed D1 and D2. The presence of the phosphorothioate moiety at the γ -position in the diastereomer D1 (m^7 Gp_sppG (D1)) causes a 2-fold increase in K_{AS} for AtelF4E or AtelF(iso)4E,

Table 1
Equilibrium association constants (K_{AS}) of *A. thaliana* eIF4E proteins and mouse eIF4E in complex with cap analogs.

Cap analog	AtelF4E	AtelF4E1b	AtelF4E1c	AtelF(iso)4E	AtnCBP	melf4E(28–217)
	K_{AS} (μM^{-1})					
m^7 GMP	0.319 ± 0.016	0.120 ± 0.007	0.118 ± 0.003	<0.050	0.064 ± 0.004	0.82 ± 0.02
m^7 GDP	3.54 ± 0.03	1.23 ± 0.03	2.47 ± 0.04	0.280 ± 0.019	0.148 ± 0.005	20.6 ± 1.9
m^7 GTP	5.79 ± 0.13	3.12 ± 0.06	5.72 ± 0.10	1.07 ± 0.05	0.684 ± 0.014	81.5 ± 1.7
m^7 Gp ₄	6.9 ± 0.4	3.75 ± 0.07	9.02 ± 0.17	1.62 ± 0.04	1.81 ± 0.03	500 ± 30
m^7 GpppG	2.08 ± 0.04	1.219 ± 0.012	2.48 ± 0.08	0.188 ± 0.011	0.263 ± 0.012	7.20 ± 0.18
m^7 Gp _s ppG (D1)	4.69 ± 0.13	ND	ND	0.401 ± 0.008	0.229 ± 0.011	22.3 ± 0.4
m^7 Gp _s ppG (D2)	2.29 ± 0.05	ND	ND	0.213 ± 0.007	0.295 ± 0.010	7.60 ± 0.19
m^7 Gpp _s pG (D1)	2.7 ± 0.3	ND	ND	0.262 ± 0.007	0.350 ± 0.015	37.0 ± 1.7
m^7 Gpp _s pG (D2)	2.68 ± 0.18	ND	ND	0.233 ± 0.003	0.326 ± 0.007	17.4 ± 1.1
GTP	<0.015	<0.053	<0.033	<0.015	0.079 ± 0.007	<0.030
$m^3_{2,7}$ GTP	<0.040	<0.024	<0.038	<0.022	<0.050	ND

ND – not determined.

Table 2

Changes of standard Gibbs free energy ($\Delta\Delta G^\circ$) for the interaction of *A. thaliana* eIF4E isoforms with structurally modified cap analogs. $\Delta\Delta G^\circ$ values were calculated as the difference in the energy of binding (ΔG°) of eIF4E to cap analogs which vary in single structural alteration. ΔG° values were obtained from K_{AS} values according to the standard equation: $\Delta G^\circ = -RT \ln K_{AS}$.

Modifications in cap structure	AtelF4E	AtelF4E1b	AtelF4E1c	AtelF(iso)4E	AtnCBP	melF4E(28–217)
$\Delta\Delta G^\circ$ (kcal/mol)						
Successive addition of the phosphate groups: $\Delta\Delta G^\circ = \Delta G^\circ(m^7Gp_{n+1}) - \Delta G^\circ(m^7Gp_n)$						
$m^7GMP \rightarrow m^7GDP$	-1.40 ± 0.03	-1.35 ± 0.05	-1.77 ± 0.02	> -1.0	-0.49 ± 0.06	-1.88 ± 0.07
$m^7GDP \rightarrow m^7GTP$	-0.287 ± 0.018	-0.54 ± 0.02	-0.49 ± 0.02	-0.78 ± 0.07	-0.89 ± 0.03	-0.80 ± 0.07
$m^7GTP \rightarrow m^7Gp_4$	-0.10 ± 0.05	-0.11 ± 0.02	-0.27 ± 0.02	-0.24 ± 0.04	-0.57 ± 0.02	-1.06 ± 0.04
Addition of the second nucleoside: $\Delta\Delta G^\circ = \Delta G^\circ(m^7Gp_nG) - \Delta G^\circ(m^7Gp_n)$						
$m^7GTP \rightarrow m^7GpppG$	0.60 ± 0.02	0.547 ± 0.017	0.49 ± 0.03	1.01 ± 0.06	0.56 ± 0.04	1.41 ± 0.03
Substitution of the phosphate moiety with the phosphorothioate: $\Delta\Delta G^\circ = (\Delta G^\circ(m^7Gp_{pp}G) \text{ or } \Delta G^\circ(m^7Gp_{ps}G)) - \Delta G^\circ(m^7Gp_{pp}G)$						
$m^7GpppG \rightarrow m^7Gp_{ps}ppG$ (D1)	-0.47 ± 0.03	ND	ND	-0.44 ± 0.05	0.08 ± 0.05	-0.66 ± 0.02
$m^7GpppG \rightarrow m^7Gp_{ps}ppG$ (D2)	-0.05 ± 0.02	ND	ND	-0.07 ± 0.05	-0.07 ± 0.05	-0.03 ± 0.03
$m^7GpppG \rightarrow m^7Gp_{ps}pG$ (D1)	-0.15 ± 0.08	ND	ND	-0.19 ± 0.05	-0.16 ± 0.05	-0.95 ± 0.04
$m^7GpppG \rightarrow m^7Gp_{ps}pG$ (D2)	-0.15 ± 0.05	ND	ND	-0.12 ± 0.04	-0.12 ± 0.04	-0.51 ± 0.05

ND – not determined.

while the diastereomer D2 is bound with a comparable affinity as the unmodified analog. No significant changes in K_{AS} are observed upon the oxygen-to-sulfur substitution at the γ -position for AtnCBP or at the β -position for any of *Arabidopsis* proteins. It was proposed that the increased polarizability of the phosphorothioate moiety, in comparison to the phosphate group, causes a shift of the negative charge from the sulfur to the oxygen. In consequence, when this oxygen is engaged in a hydrogen bond with the protein, i.e. for only one of diastereomers, the interaction is enhanced [27]. The lack of a significant increase in the binding affinity of AtnCBP for $m^7Gp_{ps}ppG$ analogs suggests that oxygens from the γ -phosphate group of the cap (or α -phosphate in the mononucleotide cap) interact weakly with this protein, while AtelF4E and AtelF(iso)4E bind the γ -phosphate comparably strongly as mouse protein. The detected smaller changes in the binding affinity of AtelF4E, AtelF(iso)4E and AtnCBP for the cap upon its modification with the phosphorothioate at the β -position, in comparison to mouse eIF4E, are in accordance with the conclusion of a weaker interaction of these plant proteins with the cap's β -phosphate, which was based on the results for mononucleotide cap analogs.

4. Discussion

In this work we were able to successfully address in a quantitative manner the question of the ability of *A. thaliana* proteins from the eIF4E family to interact with the cap structure. Although there are reports of binding studies for plant eIF4Es with cap analogs or oligonucleotides [13,24,31], these studies are not free of methodological faults (discussed in [20]), the main one being the use of eIF4E protein samples which were contaminated with remnants of a cap analog applied during protein purification procedures. Here we obtained eIF4E samples free of any cap analog contaminants. In general, binding affinities of plant eIF4Es for the cap can be arranged in the rank of a diminishing affinity: AtelF4E $\approx$ AtelF4E1c $>$ AtelF4E1b $>$ AtelF(iso)4E $\approx$ AtnCBP. The result of a 4-fold to 10-fold difference in binding affinities between the two canonical translation initiation 4E-type factors, eIF4E and eIF(iso)4E, provokes an important question of how these two proteins are able to compete for the mRNA 5'-end *in vivo*. One plausible explanation is the reduction of the difference once these proteins are assembled into appropriate 4F-type cap-binding complexes, based on known data supporting the increase of the binding affinity for the cap of eIF4E upon interaction with eIF4G [32]. Moreover, eIF4G and eIF(iso)4G might vary in the extent of their influence on cap-binding because of an additional N-terminal

domain in eIF4G which is not present in eIF(iso)4G [3]. Another explanation would be that other than cap specific elements of a transcript (such as secondary structure elements) could enhance the recruitment of either eIF4F or eIF(iso)4F to the mRNA.

Comparable binding affinities of AtelF4E and its closest two homologs: AtelF4E1b and AtelF4E1c, confirm AtelF4E1b and AtelF4E1c should be able to efficiently bind transcripts *in vivo*, which is in accordance with their ability to support translation in yeast [12]. Although AtnCBP binds cap analogs with similar strength as the second canonical factor, AtelF(iso)4E protein, and was shown to interact with eIF4G and eIF(iso)4G in wheat [13], in *Arabidopsis* it probably does not form a complex with eIF4G (Patrick and Browning, unpublished observation mentioned in [3]). Also AtelF4E1b and AtelF4E1c do not seem likely to form respective 4F-type complexes due to their very low affinity for eIF4G [12]. Therefore, it is more probable that eIF4E1b, eIF4E1c and nCBP interact with yet unknown mRNA-specific factors and function rather as translational regulators under specific developmental (i.e. the development of pollen and embryos for eIF4E1b [12]) or stress-related conditions, maybe in a manner similar to *Xenopus* oocyte-specific class I eIF4E1b [33], *Drosophila* testis-specific class I eIF4E3 [34], or mammalian and *Drosophila* 4EHP [35–39].

Acknowledgments

We are grateful for providing cDNAs for AtelF4E, AtelF(iso)4E and AtnCBP by Christophe Robaglia (CNRS-CEA-Université Aix-Marseille, France), AtelF4E1b and AtelF4E1c by Carole Caranta (INRA, France) and mouse eIF4E(28–217) by Nahum Sonenberg (McGill University, Canada). This research was conducted at the Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw (AK) and at the Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw (KK) and was supported by the National Science Centre (Poland, UMO-2012/07/B/NZ1/00118 to ED, UMO-2011/02/A/NZ2/00014 to KG, UMO-2012/05/E/ST5/03893 to JJ), Foundation For Polish Science (TEAM/2010-6 to KG), the National Centre for Research and Development (Poland, PBS1/A9/16/2012 to KG) and by the European Social Fund (EU, UDA-POKL.04.01.01-00-072/09-00 to AK and KK).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bbrc.2014.11.032>.

References

- [1] N. Sonenberg, J.W.B. Hershey, M. Mathews (Eds.), *Translational Control in Biology and Medicine*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y., 2007.
- [2] B. Joshi, K. Lee, D.L. Maeder, et al., Phylogenetic analysis of eIF4E-family members, *BMC Evol. Biol.* 5 (2005) 48, <http://dx.doi.org/10.1186/1471-2148-5-48>.
- [3] R.M. Patrick, K.S. Browning, The eIF4F and eIFiso4F complexes of plants: an evolutionary perspective, *Comp. Funct. Genomics* 2012 (2012) 287814, <http://dx.doi.org/10.1155/2012/287814>.
- [4] C.M. Rodriguez, M.A. Freire, C. Camilleri, et al., The *Arabidopsis thaliana* cDNAs coding for eIF4E and eIF(iso)4E are not functionally equivalent for yeast complementation and are differentially expressed during plant development, *Plant J.* 13 (1998) 465–473, <http://dx.doi.org/10.1046/j.1365-313X.1998.00047.x>.
- [5] M.S. Bush, A.P. Hutchins, A.M.E. Jones, et al., Selective recruitment of proteins to 5' cap complexes during the growth cycle in *Arabidopsis*, *Plant J.* 59 (2009) 400–412, <http://dx.doi.org/10.1111/j.1365-313X.2009.03882.x>.
- [6] D.R. Gallie, H. Le, R.L. Tanguay, et al., Translation initiation factors are differentially regulated in cereals during development and following heat shock, *Plant J.* 14 (1998) 715–722, <http://dx.doi.org/10.1046/j.1365-313X.1998.00175.x>.
- [7] J.P. Combe, M.E. Petracek, G. van Eldik, et al., Translation initiation factors eIF4E and eIFiso4E are required for polysome formation and regulate plant growth in tobacco, *Plant Mol. Biol.* 57 (2005) 749–760, <http://dx.doi.org/10.1007/s11103-005-3098-x>.
- [8] D.R. Gallie, K.S. Browning, EIF4G functionally differs from eIFiso4G in promoting internal initiation, cap-independent translation, and translation of structured mRNAs, *J. Biol. Chem.* 276 (2001) 36951–36960, <http://dx.doi.org/10.1074/jbc.M103869200>.
- [9] L.K. Mayberry, M.L. Allen, M.D. Dennis, et al., Evidence for variation in the optimal translation initiation complex: plant eIF4B, eIF4F, and eIF(iso)4F differentially promote translation of mRNAs, *Plant Physiol.* 150 (2009) 1844–1854, <http://dx.doi.org/10.1104/pp.109.138438>.
- [10] L.K. Mayberry, M.L. Allen, K.R. Nitka, et al., Plant cap-binding complexes eukaryotic initiation factors eIF4F and eIFiso4F: molecular specificity of subunit binding, *J. Biol. Chem.* 286 (2011) 42566–42574, <http://dx.doi.org/10.1074/jbc.M111.280099>.
- [11] A.V. Martínez-Silva, C. Aguirre-Martínez, C.E. Flores-Tinoco, et al., Translation initiation factor AtelF(iso)4E is involved in selective mRNA translation in *Arabidopsis thaliana* seedlings, *PLoS One* 7 (2012) e31606, <http://dx.doi.org/10.1371/journal.pone.0031606>.
- [12] R.M. Patrick, L.K. Mayberry, G. Choy, et al., Two *Arabidopsis thaliana* loci encode novel eIF4E isoforms that are functionally distinct from the conserved plant eIF4E, *Plant Physiol.* 164 (2014) 1820–1830, <http://dx.doi.org/10.1104/pp.113.227785>.
- [13] K.A. Ruud, C. Kuhlow, D.J. Goss, et al., Identification and characterization of a novel cap-binding protein from *Arabidopsis thaliana*, *J. Biol. Chem.* 273 (1998) 10325–10330, <http://dx.doi.org/10.1074/jbc.273.17.10325>.
- [14] S.F. Altschul, T.L. Madden, A.A. Schäffer, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402, <http://dx.doi.org/10.1093/nar/25.17.3389>.
- [15] J. Pei, R. Sadreyev, N.V. Grishin, PCMA: fast and accurate multiple sequence alignment based on profile consistency, *Bioinformatics* 19 (2003) 427–428, <http://dx.doi.org/10.1093/bioinformatics/btg008>.
- [16] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, *Bioinformatics* 16 (2000) 404–405, <http://dx.doi.org/10.1093/bioinformatics/16.4.404>.
- [17] A. Fiser, A. Sali, Modeller: generation and refinement of homology-based protein structure models, *Methods Enzymol.* 374 (2003) 461–491, [http://dx.doi.org/10.1016/S0076-6879\(03\)74020-8](http://dx.doi.org/10.1016/S0076-6879(03)74020-8).
- [18] K. Ginalski, L. Rychlewski, Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment, *Proteins* 53 (Suppl. 6) (2003) 410–417, <http://dx.doi.org/10.1002/prot.10548>.
- [19] J. Zuberek, A. Wyslouch-Cieszyńska, A. Niedzwiecka, et al., Phosphorylation of eIF4E attenuates its interaction with mRNA 5' cap analogs by electrostatic repulsion: intein-mediated protein ligation strategy to obtain phosphorylated protein, *RNA* 9 (2003) 52–61, <http://dx.doi.org/10.1261/rna.2133403.tiation>.
- [20] A. Niedzwiecka, J. Marcotrigiano, J. Stepinski, et al., Biophysical studies of eIF4E cap-binding protein: Recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins, *J. Mol. Biol.* 319 (2002) 615–635, [http://dx.doi.org/10.1016/S0022-2836\(02\)00328-5](http://dx.doi.org/10.1016/S0022-2836(02)00328-5).
- [21] J. Marcotrigiano, A.-C. Gingras, N. Sonenberg, et al., Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP, *Cell* 89 (1997) 951–961, [http://dx.doi.org/10.1016/S0092-8674\(00\)80280-9](http://dx.doi.org/10.1016/S0092-8674(00)80280-9).
- [22] K. Tomoo, X. Shen, K. Okabe, et al., Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations, *J. Mol. Biol.* 328 (2003) 365–383, [http://dx.doi.org/10.1016/S0022-2836\(03\)00314-0](http://dx.doi.org/10.1016/S0022-2836(03)00314-0).
- [23] P. Rosettani, S. Knapp, M.G. Vismara, et al., Structures of the human eIF4E homologous protein, h4EHP, in its m7GTP-bound and unliganded forms, *J. Mol. Biol.* 368 (2007) 691–705, <http://dx.doi.org/10.1016/j.jmb.2007.02.019>.
- [24] S.E. Carberry, D.J. Goss, E. Darzynkiewicz, A comparison of the binding of methylated cap analogs to wheat germ protein synthesis initiation factors 4F and (iso) 4F, *Biochemistry* 30 (1991) 1624–1627, <http://dx.doi.org/10.1021/bi00220a026>.
- [25] J. Zuberek, D. Kubacka, A. Jablonowska, et al., Weak binding affinity of human 4EHP for mRNA cap analogs, *RNA* 13 (2007) 691–697, <http://dx.doi.org/10.1261/rna.453107>.
- [26] B.A. Shirley, P. Stanssens, U. Hahn, et al., Contribution of hydrogen bonding to the conformational stability of ribonuclease T1, *Biochemistry* 31 (1992) 725–732, <http://dx.doi.org/10.1021/bi00118a013>.
- [27] J. Kowalska, M. Lewdorowicz, J. Zuberek, et al., Synthesis and characterization of mRNA cap analogs containing phosphorothioate substitutions that bind tightly to eIF4E and are resistant to the decapping pyrophosphatase DcpS, *RNA* 14 (2008) 1119–1131, <http://dx.doi.org/10.1261/rna.990208>.
- [28] E. Grudzien-Nogalska, J. Jemielity, J. Kowalska, et al., Phosphorothioate cap analogs stabilize mRNA and increase translational efficiency in mammalian cells, *RNA* 13 (2007) 1745–1755, <http://dx.doi.org/10.1261/rna.701307>.
- [29] J. Kowalska, M. Lukaszewicz, J. Zuberek, et al., Phosphorothioate analogs of m7GTP are enzymatically stable inhibitors of cap-dependent translation, *Bioorg. Med. Chem. Lett.* 19 (2009) 1921–1925, <http://dx.doi.org/10.1016/j.bmcl.2009.02.053>.
- [30] A.N. Kuhn, M. Diken, S. Kreiter, et al., Phosphorothioate cap analogs increase stability and translational efficiency of RNA vaccines in immature dendritic cells and induce superior immune responses in vivo, *Gene Ther.* 17 (2010) 961–971, <http://dx.doi.org/10.1038/gt.2010.52>.
- [31] S.E. Carberry, D.J. Goss, Wheat germ initiation factors 4F and (iso)4F interact differently with oligoribonucleotide analogs of rabbit alpha-globin mRNA, *Biochemistry* 30 (1991) 4542–4545, <http://dx.doi.org/10.1021/bi00232a025>.
- [32] T. von Der Haar, P.D. Ball, J.E. McCarthy, Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5'-cap by domains of eIF4G, *J. Biol. Chem.* 275 (2000) 30551–30555, <http://dx.doi.org/10.1074/jbc.M004565200>.
- [33] N. Minshall, M.H. Reiter, D. Weil, et al., CPEB interacts with an ovary-specific eIF4E and 4E-T in early *Xenopus* oocytes, *J. Biol. Chem.* 282 (2007) 37389–37401, <http://dx.doi.org/10.1074/jbc.M704629200>.
- [34] G. Hernandez, H. Han, V. Gandin, et al., Eukaryotic initiation factor 4E-3 is essential for meiotic chromosome segregation, cytokinesis and male fertility in *Drosophila*, *Development* 139 (2012) 3211–3220, <http://dx.doi.org/10.1242/dev.073122>.
- [35] P.F. Cho, F. Poulin, Y.A. Cho-Park, et al., A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP, *Cell* 121 (2005) 411–423, <http://dx.doi.org/10.1016/j.cell.2005.02.024>.
- [36] P.F. Cho, C. Gamberi, Y.A. Cho-Park, et al., Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos, *Curr. Biol.* 16 (2006) 2035–2041, <http://dx.doi.org/10.1016/j.cub.2006.08.093>.
- [37] A. Yarunin, R.E. Harris, M.P. Ashe, et al., Patterning of the *Drosophila* oocyte by a sequential translation repression program involving the d4EHP and Belle translational repressors, *RNA Biol.* 8 (2011) 904–912, <http://dx.doi.org/10.4161/rna.8.5.16325>.
- [38] J.C. Villaescusa, C. Buratti, D. Penkov, et al., Cytoplasmic Prep1 interacts with 4EHP inhibiting Hoxb4 translation, *PLoS One* 4 (2009) e5213, <http://dx.doi.org/10.1371/journal.pone.0005213>.
- [39] J. Uniacke, C.E. Holterman, G. Lachance, et al., An oxygen-regulated switch in the protein synthesis machinery, *Nature* 486 (2012) 126–129, <http://dx.doi.org/10.1038/nature11055>.

Supplementary information

Five eIF4E isoforms from *Arabidopsis thaliana* are characterized by distinct features of cap analogs binding

Anna Kropiwnicka^{a,b}, Krzysztof Kuchta^{a,c}, Maciej Lukaszewicz^b, Joanna Kowalska^b, Jacek Jemielity^{b,c}, Krzysztof Ginalski^c, Edward Darzynkiewicz^{b,c}, Joanna Zuberek^{b,*}

^a College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland

^b Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland

^c Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland

* Corresponding author: e-mail: jzuberek@biogeo.uw.edu.pl, tel.: +48225540781, fax: +48225540771

S1. Supplementary Materials and Methods

S1.1. Protocol modifications for purification of AteIF4E1b, AteIF4E1c and AtnCBP

For AtnCBP nucleic acids contaminants were removed using QIAquick Spin columns (QIAGEN) before protein refolding step.

For AteIF4E1b and AteIF4E1c nucleic acids were cleaved by Benzonase nuclease (Novagen) during inclusion bodies washing and adsorbed to silica particles (50% v/v silicon dioxide in 6M guanidine hydrochloride) after solubilization of inclusion bodies. Proteins were refolded during a two-step dialysis against a buffer with a decreasing concentration of guanidine hydrochloride and KCl. During ion-exchange chromatography nucleic acids were

bound to a HiTrap Q HP column (GE Healthcare) and the protein in the flow-through was applied to and eluted from a HiTrap SP HP column (GE Healthcare).

S1.2. Synthesis of cap analogs

Cap analogs (Table S1) were synthesized as previously described [1–5] and their concentration was determined spectrophotometrically assuming appropriate molar extinction coefficients [6,7].

S2. Supplementary References

- [1] E. Darzynkiewicz, I. Ekiel, S.M. Tahara, et al., Chemical synthesis and characterization of 7-methylguanosine cap analogs, *Biochemistry*. 24 (1985) 1701–1707. doi:10.1021/bi00328a020.
- [2] J. Zuberek, J. Jemielity, A. Jablonowska, et al., Influence of electric charge variation at residues 209 and 159 on the interaction of eIF4E with the mRNA 5' terminus., *Biochemistry*. 43 (2004) 5370–9. doi:10.1021/bi030266t.
- [3] E. Darzynkiewicz, J. Stepinski, S.M. Tahara, et al., Synthesis, Conformation and Hydrolytic Stability of p 1 ,p 3 –Dinucleoside Triphosphates Related to mRNA 5'-cap, and Comparative Kinetic Studies on their Nucleoside and Nucleoside Monophosphate Analogs, *Nucleosides Nucleotides*. 9 (1990) 599–618. doi:10.1080/07328319008045191.
- [4] J. Kowalska, M. Lewdorowicz, J. Zuberek, et al., Synthesis and characterization of mRNA cap analogs containing phosphorothioate substitutions that bind tightly to eIF4E and are resistant to the decapping pyrophosphatase DcpS., *RNA*. 14 (2008) 1119–1131. doi:10.1261/rna.990208.

- [5] M. Jankowska, J. Stepinski, R. Stolarski, et al., Synthesis and properties of new NH₂ and N⁷ substituted GMP and GTP 5'-mRNA cap analogues, *Collect. Czechoslov. Chem. Commun.* 58 (1993) 138–141.
- [6] A. Cai, M. Jankowska-Anyszka, A. Centers, et al., Quantitative assessment of mRNA cap analogues as inhibitors of in vitro translation., *Biochemistry.* 38 (1999) 8538–47. doi:10.1021/bi9830213.
- [7] Z. Wieczorek, J. Stepinski, M. Jankowska, et al., Fluorescence and absorption spectroscopic properties of RNA 5'-cap analogues derived from 7-methyl-, N²,7-dimethyl- and N²,N²,7-trimethyl-guanosines, *J. Photochem. Photobiol. B Biol.* 28 (1995) 57–63. doi:10.1016/1011-1344(94)07098-9.
- [8] H. Mach, C.R. Middaugh, N. Denslow, Determining the identity and purity of recombinant proteins by UV absorption spectroscopy., *Curr. Protoc. Protein Sci.* Chapter 7 (2001) Unit 7.2. doi:10.1002/0471140864.ps0702s01.

S3. Supplementary Figures and Table Legends

Figure S1.

Amino acid sequences of *Arabidopsis thaliana* eIF4E isoforms.

Figure S2.

Purified recombinant *Arabidopsis thaliana* eIF4E isoforms are free of nucleic acids contaminations. Absorption spectra of recombinant proteins AteIF4E (A), AteIF(iso)4E (B) and AtnCBP (C) after purification on the ion-exchange column. Absorption spectra of AteIF4E1b (D) and AteIF4E1c (E) purified by the standard protocol (samples contaminated with nucleic acids) and by the protocol with added steps of nucleic acids removal (samples

free of nucleic acids). For non-contaminated samples the measured A_{260}/A_{280} ratio was near the value reported for pure protein samples ($A_{260}/A_{280} = 0.57 \pm 0.06$; [8]).

Table S1.

Chemical structures of cap analogs used for binding studies with eIF4E proteins.

Figure S1

AtelF4E

```
1 mavedtpksv vteeakpnsi enpidryhee gddaegegia ggegdnvde ssksgvpesh
61 plehswtfwf dnpavkskqt swgsslrpvf tfstveefws lynnmkhpsk lahgadfyf
121 khiiepkwed picanggkw mtftpkeksdk swlytllali geqfdhgdei cgavvnirgk
181 qerisiwtkn asneaaqvsi gkqwkefldy nnsigfiihe dakkldrak nayta
```

AtelF4E1b

```
1 mvvtdspvsg imadqnidpn tttspspkek hvsaikaisg dekapskekk nyaskksttv
61 iqkshcfqns wtfwfdnpss ksnqviwgss lrslytfgti eefwslynni hpptkwvsga
121 dlycfkdkie pkwedpican ggkwsmmfpr atlecnwlnl llalvgeqfd qgdeicgavl
181 nfrargdris lwtknaanee aqlsigkqwk ellgynetig fivhedaktl drdakrrytv
```

AtelF4E1c

```
1 mvvmdspvsg rmadqnidpn tttspspiek hvsaikaisg dekapskekk nyaskksttv
61 iqkshcfqns wtfwfdnpss ksnqviwgss lrslytfati eefwslynni hpptkwvsgs
121 dlycfkdkie pkwedpican ggkwtmffpr atlesnwlnt llalvgeqfd qgdeicgavl
181 nfrtrgdris lwtkkaanee aqlsigkqwk ellgyndtig fivhedaktl drdakrrytv
```

AtelF(iso)4E

```
1 matddvnepl paaael pate aekqphkler kwsfwfdnqs kkgaaawgasl rkaytfdtve
61 dfwglhetif qtskltanae ihlfkagvep kwedpecang gkwtwvvtan rkealdkgwl
121 etlmaligeq fdeadeicgv vasvrpqskq dklslwtrtk sneavlmgig kkweildvt
181 dkitfnnhdd srrsrftv
```

AtnCBP

```
1 mevldrrdde irdsgnm dsi kshyvtdsvs eerrsrrelkd glhplrykfs iwytrrtpgv
61 rnqsyednik kmvefstveg fwacychlar ssllpsptdl hffkdgirpl wedgancngg
121 kwiirfskv sarfwdlll alvgdqldda dnicgavlsv rfmediisv nrrnasdhqav
181 mglrdsikrh lklphayvme ykphdaslrd nssyrntwlr g
```

Figure S2

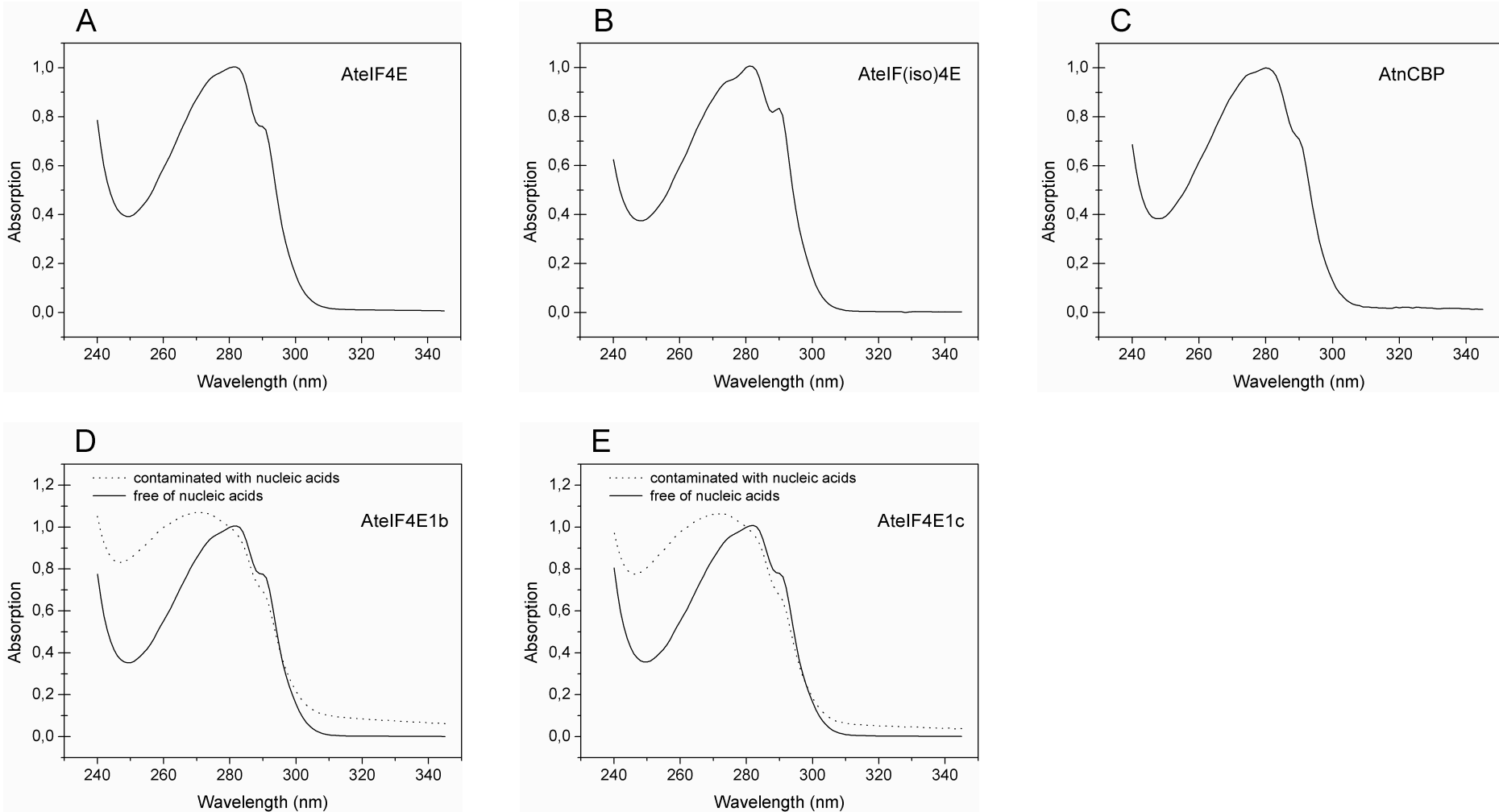
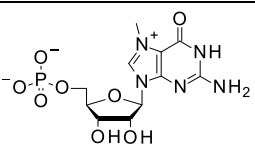
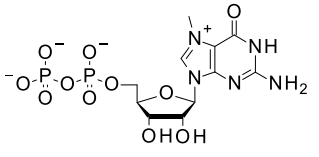
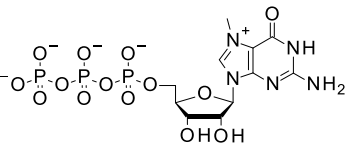
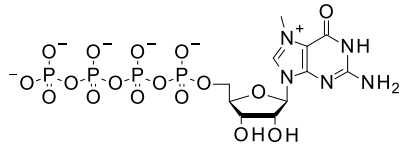
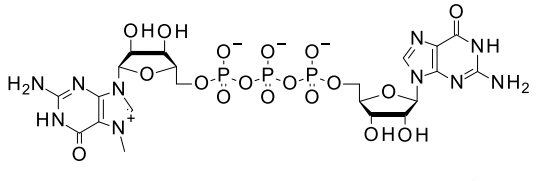
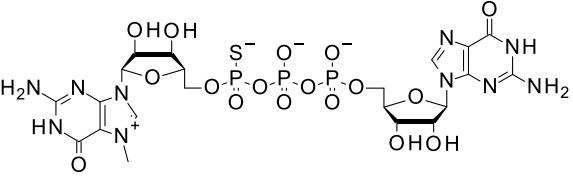
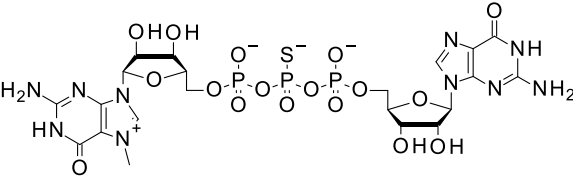
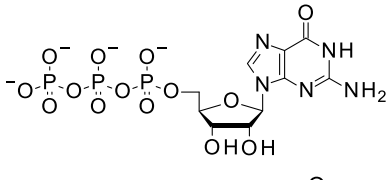
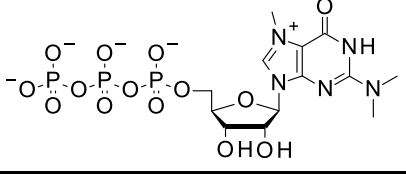


Table S1

Cap analog	Structure	Reference for the synthesis
m ⁷ GMP		[1]
m ⁷ GDP		[1]
m ⁷ GTP		[1]
m ⁷ Gp ₄		[2]
m ⁷ GpppG		[3]
m ⁷ GpsppG (D1) and (D2) (diastereomer 1 and 2)		[4]
m ⁷ GppspG (D1) and (D2) (diastereomer 1 and 2)		[4]
GTP		Commercial product (MP Biomedicals LLC)
m ₃ ^{2,2,7} GTP		[5]

Krzysztof Kuchta
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalski K,
Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are
characterized by distinct features of cap analogs binding. 2015 Biochemical
Biophysical Research Communication 456:47-52*

udział **Anny Kropiwnickiej (Biofizyka, Instytut Fizyki Doświadczalnej, Wydział
Fizyki, Uniwersytet Warszawski)** polegał na:

- uzyskaniu konstruktów ekspresyjnych dla białek AteIF4E1b i AteIF4E1c w postaci pochodnych plazmidu pET11d z wklonowanymi odpowiednimi cDNA kodującymi wyżej wymienione białka,
- ekspresji w komórkach bakteryjnych, oczyszczaniu z ciałek inkluzyjnych i uzyskaniu w formie zwiniętej białek AteIF4E, AteIF4E1b, AteIF4E1c, AteIF(iso)4E i AtnCBP,
- opracowaniu protokołu oczyszczania białek AteIF4E1a, AteIF4E1b i AtnCBP, wykorzystującego trawienie enzymatyczne kwasów nukleinowych, krzemionkę i chromatografię jonowymienną w celu usunięcia kwasów nukleinowych zanieczyszczających preparaty białka,
- wykonaniu badań oddziaływania białek AteIF4E, AteIF4E1b, AteIF4E1c, AteIF(iso)4E i AtnCBP z syntetycznymi analogami kapu techniką miareczkowań fluorescencyjnych zsynchronizowanych w czasie.

Krzysztof Kuchta.....
Podpis

20 lipiec 2017

.....
Miejscowość i data

Krzysztof Kuchta
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalski K, Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are characterized by distinct features of cap analogs binding. 2015 Biochemical Biophysical Research Communication 456:47-52

mój udział polegał na:

- Analizie sekwencyjno-strukturalnej homologów eIF4E,
- Stworzeniu modeli struktury przestrzennej izoform eIF4E z *A. thaliana*,
- Zidentyfikowaniu i opisanu w manuskrypcie różnic sekwencyjno-strukturalnych pomiędzy białkami eIF4E z *A. thaliana*, które potencjalnie tłumaczą otrzymane wyniki doświadczalne (różnice w powinowactwie wiązania izoformy eIF4E z analogiem czapeczki),
- Przygotowaniu rysunku 1.

.....
Krzysztof Kuchta
Podpis

Warszawa, 17.07.2017

Miejscowość i data

Maciej Łukaszewicz

Zakład Biofizyki, Instytut Fizyki Doświadczalnej,

Wydział Fizyki, Uniwersytet Warszawski,

Żwirki i Wigury 93,

02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kropiwnicka A, Kuchta K, Łukaszewicz M, Kowalska J, Jemielity J, Ginalski K,
Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are
characterized by distinct features of cap analogs binding. 2015 Biochemical
Biophysical Research Communication 456:47-52*

mój udział polegał na: konstrukcji wektorów plazmidowych do nadekspresji białek nCBP, eIF4E_At i eIF(iso)4E_At, bez dodatkowych sekwencji powinowactwa (tzw. affinity-tags): pET28a-nCBP_At, pET30a-4E_At, pET30a-iso4E_At.


.....
Podpis

18.07.2017

Miejscowość i data

Joanna Kowalska
Biofizyka, Instytut Fizyki Doświadczalnej,
Wydział Fizyki, Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalski K,
Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are
characterized by distinct features of cap analogs binding. 2015 Biochemical
Biophysical Research Communication 456:47-52*

mój udział polegał na syntezie analogów końca 5' mRNA (kapu) użytych w
badaniach.

Joanna Kowalska

Podpis

Warszawa 17.07.2017

Dr hab. Jacek Jemielity, Prof. UW
Centrum Nowych Technologii UW
Banacha 2c, 02-097 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy: *Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalska K, Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are characterized by distinct features of cap analogs binding. 2015 Biochemical Biophysical Research Communication 456:47-52*

mój udział polegał na syntezie analogów końca 5' mRNA

A handwritten signature in blue ink, reading "Jacek Jemielity", with a long, sweeping horizontal line extending to the right.

7 sierpień 2017

.....
Miejscowość i data

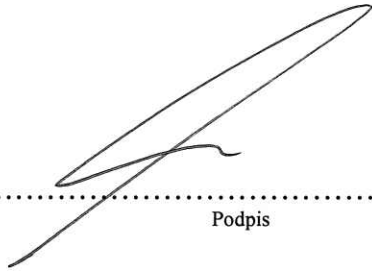
Krzysztof Ginalski
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalski K,
Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are
characterized by distinct features of cap analogs binding. 2015 Biochemical
Biophysical Research Communication 456:47-52*

mój udział polegał na zaplanowaniu badań bioinformatycznych, współuczestnictwie
w analizie otrzymanych danych doświadczalnych oraz pomocy w pisaniu
manuskryptu.

.....

Podpis

Warszawa, 7 sierpień 2017

.....
Miejscowość i data

Edward Darżynkiewicz

Biofizyka, Instytut Fizyki Doświadczalnej,
Wydział Fizyki, Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

Centrum Nowych Technologii
Banacha 2c,
02-097 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalska K,
Darżynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are
characterized by distinct features of cap analogs binding. 2015 Biochemical
Biophysical Research Communication 456:47-52*

mój udział polegał na opracowaniu syntezy związków chemicznych (analogów kapu)
użytych w badaniach dr Joanny Żuberek.


.....
Podpis

Warszawa, dn. 20 lipca 2017 r.

dr Joanna Maria Żuberek

Zakład Biofizyki, Instytut Fizyki Doświadczalnej

Wydział Fizyki

Uniwersytet Warszawski

ul. Żwirki i Wigury 93

02-089 Warszawa

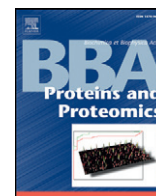
OŚWIADCZENIE

Oświadczam, że w pracy: *Kropiwnicka A, Kuchta K, Lukaszewicz M, Kowalska J, Jemielity J, Ginalska K, Darzynkiewicz E, Zuberek J. Five eIF4E isoforms from Arabidopsis thaliana are characterized by distinct features of cap analogs binding. 2015 Biochemical Biophysical Research Communication 456:47-52*

mój udział polegał na:

- na sformułowaniu problemu badawczego i stworzeniu koncepcji pracy;
- zaplanowaniu badań doświadczalnych w zakresie otrzymania rekombinowanych białek eIF4E z *Arabidopsis thaliana* oraz badań biofizycznych ich oddziaływania z analogami kapu;
- opiece merytorycznej nad metodyką oczyszczania białek i badaniami z zastosowaniem spektroskopii emisyjnej mającymi na celu scharakteryzowanie powinowactwa białek eIF4E z *A. thaliana* do analogów kapu;
- analizie i interpretacji uzyskanych równowagowych stałych asocjacji dla kompleksów izoform eIF4E z *A. thaliana* z analogami kapu;
- napisaniu pierwszej wersji manuskryptu, z wyłączeniem metodologii homologicznego modelowania białek i opisu różnic sekwencyjno-strukturalnych pomiędzy białkami eIF4E z *A. thaliana*, oraz ostatecznej redakcji manuskryptu uwzględniającej uwagi i poprawki współautorów.

J. Żuberek



Diverse cap-binding properties of *Drosophila* eIF4E isoforms

Joanna Zuberek^{a,*}, Krzysztof Kuchta^{b,c,1}, Greco Hernández^d, Nahum Sonenberg^e, Krzysztof Ginalski^b

^a Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw 02-089, Poland

^b Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, Warsaw 02-097, Poland

^c College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw 02-089, Poland

^d Division of Basic Research, National Institute of Cancer (INCan), Tlalpan, Mexico City 14080, Mexico

^e Department of Biochemistry and Goodman Cancer Research Centre, McGill University, Montreal, QC H3A 1A3, Canada

ARTICLE INFO

Article history:

Received 7 June 2016

Accepted 28 June 2016

Available online 29 June 2016

Keywords:

eIF4E family

Translation initiation

Cap analogues

Protein–ligand interactions

Fluorescence spectroscopy

ABSTRACT

The majority of eukaryotic mRNAs are translated in a cap-dependent manner, which requires recognition of the mRNA 5' cap by eIF4E protein. Multiple eIF4E family members have been identified in most eukaryotic organisms. *Drosophila melanogaster* (*Dm*) has eight eIF4E related proteins; seven of them belong to Class I and one to Class II. Their biological roles with the exception of *Dm* eIF4E-1, *Dm* eIF4E-3 and *Dm* 4EHP, remain unknown. Here, we compare the molecular basis of *Dm* eIF4E's interactions with cap and eIF4G peptide by using homology modelling and fluorescence binding assays with various cap analogues. We found that despite the presence of conserved key residues responsible for cap recognition, the differences in binding different cap analogues among Class I *Dm* eIF4E isoforms are up to 14-fold. The highest affinity for cap analogues was observed for *Dm* eIF4E-3. We suggest that *Dm* eIF4E-3 and *Dm* eIF4E-5 bind the second nucleoside of the cap in an unusual manner via stacking interactions with a histidine or a phenylalanine residue, respectively. Moreover, the analysis of ternary complexes of eIF4G peptide–eIF4E–cap analogue showed cooperativity between eIF4G and cap binding only for *Dm* eIF4E-4, which exhibits the lowest affinity for cap analogues among all *Dm* eIF4Es.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Post-transcriptional regulation of gene expression plays a key role in many cellular and developmental processes. In eukaryotes, the vast majority of mRNAs are translated via the cap-dependent mechanism, in which the cap binding protein eIF4E, complexed with a scaffold protein eIF4G, binds the cap structure (m⁷GpppN, where N is any nucleotide) at the 5' end of the mRNA to recruit the small ribosomal subunit to initiate translation. In this mechanism, cap recognition by eIF4E, as well as the formation of the eIF4E–eIF4G complex, are two of the limiting steps regulating the global process of translation [1,2]. eIF4E is negatively regulated by eIF4E-binding proteins (4E-BP), which share with eIF4G the canonical eIF4E-binding motif (C-motif) YXXXXLΦ (where X is any amino acid and Φ is a hydrophobic residue) that interacts with the convex dorsal surface of eIF4E. Binding of 4E-BPs to eIF4E precludes formation of the eIF4E–eIF4G complex, thereby repressing cap-dependent translation [3,4].

With the advent of genome-wide sequencing projects of hundreds of organisms across disparate *phyla*, a number of eIF4E genes have

been discovered and classified into three Classes [5,6]. Interestingly, some species have been reported to have several eIF4E paralogues that display different expression patterns in certain tissues, promote or inhibit translation of specific mRNAs, are involved in translation under stress responses, or show variable biochemical properties. Moreover, some eIF4E orthologues are also restricted to specific phylogenetic lineages [5–13]. Mammals have four eIF4E isoforms, eIF4E (also called eIF4E-1a), eIF4E-1b belonging to Class I, eIF4E-2 (or 4EHP) belonging to Class II and eIF4E-3 belonging to Class III. Whereas eIF4E-1a is the major translation factor, eIF4E-2 acts in a cell most probably as a translation repressor. It has been shown that mouse 4EHP binding homeodomain transcription factor Prep1 inhibits *Hoxb4* translation [14] and, along with Grb10-interacting GYF protein (GIGYF2) and zinc finger protein 598, it is a member of translational repressor complex during embryonic development [15]. However, some recent studies [16–18] have indicated an alternative role for 4EHP. They showed that under conditions of extreme oxygen depletion (hypoxia), human cells use 4EHP in the process of an alternative cap-dependent translation. The findings concerning the role of eIF4E-3, the fourth member of mammalian eIF4E family, suggest that it acts as a tissue-specific tumour suppressor [19]. The biological role of eIF4E-1b in mammalian cells remains unknown. The expression of this isoform is strictly limited to oocytes across all classes of *Tetrapoda* [9]. Its role was identified in *Xenopus* oocyte, where it is member of CPEB mRNA repressor complex [20].

* Corresponding author.

E-mail addresses: jzuberek@biogeo.uw.edu.pl (J. Zuberek), kuchtak@cent.uw.edu.pl (K. Kuchta), greco.hernandez@gmail.com (G. Hernández), nahum.sonenberg@mcgill.ca (N. Sonenberg), kginal@cent.uw.edu.pl (K. Ginalski).

¹ These authors contributed equally to the paper as first authors.

The biological role of eIF4E family members has been also extensively investigated in a nematode, *Caenorhabditis elegans*, encoding five eIF4E family members (IFE-1 to -5) [13,21], which show different expression patterns and effects of gene deletion and mutations for growth and development. The IFE-3 isoform, which is closely related to mammalian eIF4E-1a, is essential for viability and embryogenesis [13] and it regulates sex-determination in the hermaphrodite germline [22]. IFE-1, which is expressed in germ-line along with IFE-3 and IFE-5, is required for production of functional sperm, as it affects the expression of some sperm-specific proteins, such as GSP-3 and MSPs [23,24]. IFE-2, which also functions in germline, is mostly expressed in somatic tissues [13]; knock-out of the *ife-2* gene reduces somatic mRNA translation. IFE-4, the only one *C. elegans* eIF4E isoform belonging to Class II, is enriched in soma and its loss produces a pleiotropic phenotype inducing a defect in egg-laying by reducing translation IFE-4 sensitive mRNA [25].

Another species with several eIF4E genes is *Drosophila melanogaster*, a model organism that has provided extensive insight into post-transcriptional mechanisms for regulating gene expression. It contains seven genes encoding eight eIF4E paralogues, namely *Dm* eIF4E-1, -2, -3, -4, -5, -6, and -7 belonging to Class I, and *Dm* 4EHP belonging to Class II. The expression of these proteins varies throughout the development [26,27]. For some of them, a tissue-specific role has been reported. Whereas eIF4E-1 is ubiquitous and required for general translation in all tissues [5,26], eIF4E-3 is a testis-specific protein required only for spermatogenesis [28,29], and 4EHP is a translational repressor that inhibits translation of specific mRNAs during early embryogenesis [30–32]. Despite these few cases, the biological roles played by the majority of the eIF4E-related proteins, as well as their structural and biochemical properties, remain unknown.

The structures of several eIF4E family proteins from different species in the apo form, and either in binary or ternary complex with various cap analogues and a fragment of eIF4G or a 4E-BP that contain the eIF4E-binding site, have been solved [33–39]. The eIF4E-like fold is characterised by the presence of an eight-stranded antiparallel β -sheet backed by α -helices (with $\beta 1\beta 2\alpha 1\beta 3\beta 4\alpha 2\beta 5\beta 6\alpha 3\beta 7\alpha 4\beta 8$ topology, as observed in mouse eIF4E (PDB: 1EJ1) [33]), forming a structure described as a cupped hand [33]. These studies have demonstrated that eIF4G or 4E-BP proteins bind to the dorsal surface of eIF4E via a canonical motif (C-motif) and as shown recently for some 4E-BPs, by a non-canonical motif (NC-motif) to a lateral surface of eIF4E [37–39]. On the other side, they have shown that cap nucleotides interact with a narrow slot on the concave surface of eIF4E. The interatomic contacts between eIF4E proteins and cap analogues can be divided into three classes: (i) sandwiching of the 7-methyl guanine between two tryptophan residues (in Class I eIF4E isoforms) or a tryptophan and a tyrosine residues (in Class II); (ii) hydrogen bonds and van der Waals contacts with the 7-methylguanosine; and (iii) direct interactions and water-mediated contacts with the phosphate chains of a cap structure and a positively charged pocket of the cap-binding slot of eIF4E formed by the side chains of several Lys and Arg residues [12,33,35].

Although all members from each eIF4E multigenic family are structurally- and sequence-related proteins, we need to understand the features of various eIF4Es that account for their differences in activity and regulation. Here, using biophysical methods, we characterise the structural properties of all eIF4E paralogues from *D. melanogaster* (*Dm* eIF4Es) both, in a binary complex with diverse cap analogues, and in a ternary complex with cap analogues and the eIF4G C-motif.

2. Material and methods

2.1. Synthesis of cap analogues

Mono- and dinucleotide cap analogues were synthesised as described previously [40–42]. Cap analogues concentrations were determined using spectrophotometric methods [43].

2.2. Synthesis of eIF4G peptides

Peptides spanning residues 609–621 of human eIF4GI [44] (sequence: KKRYDREFLLGFQ), referred to as *Hs* eIF4G peptide; residues 618–630 of *D. melanogaster* eIF4G [45] (sequence: KKQYDREQLQLR), referred to as *Dm* eIF4G peptide; and *Drosophila* eIF4G peptide with mutations Tyr621A, Leu626A and Leu627A (sequence: KKQADREQAAQLR), referred to as *Dm* eIF4G3A peptide, were synthesized by Metabion Company (Germany) using solid-phase peptide synthesis method with Fmoc-chemistry. The peptides were purified by RP-HPLC and quantified by MALDI-TOF: (*Hs* eIF4G peptide predicted mass: 1699.99 Da, measured mass: 1700.96 Da; *Dm* eIF4G peptide predicted mass: 1718 Da, measured mass: 1719.97 Da; *Dm* eIF4G3A peptide predicted mass: 1541.74 Da, measured mass: 1542.50 Da). Peptides were dissolved in 0.1% TFA in water and their concentrations in the case of *Hs* eIF4G and *Dm* eIF4G were also determined spectrophotometrically assuming $\epsilon_{280} = 1490 \text{ cm}^{-1} \text{ M}^{-1}$.

2.3. Cloning, mutagenesis, expression and purification of recombinant proteins

cDNAs of *Drosophila* eIF4Es were PCR-amplified and subcloned into pET30a expression vector (Novagen) in *NdeI*–*BamHI* sites in the case of *Dm* eIF4E-1, *Dm* eIF4E-2, *Dm* eIF4E-4, and *Dm* 4EHP, and in *NdeI*–*EcoRI* sites in the case of *Dm* eIF4E-3, *Dm* eIF4E-5(12–232) and *Dm* eIF4E-7(9–429) without any affinity tag. Point mutations Cys217Ala and Cys228Ala in *Dm* eIF4E-4 were obtained using a PCR-based site-directed mutagenesis method.

For *Dm* eIF4E expression, constructs in pET30a were transformed into *Escherichia coli* BL21(DE3) or Rosetta 2(DE3) strains (Novagen). Bacteria were grown in LB medium to OD_{600 nm} of 1.0 and induced for 3 h at 37 °C by adding 0.5 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG). Cells were harvested, resuspended in lysis buffer (20 mM HEPES/KOH (pH 7.5), 100 mM KCl, 1 mM EDTA, 2 mM DTT, 10% glycerol) and disrupted by sonication. After lysate centrifugation (30,000g for 30 min), supernatant was removed and the pellet was washed three or two times with wash buffer (20 mM HEPES/KOH (pH 7.2), 1 M guanidine hydrochloride, 2 mM DTT, 10% glycerol). Inclusion bodies were dissolved in 50 mM HEPES/KOH (pH 7.2), 6 M guanidine hydrochloride, 10% glycerol, and 2 mM DTT, and cell debris was removed by centrifugation (43,000g for 30 min). Proteins (diluted to a concentration lower than 0.1 mg/mL) were refolded using a one-step dialysis against 50 mM HEPES/KOH (pH 7.2), 100 mM KCl (or 50 mM KCl in the case of *Dm* eIF4E-3), 1.0 mM EDTA, and 2 mM DTT, and purified by ion exchange chromatography on a HiTrap SP column (GE Healthcare). Purified proteins were analysed on Bio-Safe Coomassie stained 15% acrylamide gel (in Suppl. Fig. S1) and their concentrations were determined based on absorption of protein samples using molar absorption coefficients as follows: $\epsilon_{280} = 52,940 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* eIF4E-1; $\epsilon_{280} = 52,940 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* eIF4E-2; $\epsilon_{280} = 65,890 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* eIF4E-3; $\epsilon_{280} = 51,450 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* eIF4E-4 and mutant *Dm* eIF4E-4_C217A/C228A; $\epsilon_{280} = 55,920 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* eIF4E-5; $\epsilon_{280} = 54,430 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* eIF4E-7; and $\epsilon_{280} = 55,920 \text{ M}^{-1} \text{ cm}^{-1}$ for *Dm* 4EHP, as calculated based on amino acid composition using an algorithm on ExPASy Server [46].

Human eIF4E protein was expressed in *E. coli* BL21(DE3) strain (Novagen) and purified from inclusion bodies as described above for *Drosophila* eIF4E proteins. Concentration of the purified protein was determined based on absorption of the protein sample using $\epsilon_{280} = 52,940 \text{ M}^{-1} \text{ cm}^{-1}$ calculated based on amino acid composition [46].

2.4. Protein structure homology modelling

Homologues of *D. melanogaster* eIF4E proteins were identified with PSI-Blast [47] searches (E-value threshold of 0.005) performed against

the NCBI nonredundant protein sequence database using sequences of *Dm* eIF4E-1, *Dm* eIF4E-2, *Dm* eIF4E-3, *Dm* eIF4E-4, *Dm* eIF4E-5, *Dm* eIF4E-7 and *Dm* 4EHP as queries. Multiple sequence alignment was derived using PCMA program [48] followed by some manual adjustments. Secondary structures were predicted with PSIPRED [49]. 3D models of *D. melanogaster* eIF4E proteins with cap analogues were constructed with MODELLER [50] using selected homologues of known structure as templates [51]. To obtain eIF4E isoforms bound to m⁷GDP, *Dm* eIF4E-1/*Dm* eIF4E-2, *Dm* eIF4E-3, *Dm* eIF4E-4, *Dm* eIF4E-5, *Dm* eIF4E-7 were modelled based on *Dm* eIF4E-2 (PDB: 4AXG) [37] and human eIF4E (PDB: 2W97) [36], while *Dm* 4EHP was modelled based on human 4EHP (PDB: 2JGB) [52], *Pisum sativum* eIF4E (PDB: 2WMC) [53] and *Mus musculus* eIF4E (PDB: 1EJ1) [33]. 3D models of *Dm* eIF4E-4 and *Dm* eIF4E-5 in complex with eIF4G peptide were generated using the structure of *Dm* eIF4E-2 (PDB: 4UEC) [39]. Additionally, *Dm* eIF4E-1/*Dm* eIF4E-2, *Dm* eIF4E-3 and *Dm* eIF4E-5 were also modelled to obtain a 3D structure with m⁷GpppA using *Dm* eIF4E-2 (PDB: 4AXG) and human eIF4E (PDB: 1IPB) [54] as templates.

2.5. Fluorescence binding assays

Fluorescence measurements leading to determination of association constants for eIF4E with cap analogues and peptides were performed as described previously [34,55]. Fluorescence titration curves were carried out on an LS-55 spectrofluorometer (Perkin Elmer Co., Norwalk, CT., USA), in 50 mM HEPES/KOH (pH 7.2), 0.5 mM EDTA and 1 mM DTT, adjusted with KCl to an ionic strength of 150 mM. The sample was thermostated at a temperature of 20.0 ± 0.3 °C and controlled with a thermocouple inside the cuvette. In the case of eIF4E–cap analogue titration, the protein fluorescence was excited at 280 nm or 290 nm and the fluorescence intensity was monitored at a single wavelength 340 nm or 345 nm. For titration of eIF4E with eIF4G peptides, the protein fluorescence was excited at 290 nm and observed at 345 nm, where the relative quenching of protein fluorescence upon peptide binding was the highest. Continuous titrations were performed by adding 1 µL aliquots of a cap analogue or peptide solution to 1400 µL of 0.1 µM solution of eIF4E alone or saturated with the peptide or cap analogue. The time for the integration of fluorescence signal upon each ligand addition was 30 s. The solution was mixed using magnetic stirring. Each titration consisted of 35–45 data points. Measured fluorescence intensities were corrected for dilution (<4%) and for the inner filter effect. Association equilibrium constants (K_{as}) were obtained by fitting a theoretical curve of corrected fluorescence intensity (F) upon total concentration of ligand ($[L]$) to the titration data according to the equation:

$$F = F(0) - [cx] \cdot (\Delta\phi + \phi_{\text{lig-free}}) + [L] \cdot \phi_{\text{lig-free}}$$

with the concentration of the cap analogue–eIF4E complex $[cx]$ given by:

$$[cx] = \frac{[L] + [P_{\text{act}}]}{2} + \frac{1 - \sqrt{(K_{as}([L] - [P_{\text{act}}]) + 1)^2 + 4K_{as} \cdot [P_{\text{act}}]}}{2K_{as}}$$

where $\Delta\phi$ stands for the difference between the fluorescence efficiencies of the apo and ligand-bound protein; $\phi_{\text{lig-free}}$ is the fluorescence efficiency of the free ligand; $F(0)$ is the initial fluorescence, and $[P_{\text{act}}]$ is the total concentration of the active protein [34]. The final K_{as} was calculated as a weighted average from three to ten independent titration series carried out for at least two independent preparations of the protein. Numerical least-squares nonlinear regression analysis was performed using ORIGIN 6.0 from Microcal Software Inc., USA.

The Gibbs free energy of binding was calculated based on the K_{as} value according to the standard equation $\Delta G^\circ = -RT \ln K_{as}$.

3. Results and discussion

To compare the cap- and eIF4G-binding properties of *Drosophila* eIF4E family proteins, we applied the fluorescence titration, which is widely used to investigate the eIF4E–cap analogue association [34,55–58]. In this technique, quenching of the intrinsic fluorescence of conserved tryptophan residues in eIF4E (Fig. 1) due to binding of a cap analogue as well as eIF4G peptide is measured [34]. Next, we built homology models of *Dm* eIF4E structures and analysed the sequence and structure differences that may influence interaction with the cap and the eIF4G peptide in relation to the results obtained from binding studies.

3.1. The canonical eIF4E-1 from *Drosophila* binds cap analogues with significantly weaker affinity than its human counterpart

The measured K_{as} values for complexes of *Dm* eIF4E-1 and human eIF4E with different cap analogues are presented in Table 1. Similar to plant [58] and yeast eIF4Es [59], *Dm* eIF4E-1 bound all investigated cap analogues from 2.4- to 10-fold weaker than human eIF4E (Fig. 2A), keeping high specificity towards a methylated guanine at the N7 position and an extension of the phosphate chain (Table 2) [34,42]. For m⁷GMP, the K_{as} is $0.82 \pm 0.05 \mu\text{M}^{-1}$ for human eIF4E and $0.34 \pm 0.01 \mu\text{M}^{-1}$ for *Dm* eIF4E-1, and for m⁷GTP the K_{as} is $70.1 \pm 1.2 \mu\text{M}^{-1}$ and $8.94 \pm 0.27 \mu\text{M}^{-1}$ for human and *Drosophila* proteins, respectively.

Sequence comparison of human eIF4E and *Dm* eIF4E-1 showed that both proteins share all key residues that are required for cap binding (Fig. 1 and Table S1 for numbering). The main difference, which might affect cap binding affinity, seems to be an additional α -helix (amino acids 119–124) present in human eIF4E (Fig. 2B) and other vertebrates, but absent in *Drosophila*, yeast [60] and plant eIF4Es [61]. This region consists of large amino acids (KQRRS) and lifts the position of the neighbouring β 5 β 6 loop towards the cap binding pocket. As a consequence, in human eIF4E Lys159 located at the tip of β 5 β 6 loop interacts tightly with the cap phosphate moiety, in contrast to its counterpart Lys201 in *Dm* eIF4E-1 (Fig. 2B). Crystal and NMR structures of mammalian eIF4E showed that the α - and β -phosphate groups of cap analogues are stabilised in the binding slot by a hydrogen bond formed with Arg157 and a salt bridge with Lys162 (numbering of human eIF4E) [34,35]. Additionally, binding studies with mutated mouse eIF4E showed that lysine 159 influences the binding of the γ -, δ - and ϵ -phosphate groups of cap to eIF4E [42]. Moreover, NMR studies of human eIF4E showed that closing of the β 5 β 6 loop upon cap binding is important for interaction between the phosphate chain of the cap and the lateral chains of Arg157, Lys 159 and Lys162 [62]. Analysis of the changes in the standard Gibbs free energy of binding ($\Delta\Delta G^\circ$) (Table 2) revealed significant differences in binding of β -phosphate group of cap analogues by both eIF4E proteins. In the case of *Dm* eIF4E-1, the energy gain corresponding to addition of a β -phosphate group to m⁷GMP is ca. 0.6 kcal/mol higher than for the human counterpart, whereas the differences in the energy changes $\Delta\Delta G^\circ$ between the proteins for binding of γ - and δ -phosphate groups are only 0.1 and 0.2 kcal/mol, respectively. The observed differences in the energy changes suggest weaker closing of *Dm* eIF4E-1 structure upon cap binding and weaker interaction of Lys201 with the phosphate chain of the cap, as compared to the human counterpart.

Besides N7 guanine methylation and presence of the phosphate moiety in the cap structure, which influence its binding to canonical vertebrate eIF4E factors, the third characteristic element of the cap structure is the influence of the second nucleoside, N (m⁷GpppN). Addition of the second nucleoside to the mononucleotide in a cap analogue reduces significantly (ca. 10–12 fold) its ability to bind human eIF4E [34,42]. The same destabilisation effect caused by the presence of a second nucleoside was observed for *Dm* eIF4E-1. The K_{as} value for m⁷GpppG is ca. 7-fold lower than the K_{as} observed for m⁷GTP (Table 1).

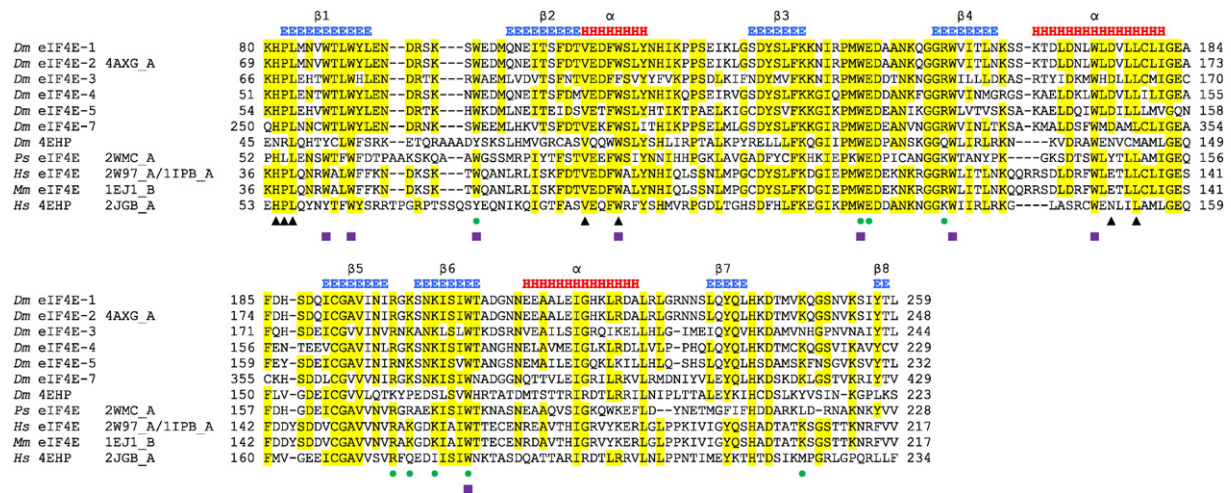


Fig. 1. Sequence comparison for eIF4E proteins. Multiple sequence alignment is shown for all *Drosophila melanogaster* eIF4E proteins and selected homologues of known structure (pea eIF4E, PDB: 2WMC; mouse eIF4E, PDB: 1EJ1; human eIF4E, PDB: 2W97/1IPB; human 4EHP, PDB: 2JGB). Conserved residues (50% or greater) are highlighted in yellow. Locations of secondary structure elements (E, β -strand; H, α -helix) in *Dm* eIF4E-2 (PDB: 4AXG) are marked above the sequences. Positions of amino acids in human/mouse eIF4E forming intermolecular contacts with cap [33,35] and eIF4G peptide [39,79] are indicated with green circles and black triangles, respectively. Positions of conserved eIF4E Trp residues are marked with violet squares.

3.2. *Dm* eIF4E-3 and *Dm* eIF4E-4 show the highest and lowest cap binding affinities, respectively, among Class I *Dm* eIF4Es

We next analysed the binding affinity of Class I *Drosophila* eIF4E proteins to different cap analogues. Despite the presence of conserved key residues responsible for cap recognition (Fig. 1, Table S1), *Drosophila* proteins bound cap analogues with diverse affinities (Fig. 3A). The K_{as} values for complexes of *Drosophila* eIF4E isoforms with a series of various cap analogues are presented in Table 1, and the corresponding standard Gibbs free energies of binding (ΔG°) are shown in Table S2.

Among the seven analysed *Dm* eIF4E proteins, the highest cap binding affinities are found for *Dm* eIF4E-3, with K_{as} values higher or comparable to those obtained for the human eIF4E, with the exception of m^7 GTP, m^7 Gp4 and m^7 GppppG, which are bound with higher affinity by the human eIF4E. *Dm* eIF4E-3 binds m^7 GTP ca. 3.5-fold weaker than human eIF4E, but 2-fold stronger than *Dm* eIF4E-1 and ca. 9-fold stronger than *Dm* eIF4E-4, which among all *Dm* eIF4E proteins shows the lowest ability to bind cap analogues. A characteristic feature of *Dm* eIF4E-3 is its high specificity towards dinucleotide cap analogues (Tables 1 and 2). Previous studies showed that in vertebrate eIF4E proteins [34,56,57], addition of a second guanosine to m^7 GTP and m^7 Gp4 gives a disadvantageous entropic effect of about +1.4 kcal/mol and +1.8 kcal/mol calculated as the change in the Gibbs free energy of binding ($\Delta\Delta G^\circ$). The observed K_{as} values for m^7 GpppG in complex with vertebrate eIF4E were even 2.5–3.0-fold lower than those observed for

eIF4E- complexes with m^7 GDP [34,56,57]. Addition of further three nucleotides to the cap structure (m^7 GpppNpNpNpN) restored the K_{as} values to the level observed for m^7 GTP [63]. These results show that the second nucleoside of the cap is not well stabilised within the cap binding slot itself and, additionally, its presence destabilises intermolecular contacts between the second, third and fourth phosphate groups of dinucleotide cap analogues and positively charged amino acids located at the entrance to the cap binding slot. Moreover, the short α -helix following the β 7 strand does not exist in human eIF4E complexes with dinucleotide cap analogues, as compared to complexes with mononucleotide cap analogues [35,36]. The presence of the long loop β 7 β 8 may make the structure of eIF4E in complex with a dinucleotide cap analogue more open as compared to that with a mononucleotide (Fig. 3B and C). For *Dm* eIF4E-3, the destabilising effect on complex formation resulting from the addition of the second guanosine to m^7 Gppp is only +0.5 kcal/mol, whereas m^7 GppppG and m^7 Gp4 are bound with similar K_{as} values of $35.7 \pm 1.1 \mu\text{M}^{-1}$ and $38.4 \pm 2.0 \mu\text{M}^{-1}$, respectively. For the other Class I *Dm* eIF4E proteins, complex destabilisation resulting from the presence of the second nucleoside was observed. The energetic cost of adding a guanosine moiety to m^7 GTP is ca. +1.0 kcal/mol (Table 2), similar to what is observed for the human eIF4E. Strong binding of the cap α -phosphate group by *Dm* eIF4E-3, with very low binding to the β -, γ - and δ -phosphate moieties, can explain the high specificity of this isoform towards dinucleotide cap analogues. The K_{as} value for complex of *Dm* eIF4E-3 with m^7 GMP is 3.6-

Table 1

Equilibrium association constants (K_{as}) for the complexes of *Drosophila* and human eIF4E proteins from Class I and Class II with a series of various cap analogues.

Cap analogue	Class I							Class II	
	<i>Hs</i> eIF4E	<i>Dm</i> eIF4E-1	<i>Dm</i> eIF4E-2	<i>Dm</i> eIF4E-3	<i>Dm</i> eIF4E-4	<i>Dm</i> eIF4E-5	<i>Dm</i> eIF4E-7	<i>Hs</i> 4EHP*	<i>Dm</i> 4EHP
	$K_{as} (\mu\text{M}^{-1})$								
m^7 GMP	0.82 ± 0.05	0.34 ± 0.01	0.30 ± 0.02	2.96 ± 0.06	0.10 ± 0.01	0.83 ± 0.03	0.44 ± 0.01	0.07 ± 0.01	0.35 ± 0.01
m^7 GDP	18.24 ± 0.24	2.70 ± 0.05	2.58 ± 0.03	11.94 ± 0.19	0.66 ± 0.02	7.13 ± 0.10	1.72 ± 0.02	0.23 ± 0.03	0.94 ± 0.03
m^7 GTP	70.1 ± 1.2	8.94 ± 0.27	8.47 ± 0.28	19.29 ± 0.52	2.16 ± 0.07	17.9 ± 1.4	5.3 ± 0.1	0.70 ± 0.04	2.89 ± 0.12
m^7 Gp4	320 ± 18	29.9 ± 0.9	28.4 ± 1.6	38.4 ± 2.0	7.55 ± 0.25	106.3 ± 2.6	30.2 ± 0.8	1.20 ± 0.04	4.4 ± 0.3
m^7 GpppG	6.25 ± 0.07	1.29 ± 0.04	1.23 ± 0.04	8.24 ± 0.28	0.38 ± 0.01	4.44 ± 0.20	1.56 ± 0.01	0.17 ± 0.01	0.84 ± 0.03
$m^7_{3-2'}\text{-GppppG}$	5.95 ± 0.13	1.27 ± 0.01	1.04 ± 0.03	8.96 ± 0.56	0.35 ± 0.01	5.46 ± 0.10	1.17 ± 0.01	0.12 ± 0.01	0.89 ± 0.01
m^7 GpppA	3.08 ± 0.03	0.78 ± 0.07	0.73 ± 0.03	5.88 ± 0.10	0.24 ± 0.01	3.16 ± 0.09	0.64 ± 0.03	0.13 ± 0.01	0.71 ± 0.01
m^7 GpppC	3.45 ± 0.18	0.98 ± 0.02	0.94 ± 0.01	4.79 ± 0.35	0.26 ± 0.01	2.46 ± 0.08	0.92 ± 0.03	0.16 ± 0.02	0.73 ± 0.04
m^7 GppppG	49.9 ± 1.2	7.03 ± 0.10	6.61 ± 0.15	35.7 ± 1.1	1.73 ± 0.03	22.5 ± 3.0	5.5 ± 0.2	0.72 ± 0.02	2.49 ± 0.17
GTP	0.044 ± 0.002	0.036 ± 0.005	0.05 ± 0.01	0.034 ± 0.001	0.07 ± 0.01	0.056 ± 0.003	0.07 ± 0.01	0.031 ± 0.002	0.014 ± 0.002

* data from [56].

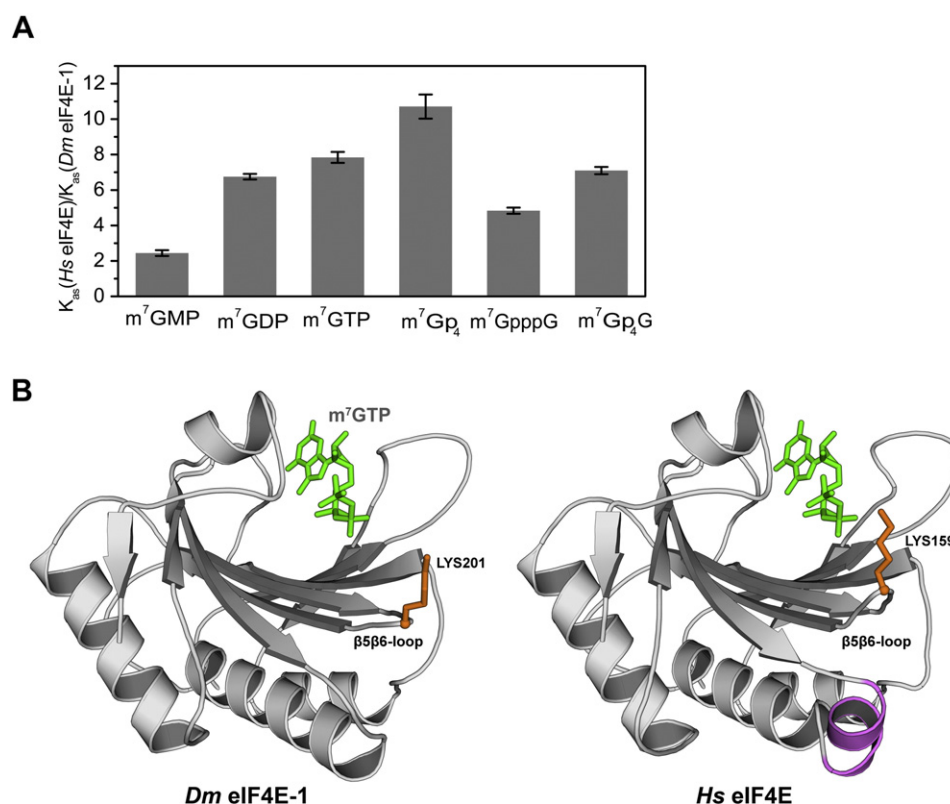


Fig. 2. Cap binding ability of *Drosophila* eIF4E-1 protein and human eIF4E. (A) Comparison of binding affinities of *Dm* eIF4E-1 and human eIF4E protein to various cap analogues on the basis of the ratio of K_{as} values as determined from fluorescence titration measurements in the same conditions. (B) Comparison of a 3D model for *Dm* eIF4E-1 and crystal structure of human eIF4E (PDB: 2W97). Additional short α -helix in human eIF4E is shown in violet, while side chains of the conserved lysine located on $\beta 5\beta 6$ loop, which is closer to the cap binding slot in human protein, are shown in orange. Cap analogue (m^7GTP) is shown in green.

fold higher than that for the human eIF4E ($K_{as} = 2.96 \pm 0.06 \mu M^{-1}$ and $0.82 \pm 0.05 \mu M^{-1}$, respectively). The energetic gain resulting from the extension of m^7GMP to m^7GDP is about -0.8 kcal/mol for *Dm* eIF4E-3, whereas for the human protein it is about -1.81 kcal/mol. Upon addition of the third and fourth phosphate groups to cap analogues, the energetic gain for *Dm* eIF4E-3 is about -0.3 kcal/mol and -0.4 kcal/mol, whereas for the human eIF4E it is about -0.8 kcal/mol and -0.9 kcal/mol, respectively (Table 2). Consequently, the second nucleotide of the cap does not disturb the interaction between the cap phosphate chain and positively charged amino acids in *Dm* eIF4E-3, in contrast to other eIF4Es, which bind β - γ - and δ -phosphate groups strongly. Additionally, according to our model, His234 in *Dm* eIF4E-3, located in the $\beta 7\beta 8$ loop (Fig. 3B), is likely to cause the observed high affinity of *Dm* eIF4E-3 for dinucleotide cap analogues. This aromatic amino acid probably takes part in the stacking interactions with the

second base of the cap and is replaced by Ser207, Gln249, Gln238, Gln219 and Leu419 in human eIF4E, *Dm* eIF4E-1, *Dm* eIF4E-2, *Dm* eIF4E-4 and *Dm* eIF4E-7, respectively (Fig. 1 and Fig. 3B). Prediction of stacking interactions between His234 of *Dm* eIF4E-3 and the second base of the cap is also confirmed by weaker binding to dinucleotide cap analogues containing pyrimidines instead of purines (Table 1). *Dm* eIF4E-5 has an aromatic residue, Phe222, at the position corresponding to His234 in *Dm* eIF4E-3. The K_{as} values for complexes of *Dm* eIF4E-5 with dinucleotide cap analogues (m^7GpppN) are ca. 3-fold higher than those observed for *Dm* eIF4E-1 and comparable to those of the human eIF4E. However, the energetic cost resulting from the addition of a second nucleoside to m^7Gppp is about $+1.0$ kcal/mol, 2-fold higher than that observed for *Dm* eIF4E-3. *Dm* eIF4E-5 shows stronger binding to the β -, γ - and δ -phosphate groups of mononucleotide cap analogues (Table 2), and probably opening of the eIF4E structure after dinucleotide

Table 2
Changes in standard Gibbs free energy ($\Delta\Delta G^\circ$), showing the contribution of structural elements of the 5' mRNA cap to binding free energy of eIF4E–cap complexes. $\Delta\Delta G^\circ$ values were calculated as the difference in the binding free energy (ΔG°) of eIF4E to cap analogues, which varies in single structural modifications.

Structural changes	Class I							Class II	
	<i>Hs</i> eIF4E	<i>Dm</i> eIF4E-1	<i>Dm</i> eIF4E-2	<i>Dm</i> eIF4E-3	<i>Dm</i> eIF4E-4	<i>Dm</i> eIF4E-5	<i>Dm</i> eIF4E-7	<i>Hs</i> 4EHP	<i>Dm</i> 4EHP
$\Delta\Delta G^\circ$ (kcal/mol)									
<i>Methylation of the guanosine ring in N⁷ position</i> $\Delta\Delta G^\circ = \Delta G^\circ (m^7Gp_n) - \Delta G^\circ (Gp_n)$									
GTP $\rightarrow$ m^7GTP	-4.29 ± 0.03	-3.20 ± 0.09	-2.98 ± 0.12	-3.69 ± 0.03	-2.02 ± 0.04	-3.36 ± 0.05	-2.49 ± 0.04	-1.80 ± 0.05	-3.07 ± 0.09
<i>Successive addition of the phosphate groups</i> $\Delta\Delta G^\circ = \Delta G^\circ (m^7Gp_{n+1}) - \Delta G^\circ (m^7Gp_n)$									
$m^7GMP \rightarrow m^7GDP$	-1.81 ± 0.03	-1.21 ± 0.03	-1.25 ± 0.04	-0.81 ± 0.02	-1.08 ± 0.06	-1.24 ± 0.02	-0.80 ± 0.02	-0.71 ± 0.15	-0.58 ± 0.02
$m^7GDP \rightarrow m^7GTP$	-0.78 ± 0.02	-0.70 ± 0.02	-0.69 ± 0.02	-0.28 ± 0.02	-0.69 ± 0.03	-0.54 ± 0.05	-0.65 ± 0.01	-0.64 ± 0.09	-0.66 ± 0.03
$m^7GTP \rightarrow m^7Gp_4$	-0.88 ± 0.04	-0.70 ± 0.02	-0.70 ± 0.04	-0.40 ± 0.03	-0.73 ± 0.03	-1.04 ± 0.05	-1.02 ± 0.02	-0.31 ± 0.03	-0.25 ± 0.05
<i>Addition of the second nucleoside</i> $\Delta\Delta G^\circ = \Delta G^\circ (m^7Gp_nG) - \Delta G^\circ (m^7Gp_n)$									
$m^7GTP \rightarrow m^7Gp_3G$	1.41 ± 0.02	1.13 ± 0.03	1.13 ± 0.03	0.50 ± 0.03	1.01 ± 0.02	0.981 ± 0.05	0.71 ± 0.01	0.82 ± 0.04	0.72 ± 0.03
$m^7Gp_4 \rightarrow m^7Gp_4G$	1.08 ± 0.04	0.84 ± 0.02	0.85 ± 0.03	0.04 ± 0.03	0.86 ± 0.02	0.90 ± 0.08	0.99 ± 0.03	0.29 ± 0.02	0.34 ± 0.06

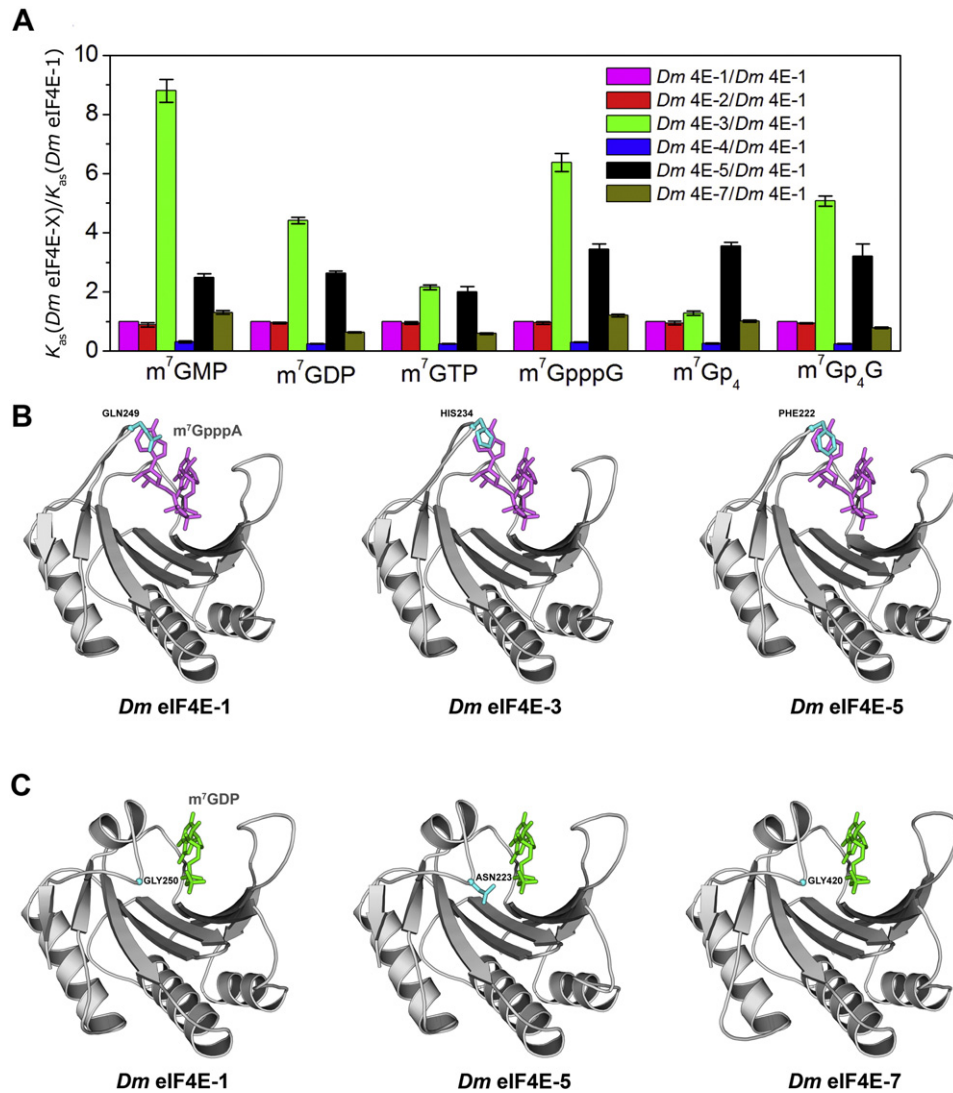


Fig. 3. Differences in the cap binding ability of Class I *Drosophila* eIF4Es. (A) Comparison of the binding affinities of *Dm* eIF4Es to various cap analogues determined from fluorescence titration measurements in the same conditions. Values are given as the K_{as} ratio of the given isoform over that of *Dm* eIF4E-1. (B) Possible stabilisation of the second base of the cap by stacking interactions with His243 in *Dm* eIF4E-3 and Phe222 in *Dm* eIF4E-5. Cap analogue (m⁷GpppA) is shown in violet. (C) Additional potential stabilisation of the phosphate group of cap by interaction with Asn223 in *Dm* eIF4E-5. Cap analogue (m⁷GDP) is shown in green.

cap analogue binding is not compensated by stacking interactions between the second base of the cap and Phe222.

Dm eIF4E-5 shows the highest affinities for dinucleotide cap analogues with methyl at the 2'-hydroxyl group of the ribose moiety of m⁷Guo. The K_{as} value for m⁷2'-O-GpppG is 1.4 fold higher than for m⁷GpppG. For the other *Dm* eIF4E isoforms and the human eIF4E, the positive influence of ribose methylation in C2' position on cap-binding was not observed (Table 1). The structures of mammalian eIF4Es show that the ribose hydroxyl group of m⁷Guo is located outside the protein surface and does not interact directly with eIF4E [33, 35].

eIF4E-7 is the *Drosophila* isoform with the highest molecular weight and the longest N-terminal region (31–249), which seems to be unstructured, as indicated by our sequence and structure analysis. We observed that eIF4E-7 binds cap analogues with affinities comparable to those of *Dm* eIF4E-1. Similarly, *Dm* eIF4E-2, obtained via alternative splicing of the product of the gene encoding *Dm* eIF4E-1, showed binding affinity for cap analogues similar to that of *Dm* eIF4E-1 (Fig. 3A). These two isoforms differ only in the N-terminal region, as the core and the C-terminal region are identical.

The lowest cap-binding affinity among the *Dm* eIF4E isoforms was observed for *Dm* eIF4E-4, despite the conservation of all key residues involved in cap binding (Fig. 1). This isoform shows a similar decrease of K_{as} values for all investigated monomethylated cap analogues, as compared to K_{as} observed for *Dm* eIF4E-1 (Table 1). *Dm* eIF4E-4 binds to monomethylated cap analogues 3.2- to 4.1-fold weaker than *Dm* eIF4E-1. The energetic effects corresponding to the extension of the cap phosphate chain and to the addition of the second nucleoside are similar for both proteins (Table 2). The main difference observed for the *Dm* eIF4E-4 isoform was its lower specificity to N7 guanosine methylation. The energetic gain observed upon GTP methylation was only -2.02 ± 0.04 kcal/mol, whereas for *Dm* eIF4E-1 it was -3.20 ± 0.09 kcal/mol and for *Dm* eIF4E-3 -3.69 ± 0.03 kcal/mol (Table 2). For this isoform, we observed the tendency for dimerisation (in Suppl. Fig. S2 A); it is likely that two cysteines in *Dm* eIF4E-4 (Cys217 and Cys228) are responsible for this effect. The cysteines are located in the β 7/ β 8 loop and the last strand β 8, which may form cross disulphide bonds. Replacement of these cysteines by alanine residues abrogates the protein dimerisation (in Suppl. Fig. S2 A) but does not affect its cap and eIF4G binding affinities (in Suppl. Fig. S2B).

3.3. Different contribution of cap's phosphate groups on stabilization of *Dm* Class I eIF4E–cap complexes

The interactions between the phosphate moiety of the cap and the side chains of basic residues located at the entrance to the eIF4E cap binding slot play a critical role in eIF4E–cap complex formation. These interactions are mediated by networks of salt bridges and hydrogen bonds [33, 35]. The extension of the phosphate chain of mono- and dinucleotide cap analogues results in an increase of vertebrate eIF4E binding affinities [34,42,56,57]. In mammalian eIF4Es, the strongest influence on cap binding is exerted by the cap α - and β -phosphate groups that interact with the side chains of Arg112, Arg157, and Lys162 (numbering of human eIF4E). The total energy of α -phosphate and β -phosphate stabilisation is ca. -3.0 kcal/mol and ca. -1.8 kcal/mol, respectively (Table 2) [34,35]. The extension of the phosphate chain by the γ - and δ -moieties gives an energetic gain of -0.8 kcal/mol and -0.9 kcal/mol, respectively, similar to what has been described for hydrogen bond formation [64] (Table 2). For all *Dm* eIF4E isoforms, an increase of K_{as} values as a result of the phosphate chain extension was also observed. However, the stabilising effect of individual phosphate groups on binding with proteins was very diverse (Tables 1 and 2).

Dm eIF4E-3 recognizes the α -phosphate group with the highest affinity, with K_{as} value for m^7 GMP about 9-fold higher than *Dm* eIF4E-1 and about 3.6-fold higher than human eIF4E and *Dm* eIF4E-5. Addition of the β -phosphate (m^7 GMP $\rightarrow$ m^7 GDP) causes a significant increase of the binding affinity in the case of *Dm* eIF4E-1, *Dm* eIF4E-2, *Dm* eIF4E-4, and *Dm* eIF4E-5. K_{as} values for the respective complexes with m^7 GDP are from 6.6 to 8.6-fold higher than with m^7 GMP. However, the observed increase in cap binding affinity for these proteins is two-fold lower, as compared to the human protein. The energetic gain of introducing the β -phosphate group into the cap is about -1.2 kcal/mol, which suggests that these isoforms can form one or two hydrogen bonds/salt bridges with the β -phosphate moiety. These results are in contrast to what we observed for *Dm* eIF4E-3 and *Dm* eIF4E-7, with a $\Delta\Delta G^\circ$ of -0.8 kcal/mol, corresponding to the creation of only one hydrogen bond/salt bridge. The changes in the binding energy accompanying the cap phosphate chain extension in mononucleotide analogues (m^7 GDP $\rightarrow$ m^7 GTP $\rightarrow$ m^7 Gp₄) is ca. -0.70 kcal/mol for *Dm* eIF4E-1, *Dm* eIF4E-2 and *Dm* eIF4E-4, implying that these proteins probably form one hydrogen bond with the cap γ - and δ -phosphate groups, similar to what has been reported for mammalian eIF4Es (Table 2) [34,42]. In the case of *Dm* eIF4E-3, the corresponding $\Delta\Delta G^\circ$ value associated with the addition of γ - and δ -phosphate groups to the cap is ca. -0.28 kcal/mol and -0.4 kcal/mol, respectively, which is weaker than the typical values reported for the hydrogen bond in a solvent-accessible region within proteins [64]. The reason of this might be a partial dispersion of the positive charged in the cap binding pocket.

In contrast to other *Dm* eIF4E isoforms and human eIF4E, a significant increase in cap binding affinity associated with the addition of an extra δ -phosphate group was observed for *Dm* eIF4E-5 and *Dm* eIF4E-7. They bind m^7 Gp₄ about 6-fold stronger than m^7 GTP, while other *Drosophila* isoforms bind m^7 Gp₄ up to 3.5-fold stronger than m^7 GTP. In addition, *Dm* eIF4E-5 binds m^7 Gp₄ 2.5-fold stronger than other isoforms. The main difference between *Dm* eIF4E-5 and the other isoforms, responsible for a higher affinity to m^7 Gp₄ might be Asn223, which may interact with the δ -phosphate group (Fig. 3C). This residue is replaced by Gly in *Dm* eIF4E-1, -2, -3, -4 and -7. Based on our modelling, we see no significant changes in *Dm* eIF4E-7 that might be responsible for its higher affinity to the m^7 Gp₄ cap analogue.

3.4. *Drosophila* 4EHP shows a low cap binding affinity but still higher than its human counterpart

Dm 4EHP belongs to the Class II of eIF4E family proteins [6]. Members of this class have either Tyr or Phe in the position corresponding

to Trp43 and Tyr instead of Trp56 of human eIF4E [6]. As a consequence, in *Dm* 4EHP and human 4EHP the N7-methylguanine moiety of the cap is sandwiched between tyrosine and tryptophan residues (Tyr68 and Trp114 in *Dm* 4EHP and Tyr78 and Trp124 in human 4EHP [52]) (Fig. 4). In contrast, in Class I eIF4E proteins, the N7-methylguanine is stacked between two tryptophan residues (Trp56 and Trp102 in the human protein). The lack of the ability to interact with eIF4G renders both, *Drosophila* and human 4EHP, translation repressors [26,30]. Through binding with both, the cap and diverse 4E-BPs that directly or indirectly bind to 3'-UTR elements, 4EHP inhibits translation of specific mRNAs [14,15,30,31]. Recent fluorescence binding studies showed that human 4EHP recognises cap analogues very weakly, with K_{as} for m^7 GTP 100-fold lower than that of the human eIF4E [56]. This difference in cap affinity probably prevents uncontrolled competition between 4EHP and eIF4E for cap binding in vivo.

We compared the binding affinities of *Dm* 4EHP for different cap analogues. As reported for human 4EHP, also for *Dm* 4EHP we observed lower binding affinity for cap analogues as compared to *Dm* eIF4E-1, but the differences in ability to bind the cap are not as high as it was observed for human proteins (Table 1). *Dm* 4EHP binds m^7 GTP with a K_{as} value of $2.89 \pm 0.12 \mu\text{M}^{-1}$, which is 3.1-fold lower than the value observed for *Dm* eIF4E-1 and similar to the *Dm* eIF4E-4 value. Thus, *Dm* 4EHP affinity for m^7 GTP is 4-fold higher than the K_{as} value reported for its human counterpart under the same conditions [56].

Since *Dm* 4EHP specificity towards N7-methylguanine of cap is similar to that of *Dm* eIF4E-1 (Table 2), analyses of our data suggest that the weak affinity of *Dm* 4EHP for the cap might not be caused by the Trp to Tyr substitution in the stacking interaction with m^7 G. We observed that the corresponding energetic gain upon addition of a methyl group at the guanine moiety of GTP is -3.07 ± 0.09 kcal/mol and -3.20 ± 0.09 kcal/mol for *Dm* 4EHP and *Dm* eIF4E-1, respectively. In addition, substitution of Trp56 by tyrosine residue in human eIF4E did not reduce its binding affinity [56]. Both, our sequence analysis and 3D structure modelling, suggest that the most prominent difference that might explain the weak cap binding ability of *Dm* 4EHP seems to be the lack of conserved Arg and Lys residues (Arg156, Lys201 and Lys204 in *Dm* eIF4E-1), which may interact with the phosphate groups of the cap. In *Dm* 4EHP, they are replaced by glutamine (Gln124), proline (Pro166) and serine (Ser169), respectively (Fig. 4). Lack of arginine and lysine residues in these positions can be partially compensated in *Dm* 4EHP by Arg128. This residue is substituted by an uncharged amino acid in the other *Drosophila* eIF4E isoforms (e.g. Thr160 in *Dm* eIF4E-1).

The binding studies of *Drosophila* and human 4EHP revealed some differences and similarities in their abilities to bind the cap. First of all, human 4EHP shows a very low specificity to N7 guanosine methylation, what distinguishes it from *Dm* 4EHP. The changes in binding energy corresponding to methylation of the guanine moiety in GTP is only -1.80 ± 0.05 kcal/mol for human 4EHP, whereas for *Dm* 4EHP it is -3.07 ± 0.09 kcal/mol. These differences might be associated with the extension of the β 1 β 2 loop by five amino acids in human 4EHP, as compared to eIF4E, which may increase the loop's flexibility. In *Dm* 4EHP, this loop is extended only by three amino acids (Figs. 1 and 4). As a consequence, the aromatic ring of Tyr78 in human 4EHP might adopt a slightly different alignment and may overlap with the N7-methylguanine moiety. Again, both proteins are lacking the full network of interactions created between the conserved Arg and Lys residues and the phosphate groups of the cap that are present in canonical eIF4Es. In both proteins, Lys159, Lys162 (numbering of human eIF4E) and in *Dm* 4EHP also Arg112 are substituted by non-positively charged amino acids (Fig. 4). However, in human 4EHP, the α -phosphate group is stabilised by an extra hydrogen bond with N^ε2 of His110 [52], which is replaced by Tyr100 in *Dm* 4EHP and by glycine residue in both, *Dm* eIF4E-1 and human eIF4E. What is more, the γ -phosphate group is stabilized by water-mediated hydrogen bond with the additional Arg138 in human and Arg128 in *Drosophila* 4EHP (Fig. 4). The differences in intermolecular contacts between phosphate chain of cap and

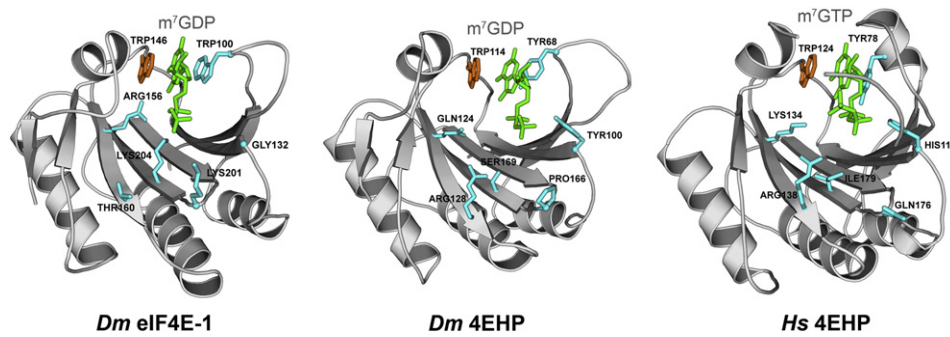


Fig. 4. Comparison of Class I and Class II eIF4E structures in complex with cap analogues. *Dm* eIF4E-1 (Class I) and *Dm* 4EHP (Class II) modelled structures were compared with reported structure of human 4EHP (Class II, PDB: 2JGB). Amino acids critical for differences in binding affinity for cap analogues are indicated in cyan. Conserved tryptophan is shown in orange. Cap analogues are depicted in green.

amino acids observed between both 4EHP proteins did not transfer to differences in the Gibbs free energy of phosphate group binding between proteins. The sequential additions of the next three phosphate groups to m⁷GMP gave the same energetic gain ($\Delta\Delta G^\circ$) for both proteins (Table 2). These results suggest that the difference in cap binding abilities between *Drosophila* and human 4EHP proteins concerns the binding of N⁷-methylguanosine.

3.5. Cooperativity between binding of the eIF4G peptide and cap analogues is observed only for *Dm* eIF4E-4 isoforms

It has been shown by both coimmunoprecipitation experiments and the yeast two-hybrid systems that all *Drosophila* Class I eIF4E proteins bind *Dm* eIF4G, in contrast to both *Drosophila* and mammalian 4EHP [7,26,28,65,66]. Moreover, recent studies [38,39] have shown that *Dm* eIF4E-2 binds *Dm* eIF4G only via the canonical 4E-binding motif 621YDREQLL627 in eIF4EG. In six Class I *Drosophila* eIF4E proteins, all residues involved in eIF4G binding by the C-motif are conserved (Fig. 1), whereas in *Dm* eIF4E-3 and *Dm* 4EHP they are not [26].

Several biophysical and biological studies have shown that the association of eIF4G to eIF4E in the *apo* form enhances cap binding by eIF4E. This cooperativity effect has been observed in the case of both, human and yeast, eIF4Es, where interaction with eIF4G increased cap affinity 2-fold [67] and up to 100-fold, respectively [68]. It has been suggested that this effect is important for formation of an active eIF4F complex and therefore for translation control [69]. In disagreement with these observations, other biophysical studies have shown that binding of a 90-residue-long fragment of eIF4G to mammalian eIF4E did not increase the cap binding affinity of eIF4E [34,70].

Next, we wanted to extend our fluorescence-based comparative analyses of cap binding to *Drosophila* eIF4E proteins in complex with a peptide (13 amino acids) consisting of the eIF4E-binding site of either

human or *Drosophila* eIF4G. As a control, we used a *Drosophila* eIF4G peptide, in which the residues important for eIF4E binding, Tyr621, Leu626 and Leu627, are substituted by Ala.

First, we determined the K_{as} values for binary complexes of *Dm* eIF4E with *Dm* eIF4G peptide, taking advantage of the fact that binding to eIF4E dorsal surface leads to quenching of the intrinsic eIF4E fluorescence through interaction with a tryptophan residue corresponding to human Trp73. The obtained K_{as} values for *Drosophila* (except for *Dm* eIF4E-3, which does not have a Trp residue in this position) and human eIF4Es are presented in Table 3. We observed K_{as} of the order of 10^6 M⁻¹, which is similar to the values published for mammalian proteins with short eIF4G peptides containing only the C-motif [34,67,71,72]. As expected, *Dm* 4EHP did not interact with *Dm* eIF4G peptide [65]. We also observed no interaction between the peptide *Dm* eIF4G3A and any of all *Dm* eIF4E isoforms. Unexpectedly, we observed the strongest binding affinity between *Dm* eIF4G peptide and *Dm* eIF4E-4, with K_{as} value for the binary complex ca. 7-fold higher than that for the corresponding complex with *Dm* eIF4E-1 (Table 3). We also observed formation of a strong complex in the case of *Dm* eIF4E-5 isoform with K_{as} value ca. 3.5-fold higher than for *Dm* eIF4E-1 and *Dm* eIF4E-2. The most crucial difference that might explain higher affinity of *Dm* eIF4E-4 and *Dm* eIF4E-5 to eIF4G peptides corresponds to the replacement of *Dm* eIF4E-2 Met73 by Glu55 and Glu58 in *Dm* eIF4E-4 and *Dm* eIF4E-5, respectively (Fig. 5C). These glutamic acids may form additional interactions with two conserved lysines, namely Lys618 and Lys619 in *Dm* eIF4G, and Lys609 and Lys610 in human eIF4G. Interestingly, both, *Dm* eIF4E-4 and human eIF4E, showed the highest affinity for their own eIF4G peptide, even though the peptides differ from each other only in 5 amino acids. The differences in binding affinities between human eIF4G and *Dm* eIF4G to the corresponding eIF4Es are by 2-fold (Table 3). These results suggest that the composition of amino acids in the C-binding motif and in the neighbouring regions, especially the presence of arginine and lysine

Table 3

Equilibrium association constants (K_{as}) of the binary complexes of the *Drosophila* and human eIF4E protein from Class I with the *Drosophila* and human eIF4G peptides that contain the respective eIF4E-binding site, with the non-binding counterpart in the case of the *Drosophila* eIF4G peptide, and for the ternary complexes: peptide-eIF4E-m⁷GTP of previously peptide-saturated eIF4E.

	Class I							Class II
	Hs eIF4E	<i>Dm</i> eIF4E-1	<i>Dm</i> eIF4E-2	<i>Dm</i> eIF4E-3 ^a	<i>Dm</i> eIF4E-4	<i>Dm</i> eIF4E-5	<i>Dm</i> eIF4E-7	<i>Dm</i> 4EHP
	K_{as} [μM^{-1}] for peptide							
<i>Dm</i> eIF4G peptide	4.44 ± 0.62	0.99 ± 0.21	1.16 ± 0.08	ND	7.39 ± 0.60	3.44 ± 0.78	1.15 ± 0.52	<0.07
<i>Dm</i> eIF4G3A peptide	<0.17	No binding	<0.03	ND	<0.03	No binding	<0.08	No binding
Hs eIF4G peptide	7.08 ± 0.30	0.74 ± 0.23	1.29 ± 0.06	ND	4.64 ± 0.14	2.33 ± 0.33	ND	ND
	K_{as} [μM^{-1}] for m ⁷ GTP when eIF4E is peptide-saturated							
Apo	73.5 ± 2.4	8.09 ± 0.25	7.34 ± 0.12	21.97 ± 0.72	2.12 ± 0.06	19.7 ± 0.8	5.70 ± 0.13	2.76 ± 0.10
<i>Dm</i> eIF4G peptide saturated	70.3 ± 3.1	7.71 ± 0.20	6.06 ± 0.08	20.86 ± 0.76	3.70 ± 0.05	17.3 ± 3.0	5.52 ± 0.07	2.71 ± 0.08
<i>Dm</i> eIF4G3A peptide saturated	73.4 ± 3.2	8.30 ± 0.45	7.04 ± 0.09	20.2 ± 2.4	2.24 ± 0.08	18.0 ± 0.7	5.56 ± 0.07	2.96 ± 0.05
Hs eIF4G peptide saturated	68.4 ± 5.5	6.77 ± 0.15	7.00 ± 0.67	ND	4.00 ± 0.06	21.4 ± 0.8	ND	ND

^a Substitution of Trp residue taking a part in binding of eIF4G and others 4E-BP proteins by Phe residue.

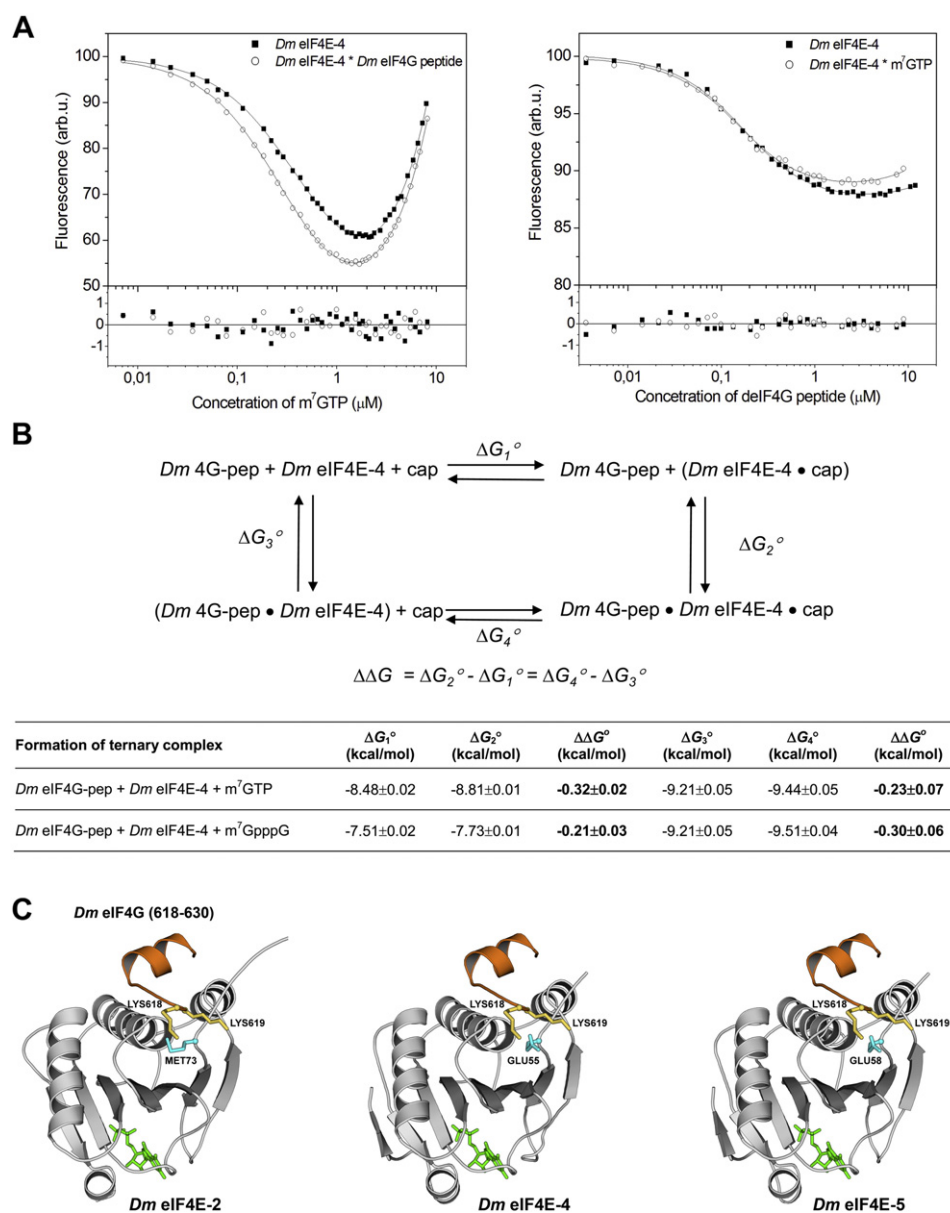


Fig. 5. Cooperativity in the formation of the ternary complex eIF4G–*Dm* eIF4E-4–cap analogue. (A) Comparison of the fluorescence titration curves for the formation of *Dm* eIF4E-4 binary and ternary complexes with m⁷GTP and the *Dm* eIF4G-peptide. (B) Thermodynamic cycle for ternary complex formation of *Dm* eIF4G-peptide, *Dm* eIF4E-4 and either m⁷GTP or m⁷GpppG cap analogues with the changes in the standard free energy ($\Delta\Delta G^\circ$) as calculated from free energy of binding (ΔG°) determined for sequential binding of each member of the complex. (C) Comparison of the reported structure of *Dm* eIF4E-2 (PDB: 4UEC) and the modelled structures of *Dm* eIF4E-4 and *Dm* eIF4E-5 in complex with the *Dm* eIF4G peptide and cap analogue. *Dm* eIF4G peptide and cap analogue are indicated in orange and green, respectively. Critical *Dm* eIF4G lysines are shown in yellow. Amino acids critical for the difference in the binding affinity for *Dm* eIF4G are depicted in cyan.

residues (KKXYXR/KX₂LΦX₂R/K), play an important role in diversification of eIF4G and other 4E-BPs binding affinities among eIF4Es, as also suggested Peter and co-workers [39].

Then, we asked whether eIF4G binding to *Drosophila* eIF4E proteins exerts a cooperative effect on their ability to bind cap analogues. To address this question, we determined K_{as} values for m⁷GTP binding, when the eIF4G peptide was already bound to *Dm* eIF4Es. The saturation concentration of eIF4G peptide for each eIF4E isoform was calculated based on the equation for concentration of eIF4E–eIF4G peptide complex using the determined K_{as} values. A short (13 amino acids) eIF4G peptide enhanced binding to m⁷GTP only in the case of *Dm* eIF4E-4, by 2-fold (Table 3). To probe the synergy in formation of the ternary complex of eIF4G peptide–eIF4E-4–cap analogues, we conducted fluorescence binding titration for all four binding reactions involving *Dm* eIF4G peptides, *Dm* eIF4E-4 and one of two cap analogues, namely, m⁷GTP and

m⁷GpppG (Fig. 5A). The obtained results are summarised in Fig. 5A, B and in Suppl. Table S3. In the case of *Dm* eIF4E-4, we observed a cooperative activity in ternary complexes formation. Both, eIF4G peptides and cap analogues, increased *Dm* eIF4E-4 affinity for the other ligand by an average of 0.2–0.3 kcal/mol in the case of *Dm* eIF4G peptide and 0.3–0.4 kcal/mol in the case of human eIF4G (1.5–2.0-fold increase of K_{as} values). Equivalent experiments for human complexes did not show the cooperativity in formation of ternary complex (in Suppl. Table S3). Previous studies on human complex formation showed both eIF4E cap-affinity enhancement by eIF4G binding [67], as well as lack of influence of eIF4G on eIF4E cap binding [34,70]. However, the eIF4E–eIF4G binding studies for human protein were performed with different length of eIF4GI and eIF4GII fragments (short 12–20 amino acid peptides and 90-residue-long eIF4G fragment), and according to our and previous studies the length and sequences composition of eIF4G peptide

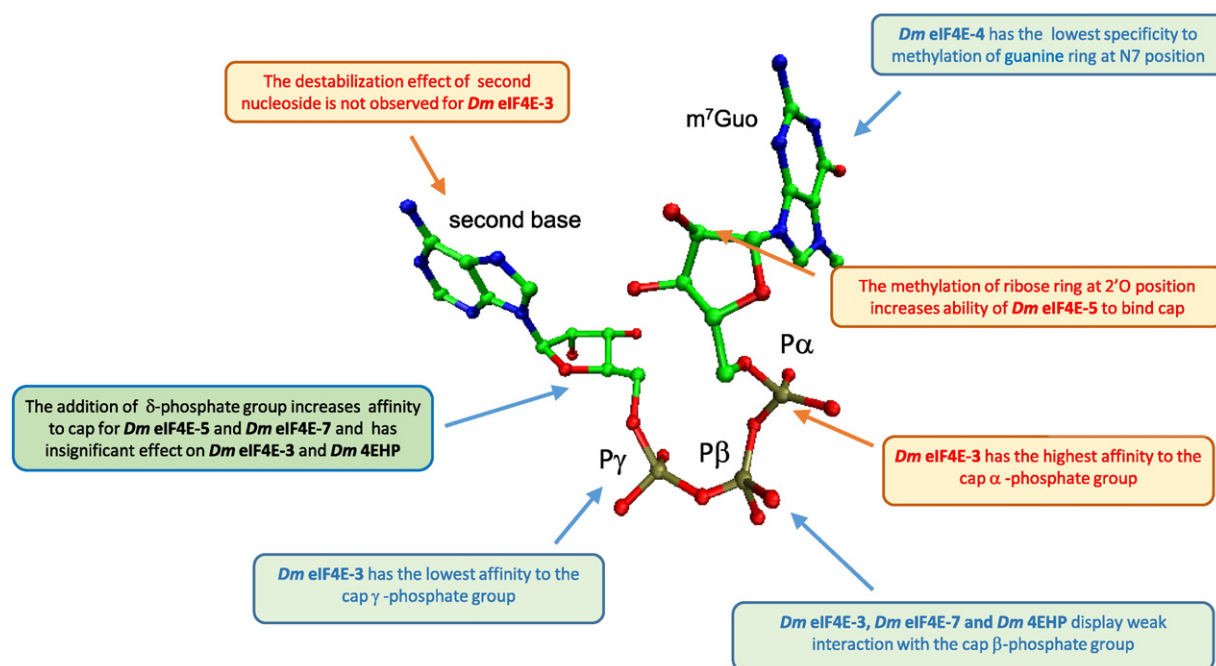


Fig. 6. Schematic representation of the cap structure and its interactions with *Dm* eIF4Es. The differences in ability of the *Drosophila* eIF4E proteins to bind the individual elements of a 5' mRNA cap structure are indicated. The cap is represented by m⁷GpppA. For consistency with text, the numbering of phosphate groups applied for mononucleotide cap analogues was used.

is important for the interaction with eIF4E [39,71,72]. Furthermore, in contrast to *Drosophila* eIF4E, the human protein probably contains a second NC-motif for eIF4G binding [72,73].

Although the cooperativity in formation of ternary eIF4G peptide–eIF4E–cap analogue complexes remains controversial, some NMR, MS and crystallographic studies have shown significant structural differences between a binary and a ternary complex [36,62,74]. eIF4E binding to the cap results in structural rearrangements, not only around the region of cap binding, but also on the dorsal surface of eIF4E and, conversely, eIF4G binding alters the conformation around the cap binding slot especially in the region that interacts with phosphate moiety of the cap [62].

4. Conclusion

Using fluorescence binding studies combined with sequence/structure analyses, we demonstrate shared and distinct features of cap binding by *Drosophila* eIF4E family members. Although Class I *Dm* eIF4Es have all the key conserved residues involved in cap recognition, differences in amino acid sequences confer different binding affinities to the elements of the cap structure: N7-methylguanosine, phosphate chain and the second nucleoside (Fig. 6). The K_{as} values for complexes of *Dm* eIF4Es with individual cap analogues differed by as much as 14-fold. However, the N7-methyl group is crucial for cap recognition by all *Dm* eIF4E proteins.

According to our 3D models, *Dm* eIF4E-3 and eIF4E-5 display, typical for nuclear cap binding complex (CBC), stabilisation of the second cap nucleoside by potential stacking interaction with an aromatic residue: His234 and Phe222, respectively. In human CBC, the second nucleoside is stabilised by stacking interaction with Tyr138 [75] and its substitution by alanine reduces CBC binding affinity for m⁷GpppG about 4-fold [76, 77]. *Dm* eIF4E-3 binds dinucleotide cap analogues 1.3–1.9 fold stronger than human eIF4E. Interestingly, *Dm* eIF4E-3 expression is restricted to testes of adult male flies and regulates translation during spermatogenesis by forming specific eIF4F complexes with eIF4G and eIF4G2, and its activity is not regulated by 4E-BP. Moreover, male flies lacking eIF4E-3 are viable but are completely sterile [28].

Other *Drosophila* eIF4E cognate, for which the biological role is known, is 4EHP that acts as a translation repressor by competing with eIF4E for binding of the mRNA cap. It represses translation of specific maternal mRNAs, such as *caudal* (*cad*) and *hunchback* (*hb*), which are important for establishing the correct anterior-posterior axis polarity in the *Drosophila* embryo [30,31]. 4EHP inhibits translation of *cad* mRNA by simultaneous binding with the cap and Bicoid, which in turn binds to 3'UTR of *cad* mRNA. In the case of *hb* mRNA, 4EHP binds the cap and a Brat protein member of NRE complex required for inhibition of *hb* mRNA [31,78]. In mammals, such inhibition of mechanism involving 4EHP was documented during maturation of mouse oocyte, where 4EHP in complex with Prep1 protein inhibits translation of *Hox4* mRNA [14]. Interactions of 4EHP with specific protein partners together with the lack of eIF4G binding, as well as its low affinity for the cap in *apo* form were considered key factors for translational inhibition of specific mRNAs. The significantly lower (100–200 fold) cap binding affinity of human 4EHP as compared to human eIF4E [56] can prevent cap recognition of all mRNA and hence uncontrolled translation inhibition. *Drosophila* 4EHP binds only 3-fold weaker to m⁷GTP and m⁷GpppG than *Dm* eIF4E-1, suggesting the existence of different additional control mechanisms of 4EHP availability for mRNA recognition in *Drosophila* and human.

Altogether, our studies showed that some local sequence differences play an important role in the eIF4E–cap interactions that might have an impact on their biological role.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgement

We are grateful to Janusz Stepinski and Jacek Jemielity from Division of Biophysics and Centre of New Technology at University of Warsaw for providing us with cap analogues and to Kamil Steczkiewicz for help in preparation of the figures. This research was supported by the Polish

National Science Centre, [UMO-2012/07/B/NZ1/00118 to Edward Darzynkiewicz and JZ, UMO-2011/02/A/NZ2/00014 and UMO-2014/15/B/NZ1/03357 to KG], and by the European Social Fund [EU, UDA-POKL.04.01.01-00-072/09-00 to KK]. We thank the Biopolymers Laboratory, Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, for use the Thermocycler PCR “CXF-96” and Autoclave “Laboklav 135V” co-financed by the European Union within the European Regional Development Fund Project [POIG.02.01.00-14-122/09].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbapap.2016.06.015>.

References

- [1] N. Sonenberg, A.G. Hinnebusch, Regulation of translation initiation in eukaryotes: mechanisms and biological targets, *Cell* 136 (2009) 731–745, <http://dx.doi.org/10.1016/j.cell.2009.01.042>.
- [2] R.J. Jackson, C.U.T. Hellen, T.V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation, *Nat. Rev. Mol. Cell Biol.* 11 (2010) 113–127, <http://dx.doi.org/10.1038/nrm2838>.
- [3] J.D. Richter, N. Sonenberg, Regulation of cap-dependent translation by eIF4E inhibitory proteins, *Nature* 433 (2005) 477–480, <http://dx.doi.org/10.1038/nature03205>.
- [4] G. Hernández, M. Altmann, P. Lasko, Origins and evolution of the mechanisms regulating translation initiation in eukaryotes, *Trends Biochem. Sci.* 35 (2010) 63–73, <http://dx.doi.org/10.1016/j.tibs.2009.10.009>.
- [5] G. Hernández, P. Vazquez-Pianzola, Functional diversity of the eukaryotic translation initiation factors belonging to eIF4 families, *Mech. Dev.* 122 (2005) 865–876, <http://dx.doi.org/10.1016/j.mod.2005.04.002>.
- [6] B. Joshi, K. Lee, D.L. Maeder, R. Jagus, Phylogenetic analysis of eIF4E-family members, *BMC Evol. Biol.* 5 (2005) 48, <http://dx.doi.org/10.1186/1471-2148-5-48>.
- [7] B. Joshi, A. Cameron, R. Jagus, Characterization of mammalian eIF4E-family members, *Eur. J. Biochem.* 271 (2004) 2189–2203, <http://dx.doi.org/10.1111/j.1432-1033.2004.04149.x>.
- [8] R.E. Rhoads, eIF4E: new family members, new binding partners, new roles, *J. Biol. Chem.* 284 (2009) 16711–16715, <http://dx.doi.org/10.1074/jbc.R900002200>.
- [9] A.V. Evisikov, C.M. de Evisikova, Evolutionary origin and phylogenetic analysis of the novel oocyte-specific eukaryotic translation initiation factor 4E in Tetrapoda, *Dev. Genes Evol.* 219 (2009) 111–118, <http://dx.doi.org/10.1007/s00427-008-0268-2>.
- [10] R.M. Patrick, K.S. Browning, The eIF4F and eIFiso4F complexes of plants: an evolutionary perspective, *Comp. Funct. Genomics* 2012 (2012) 287814, <http://dx.doi.org/10.1155/2012/287814>.
- [11] R. Jagus, T.R. Bachvaroff, B. Joshi, A.R. Place, Diversity of eukaryotic translational initiation factor eIF4E in protists, *Comp. Funct. Genomics* 2012 (2012) 134839, <http://dx.doi.org/10.1155/2012/134839>.
- [12] G.D. Jones, E.P. Williams, A.R. Place, R. Jagus, T.R. Bachvaroff, The alveolate translation initiation factor 4E family reveals a custom toolkit for translational control in core dinoflagellates, *BMC Evol. Biol.* 15 (2015) 14, <http://dx.doi.org/10.1186/s12862-015-0301-9>.
- [13] B.D. Keiper, B.J. Lamphear, A.M. Deshpande, M. Jankowska-Anyszka, E.J. Aamodt, T. Blumenthal, et al., Functional characterization of five eIF4E isoforms in *Caenorhabditis elegans*, *J. Biol. Chem.* 275 (2000) 10590–10596, <http://dx.doi.org/10.1074/jbc.275.14.10590>.
- [14] J.C. Villacusa, C. Buratti, D. Penkov, L. Mathiasen, J. Planagumà, E. Ferretti, et al., Cytoplasmic Prep1 interacts with 4EHP inhibiting Hoxb4 translation, *PLoS One* 4 (2009), e5213 <http://dx.doi.org/10.1371/journal.pone.0005213>.
- [15] M. Morita, L.W. Ler, M.R. Fabian, N. Siddiqui, M. Mullin, V.C. Henderson, et al., A novel 4EHP-GIGYF2 translational repressor complex is essential for mammalian development, *Mol. Cell. Biol.* 32 (2012) 3585–3593, <http://dx.doi.org/10.1128/MCB.00455-12>.
- [16] J. Uniacke, C.E. Holterman, G. Lachance, A. Franovic, M.D. Jacob, M.R. Fabian, et al., An oxygen-regulated switch in the protein synthesis machinery, *Nature* 486 (2012) 126–129, <http://dx.doi.org/10.1038/nature11055>.
- [17] J. Uniacke, J.K. Perera, G. Lachance, C.B. Francisco, S. Lee, Cancer cells exploit eIF4E2-directed synthesis of hypoxia response proteins to drive tumor progression, *Cancer Res.* 74 (2014) 1379–1389, <http://dx.doi.org/10.1158/0008-5472.CAN-13-2278>.
- [18] S. Timpano, J. Uniacke, Human cells cultured under physiological oxygen utilize two cap-binding proteins to recruit distinct mRNAs for translation, *J. Biol. Chem.* (2016) <http://dx.doi.org/10.1074/jbc.M116.717363>.
- [19] M.J. Osborne, L. Volpon, J.A. Kornblatt, B. Culjkovic-Kraljic, A. Baguet, K.L.B. Borden, eIF4E3 acts as a tumor suppressor by utilizing an atypical mode of methyl-7-guanosine cap recognition, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 3877–3882, <http://dx.doi.org/10.1073/pnas.1216862110>.
- [20] N. Minshall, M.H. Reiter, D. Weil, N. Standart, CPEB interacts with an ovary-specific eIF4E and 4E-T in early Xenopus oocytes, *J. Biol. Chem.* 282 (2007) 37389–37401, <http://dx.doi.org/10.1074/jbc.M704629200>.
- [21] M. Jankowska-Anyszka, B.J. Lamphear, E.J. Aamodt, T. Harrington, E. Darzynkiewicz, R. Stolarski, R.E. Rhoads, Multiple isoforms of eukaryotic protein-synthesis initiation-factor-4E in *Caenorhabditis elegans* can distinguish between monomethylated and trimethylated messenger RNA cap structure, *J. Biol. Chem.* 273 (1998) 10538–10542, <http://dx.doi.org/10.1074/jbc.273.17.10538>.
- [22] R.S. Mangio, S. Votra, D. Prunne, The canonical eIF4E isoform of *C. elegans* regulates growth, embryogenesis, and germline sex-determination, *Biology Open.* (2015) 1–9, <http://dx.doi.org/10.1242/bio.011585>.
- [23] M.A. Henderson, E. Cronland, S. Dunkelbarger, V. Contreras, S. Strome, B.D. Keiper, A germline-specific isoform of eIF4E (IFE-1) is required for efficient translation of stored mRNAs and maturation of both oocytes and sperm, *J. Cell Sci.* 122 (2009) 1529–1539, <http://dx.doi.org/10.1242/jcs.046771>.
- [24] I. Kawasaki, M.H. Jeong, Y.H. Shim, Regulation of sperm-specific proteins by IFE-1, a germline-specific homolog of eIF4E, in *C. elegans*, *Mol. Cells* 31 (2011) 191–197, <http://dx.doi.org/10.1007/s10059-011-0021-y>.
- [25] T.D. Dinkova, B.D. Keiper, L. Nadejda, E.J. Aamodt, R.E. Rhoads, N.L. Korneeva, Translation of a small subset of *Caenorhabditis elegans* mRNAs is dependent on a specific eukaryotic translation initiation factor 4E isoform translation of a small subset of *Caenorhabditis elegans* mRNAs is dependent on a specific eukaryotic translation in, *Mol. Cell. Biol.* 25 (2005) 100–113, <http://dx.doi.org/10.1128/MCB.25.1.100>.
- [26] G. Hernández, M. Altmann, J.M. Sierra, H. Urlaub, R. Diez Del Corral, P. Schwartz, et al., Functional analysis of seven genes encoding eight translation initiation factor 4E (eIF4E) isoforms in *Drosophila*, *Mech. Dev.* 122 (2005) 529–543, <http://dx.doi.org/10.1016/j.mod.2004.11.011>.
- [27] G. Tettweiler, M. Kowanda, P. Lasko, N. Sonenberg, G. Hernández, The distribution of eIF4E-family members across Insecta, *Comp. Funct. Genomics* 2012 (2012) 960420, <http://dx.doi.org/10.1155/2012/960420>.
- [28] G. Hernandez, H. Han, V. Gandin, L. Fabian, T. Ferreira, J. Zuberek, et al., Eukaryotic initiation factor 4E-3 is essential for meiotic chromosome segregation, cytokinesis and male fertility in *Drosophila*, *Development* 139 (2012) 3211–3220, <http://dx.doi.org/10.1242/dev.073122>.
- [29] S. Ghosh, P. Lasko, Loss-of-function analysis reveals distinct requirements of the translation initiation factors eIF4E, eIF4E-3, eIF4G and eIF4G2 in *Drosophila* spermatogenesis, *PLoS One* 10 (2015), e0122519 <http://dx.doi.org/10.1371/journal.pone.0122519>.
- [30] P.F. Cho, F. Poulin, Y.A. Cho-Park, I.B. Cho-Park, J.D. Chicoine, P. Lasko, et al., A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by bicoid and the eIF4E cognate 4EHP, *Cell* 121 (2005) 411–423, <http://dx.doi.org/10.1016/j.cell.2005.02.024>.
- [31] P.F. Cho, C. Gamberi, Y.A. Cho-Park, I.B. Cho-Park, P. Lasko, N. Sonenberg, Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos, *Curr. Biol.* 16 (2006) 2035–2041, <http://dx.doi.org/10.1016/j.cub.2006.08.093>.
- [32] A. Yarinun, R.E. Harris, M.P. Ashe, H.L. Ashe, Patterning of the *Drosophila* oocyte by a sequential translation repression program involving the d4EHP and Belle translational repressors, *RNA Biol.* 8 (2011) 904–912, <http://dx.doi.org/10.4161/rna.8.5.16325>.
- [33] J. Marcotrigiano, A. Gingras, N. Sonenberg, S.K. Burley, Cocystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP, *Cell* 89 (1997) 951–961.
- [34] A. Niedzwiecka, J. Marcotrigiano, J. Stepinski, M. Jankowska-Anyszka, A. Wyslouchy-Cieszyńska, M. Dadlez, et al., Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins, *J. Mol. Biol.* 319 (2002) 615–635, [http://dx.doi.org/10.1016/S0022-2836\(02\)00328-5](http://dx.doi.org/10.1016/S0022-2836(02)00328-5).
- [35] K. Tomoo, X. Shen, K. Okabe, Y. Nozoe, S. Fukuhara, S. Morino, et al., Structural features of human initiation factor 4E, studied by X-ray crystal analyses and molecular dynamics simulations, *J. Mol. Biol.* 328 (2003) 365–383, [http://dx.doi.org/10.1016/S0022-2836\(03\)00314-0](http://dx.doi.org/10.1016/S0022-2836(03)00314-0).
- [36] C.J. Brown, C.S. Verma, M.D. Walkinshaw, D.P. Lane, Crystallization of eIF4E Complexed With eIF4G Peptide and Glycerol Reveals Distinct Structural Differences Around the Cap-binding Site, 2009 1905–1911.
- [37] K. Kinkelin, K. Veith, M. Grunwald, F. Bono, Crystal structure of a minimal eIF4E-Cup complex reveals a general mechanism of eIF4E regulation in translational repression, *RNA* 18 (2012) 1624–1634, <http://dx.doi.org/10.1261/ma.033639.112>.
- [38] C. Igreja, D. Peter, C. Weiler, E. Izaurralde, 4E-BPs require non-canonical 4E-binding motifs and a lateral surface of eIF4E to repress translation, *Nat. Commun.* 5 (2014) 4790, <http://dx.doi.org/10.1038/ncomms5790>.
- [39] D. Peter, C. Igreja, R. Weber, L. Wohlbold, C. Weiler, L. Eberts, et al., Molecular architecture of 4E-BP translational inhibitors bound to eIF4E, *Mol. Cell* 1–14 (2015) <http://dx.doi.org/10.1016/j.molcel.2015.01.017>.
- [40] E. Darzynkiewicz, I. Ekiel, S.M. Tahara, L.S. Seliger, A.J. Shatkin, Chemical synthesis and characterization of 7-methylguanosine cap analogues, *Biochemistry* 24 (1985) 1701–1707, <http://dx.doi.org/10.1021/bi00328a020>.
- [41] J. Jemielity, T. Fowler, J. Zuberek, J. Stepinski, M. Lewdorowicz, A. Niedzwiecka, et al., Novel “anti-reverse” cap analogs with superior translational properties, *RNA* (New York, N.Y.) 9 (2003) 1108–1122, <http://dx.doi.org/10.1261/ma.5430403>.
- [42] J. Zuberek, J. Jemielity, A. Jablonowska, J. Stepinski, M. Dadlez, R. Stolarski, et al., Influence of electric charge variation at residues 209 and 159 on the interaction of eIF4E with the mRNA 5' terminus, *Biochemistry* 43 (2004) 5370–5379, <http://dx.doi.org/10.1021/bi030266t>.
- [43] A. Cai, M. Jankowska-Anyszka, A. Centers, L. Chlebicka, J. Stepinski, R. Stolarski, et al., Quantitative assessment of mRNA cap analogues as inhibitors of in vitro translation, *Biochemistry* 38 (1999) 8538–8547, <http://dx.doi.org/10.1021/bi9830213>.
- [44] T. Ota, Y. Suzuki, T. Nishikawa, T. Otsuki, T. Sugiyama, R. Irie, et al., Complete sequencing and characterization of 21,243 full-length human cDNAs, *Nat. Genet.* 36 (2004) 40–45, <http://dx.doi.org/10.1038/ng1285>.
- [45] G. Hernández, M. del Mar Castellano, M. Agudo, J.M. Sierra, Isolation and characterization of the cDNA and the gene for eukaryotic translation initiation factor 4G from

- Drosophila melanogaster*, Eur. J. Biochem. / FEBS. 253 (1998) 27–35 (<http://www.ncbi.nlm.nih.gov/pubmed/9578457>).
- [46] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4 (1995) 2411–2423, <http://dx.doi.org/10.1002/pro.5560041120>.
- [47] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402 <http://nar.oxfordjournals.org/content/25/17/3389.short>.
- [48] J. Pei, R. Sadreyev, N.V. Grishin, PCMA: fast and accurate multiple sequence alignment based on profile consistency, *Bioinformatics* 19 (2003) 427–428, <http://dx.doi.org/10.1093/bioinformatics/btg008>.
- [49] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server. *Bioinformatic*. (Oxford, England). 16 (2000) 404–405, <http://dx.doi.org/10.1093/bioinformatics/16.4.404>.
- [50] A. Fiser, A. Šali, MODELLER: generation and refinement of homology-based protein structure models, *Methods Enzymol.* 374 (2003) 461–491, [http://dx.doi.org/10.1016/S0076-6879\(03\)74020-8](http://dx.doi.org/10.1016/S0076-6879(03)74020-8).
- [51] K. Ginalski, L. Rychlewski, Protein structure prediction of CASP5 comparative modelling and fold recognition targets using consensus alignment approach and 3D assessment, *proteins: structure, Funct. Genet.* 53 (2003) 410–417, <http://dx.doi.org/10.1002/prot.10548>.
- [52] P. Rosettani, S. Knapp, M.-G. Vismara, L. Rusconi, A.D. Cameron, Structures of the human eIF4E homologous protein, h4EHP, in its m7GTP-bound and unliganded forms, *J. Mol. Biol.* 368 (2007) 691–705, <http://dx.doi.org/10.1016/j.jmb.2007.02.019>.
- [53] J.A. Ashby, C.E.M. Stevenson, G.E. Jarvis, D.M. Lawson, A.J. Maule, Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to Pea seed-borne mosaic virus in Pea, *PLoS One* 6 (2011), e15973 <http://dx.doi.org/10.1371/journal.pone.0015873>.
- [54] K. Tomoo, X. Shen, K. Okabe, Y. Nozoe, S. Fukuhara, S. Morino, et al., Crystal structures of 7-methylguanosine 5'-triphosphate (m(7)GTP)- and P(1)-7-methylguanosine-P(3)-adenosine-5',5'-triphosphate (m(7)GpppA)-bound human full-length eukaryotic initiation factor 4E: biological importance of the C-terminal flexible region, *Biochem. J.* 362 (2002) 539–544, <http://dx.doi.org/10.1042/0264-6021:3620539>.
- [55] A. Niedzwiecka, J. Stepinski, J.M. Antosiewicz, E. Darzynkiewicz, R. Stolarski, Biophysical Approach to Studies of Cap-eIF4E Interaction by Synthetic Cap Analogs, Elsevier Inc., First Edit, 2007 [http://dx.doi.org/10.1016/S0076-6879\(07\)30009-8](http://dx.doi.org/10.1016/S0076-6879(07)30009-8).
- [56] J. Zuberek, D. Kubacka, A. Jablonowska, J. Jemielity, J. Stepinski, N. Sonenberg, et al., Weak binding affinity of human 4EHP for mRNA cap analogs. *RNA* (New York, N.Y.) 13 (2007) 691–697, <http://dx.doi.org/10.1261/rna.453107>.
- [57] D. Kubacka, R.N. Miguel, N. Minshall, E. Darzynkiewicz, N. Standart, J. Zuberek, Distinct features of Cap binding by eIF4E1b proteins, *J. Mol. Biol.* 427 (2015) 387–405, <http://dx.doi.org/10.1016/j.jmb.2014.11.009>.
- [58] A. Kropiwnicka, K. Kuchta, M. Lukaszewicz, J. Kowalska, J. Jemielity, K. Ginalski, et al., Five eIF4E isoforms from *Arabidopsis thaliana* are characterized by distinct features of cap analogs binding, *Biochem. Biophys. Res. Commun.* 456 (2015) 47–52, <http://dx.doi.org/10.1016/j.bbrc.2014.11.032>.
- [59] K. Kiraga-Motoszko, A. Niedzwiecka, A. Modrak-Wojcik, J. Stepinski, E. Darzynkiewicz, R. Stolarski, Thermodynamics of molecular recognition of mRNA 5' cap by yeast eukaryotic initiation factor 4E, *J. Phys. Chem. B* 115 (2011) 8746–8754, <http://dx.doi.org/10.1021/jp2012039>.
- [60] J.D. Gross, N.J. Moerke, T. Von Der Haar, A.A. Lugovskoy, A.B. Sachs, J.E.G. McCarthy, et al., Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E, *Cell.* 115 (2003) 739–750, [http://dx.doi.org/10.1016/S0092-8674\(03\)00975-9](http://dx.doi.org/10.1016/S0092-8674(03)00975-9).
- [61] A.F. Monzingo, S. Dhaliwal, A. Dutt-Chaudhuri, A. Lyon, J.H. Sadow, D.W. Hoffman, et al., The structure of eukaryotic translation initiation factor-4E from wheat reveals a novel disulfide bond, *Plant Physiol.* 143 (2007) 1504–1518, <http://dx.doi.org/10.1104/pp.106.093146>.
- [62] L. Volpon, M.J. Osborne, I. Topisirovic, N. Siddiqui, K.L.B. Borden, Cap-free structure of eIF4E suggests a basis for conformational regulation by its ligands, *EMBO J.* 25 (2006) 5138–5149, <http://dx.doi.org/10.1038/sj.emboj.7601380>.
- [63] Y. Yoffe, J. Zuberek, M. Lewdorowicz, Z. Zeira, C. Keasar, I. Orr-Dahan, et al., Cap-binding activity of an eIF4E homolog from *Leishmania*, *RNA* (New York, N.Y.) 10 (2004) 1764–1775, <http://dx.doi.org/10.1261/rna.7520404>.
- [64] B.A. Shirley, P. Stanssens, U. Hahn, C.N. Pace, Contribution of hydrogen bonding to the conformational stability of ribonuclease T1, *Biochemistry* 31 (1992) 725–732, <http://dx.doi.org/10.1021/bi00118a013>.
- [65] E. Rom, H.C. Kim, a.C. Gingras, J. Marcotrigiano, D. Favre, H. Olsen, et al., Cloning and characterization of 4EHP, a novel mammalian eIF4E-related cap-binding protein, *J. Biol. Chem.* 273 (1998) 13104–13109.
- [66] A.R. Tee, J.a. Tee, J. Blenis, Characterizing the interaction of the mammalian eIF4E-related protein 4EHP with 4E-BP1, *FEBS Lett.* 564 (2004) 58–62, [http://dx.doi.org/10.1016/S0014-5793\(04\)00313-8](http://dx.doi.org/10.1016/S0014-5793(04)00313-8).
- [67] D.E. Friedland, W.N.B. Wooten, J.E. Lavoy, C.H. Hagedorn, D.J. Goss, A mutant of eukaryotic protein synthesis initiation factor eIF4E, *Biochemistry* 44 (2005) 4546–4550.
- [68] M. Ptushkina, T. Von Der Haar, M.M. Karim, J.M.X. Hughes, J.E.G. McCarthy, Repressor binding to a dorsal regulatory site traps human eIF4E in a high cap-affinity state, *EMBO J.* 18 (1999) 4068–4075, <http://dx.doi.org/10.1093/emboj/18.14.4068>.
- [69] W.C. Merrick, eIF4F: a retrospective, *J. Biol. Chem.* 290 (2015) 24091–24099, <http://dx.doi.org/10.1074/jbc.R115.675280>.
- [70] S.V. Slepnev, N.L. Korneeva, R.E. Rhoads, Kinetic mechanism for assembly of the m7GpppG · eIF4E · eIF4G complex, *J. Biol. Chem.* 283 (2008) 25227–25237, <http://dx.doi.org/10.1074/jbc.M801786200>.
- [71] C.J. Brown, J.J. Lim, T. Leonard, H.C.a. Lim, C.S.B. Chia, C.S. Verma, et al., Stabilizing the eIF4G1 α -helix increases its binding affinity with eIF4E: implications for peptidomimetic design strategies, *J. Mol. Biol.* 405 (2011) 736–753, <http://dx.doi.org/10.1016/j.jmb.2010.10.045>.
- [72] Y. Umenaga, K.S. Paku, Y. In, T. Ishida, K. Tomoo, Identification and function of the second eIF4E-binding region in N-terminal domain of eIF4G: comparison with eIF4E-binding protein, *Biochem. Biophys. Res. Commun.* 414 (2011) 462–467, <http://dx.doi.org/10.1016/j.bbrc.2011.09.084>.
- [73] D. Lama, C.J. Brown, D.P. Lane, C.S. Verma, Gating by tryptophan 73 exposes a cryptic pocket at the protein-binding interface of the oncogenic eIF4E protein, *Biochemistry* 151009161412001 (2015) <http://dx.doi.org/10.1021/acs.biochem.5b00812>.
- [74] I. Rutkowska-Wlodarczyk, J. Stepinski, M. Dadlez, E. Darzynkiewicz, R. Stolarski, A. Niedzwiecka, Structural changes of eIF4E upon binding to the mRNA 5' monomethylguanosine and trimethylguanosine cap, *Biochemistry* 47 (2008) 2710–2720, <http://dx.doi.org/10.1021/bi701168z>.
- [75] C. Mazza, A. Segref, I.W. Mattaj, S. Cusack, Large-scale induced fit recognition of an m7GpppG cap analogue by the human nuclear cap-binding complex, *EMBO J.* 21 (2002) 5548–5557, <http://dx.doi.org/10.1093/emboj/cdf538>.
- [76] R. Worch, A. Niedzwiecka, J. Stepinski, C. Mazza, M. Jankowska-Anyszka, E. Darzynkiewicz, et al., Specificity of recognition of mRNA 5' cap by human nuclear cap-binding complex. *RNA* (New York, N.Y.) 11 (2005) 1355–1363, <http://dx.doi.org/10.1261/rna.2850705>.
- [77] R. Worch, M. Jankowska-Anyszka, A. Niedzwiecka, J. Stepinski, C. Mazza, E. Darzynkiewicz, et al., Diverse role of three tyrosines in binding of the RNA 5' cap to the human nuclear cap binding complex, *J. Mol. Biol.* 385 (2009) 618–627, <http://dx.doi.org/10.1016/j.jmb.2008.10.092>.
- [78] D. Chagnovich, R. Lehmann, Poly(A)-independent regulation of maternal hunchback translation in the *Drosophila* embryo, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 11359–11364, <http://dx.doi.org/10.1073/pnas.201284398>.
- [79] J. Marcotrigiano, A.C. Gingras, N. Sonenberg, S.K. Burley, Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G, *Mol. Cell* 3 (1999) 707–716, [http://dx.doi.org/10.1016/S1097-2765\(01\)80003-4](http://dx.doi.org/10.1016/S1097-2765(01)80003-4).

Supplementary Data

Diverse cap-binding properties of *Drosophila* eIF4E isoforms

Joanna Zuberek^{a*,1}, Krzysztof Kuchta^{b,c,1}, Greco Hernández^d, Nahum Sonenberg^e,
Krzysztof Ginalski^b

^a Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, 02-089, Poland, jzuberek@biogeo.uw.edu.pl

^b Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, Warsaw, 02-097, Poland, kginal@cent.uw.edu.pl

^c College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, 02-089, Poland, kuchtak@cent.uw.edu.pl

^d Division of Basic Research, National Institute of Cancer (INCan), Tlalpan, Mexico City, 14080, Mexico, greco.hernandez@gmail.com

^e Department of Biochemistry and Goodman Cancer Research Centre, McGill University, Montreal, QC H3A 1A3, Canada, nahum.sonenberg@mcgill.ca

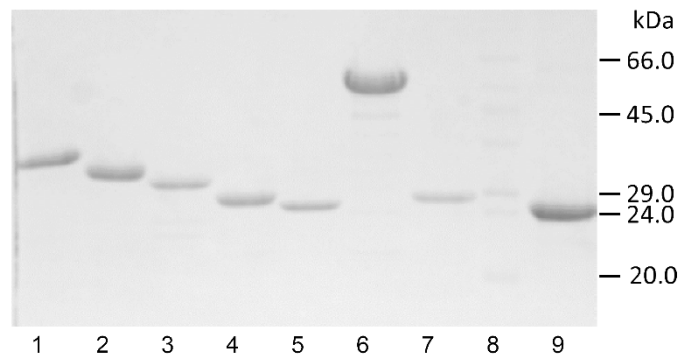
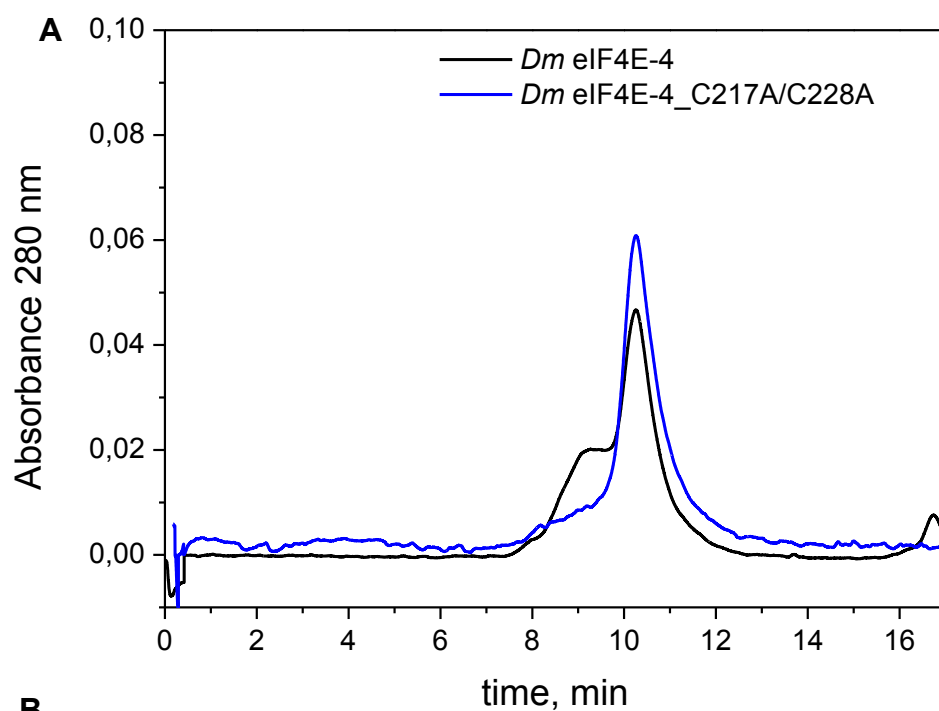


Figure S1. Analysis of recombinant eIF4E proteins (purified on ion-exchange column) by 15% SDS-PAGE and Bio-Safe Coomassie staining. Lane 1, *Dm* eIF4E-1; Lane 2, *Dm* eIF4E-2; Lane 3, *Dm* eIF4E-3; Lane 4, *Dm* eIF4E-4; Lane 5, *Dm* eIF4E-5; Lane 6, *Dm* eIF4E-7; Lane 7, *Dm* 4EHP; Lane 8, Molecular weight ladder (SIGMA); Lane 9, human eIF4E.



B

Ligand	<i>Dm</i> eIF4E-4 (apo)	<i>Dm</i> eIF4E-4_C217A/C228A (apo)
	K_{as} (μM^{-1})	
m ⁷ GTP	2.12 ± 0.06	2.32 ± 0.07
<i>Dm</i> eIF4G peptide	7.39 ± 0.60	6.6 ± 0.9
	<i>Dm</i> eIF4E-4 (saturated with <i>Dm</i> eIF4G peptide)	<i>Dm</i> eIF4E-4_C217A/C228A (saturated with <i>Dm</i> eIF4G peptide)
	K_{as} (μM^{-1})	
m ⁷ GTP	3.70 ± 0.05	3.55 ± 0.14

Figure S2. Comparison of biochemical and cap binding properties of *Dm* eIF4E-4 isoform and its mutant, in which cysteines 217 and 228 were replaced by alanine. (A) Molecular weight analysis of *Dm* eIF4E-4 isoform (26372.1 Da) and its mutant *Dm* eIF4E-4_C217A/C228A (26308.0 Da) samples, carried out on a size exclusion chromatography column – Enrich SEC70 (BIO-RAD). The separations were performed in 50 mM Hepes/KOH (pH 7.2), 300 mM KCl at 20°C and at a flow rate of 0.5 ml/min. The chromatography runs show the presence of the monomeric and dimeric state of *Dm* eIF4E-4 and only the monomeric state for its mutant. (B) Equilibrium association constants (K_{as}) of binary and ternary complex of *Dm* eIF4E-4 and its mutants with m⁷GTP and *Dm* eIF4G peptide.

Table S1. Numbering position of key residues taking part in interaction with the cap and eIF4G protein for human and *Drosophila* eIF4E

Class I							Class II	
<i>Hs</i> eIF4E	<i>Dm</i> eIF4E-1	<i>Dm</i> eIF4E-2	<i>Dm</i> eIF4E-3	<i>Dm</i> eIF4E-4	<i>Dm</i> eIF4E-5	<i>Dm</i> eIF4E-7	<i>Hs</i> 4EHP	<i>Dm</i> 4EHP
Trp56	Trp100	Trp89	Trp86	Trp71	Trp74	Trp270	Tyr78	Tyr68
Trp102	Trp146	Trp135	Trp132	Trp117	Trp120	Trp316	Trp124	Trp114
Glu103	Glu147	Glu136	Glu133	Glu118	Glu121	Glu317	Glu124	Glu115
Arg112	Arg156	Arg145	Arg142	Arg127	Arg130	Arg326	Lys134	Gln124
Arg157	Arg199	Arg188	Arg185	Arg170	Arg173	Arg369	Arg174	Lys164
Lys159	Lys201	Lys190	Lys187	Lys172	Lys175	Lys371	Glu176	Pro166
Lys162	Lys204	Lys193	Lys190	Lys175	Lys178	Lys374	Ile179	Ser169
Trp166	Trp208	Trp197	Trp194	Trp179	Trp182	Trp378	Trp183	Trp173

Table S2. Binding free energy (ΔG°) for various complexes of human and *Drosophila* eIF4E proteins with cap analogues at 20°C. Standard Gibbs free energy of binding was calculated from equilibrium association constants (K_{as}) obtained based on titration assay, according to the standard equation $\Delta G^\circ = -RT \ln K_{as}$.

Cap Analogue	Class I							Class II	
	<i>Hs</i> eIF4E	<i>Dm</i> eIF4E-1	<i>Dm</i> eIF4E-2	<i>Dm</i> eIF4E-3	<i>Dm</i> eIF4E-4	<i>Dm</i> eIF4E-5	<i>Dm</i> eIF4E-7	<i>Hs</i> 4EHP	<i>Dm</i> 4EHP
	ΔG° (kcal/mol)								
m ⁷ GMP	-7.93 ± 0.03	-7.40 ± 0.02	-7.34 ± 0.04	-8.67 ± 0.01	-6.72 ± 0.06	-7.94 ± 0.02	-7.56 ± 0.01	-6.49 ± 0.12	-7.42 ± 0.02
m ⁷ GDP	-9.73 ± 0.01	-8.62 ± 0.01	-8.59 ± 0.01	-9.49 ± 0.01	-7.80 ± 0.02	-9.19 ± 0.01	-8.36 ± 0.01	-7.20 ± 0.08	-8.00 ± 0.02
m ⁷ GTP	-10.52 ± 0.01	-9.31 ± 0.02	-9.28 ± 0.02	-9.77 ± 0.02	-8.49 ± 0.02	-9.72 ± 0.05	-9.01 ± 0.01	-7.84 ± 0.03	-8.66 ± 0.02
m ⁷ Gp ₄	-11.40 ± 0.03	-10.02 ± 0.02	-9.99 ± 0.03	-10.17 ±	-9.22 ± 0.02	-10.76 ± 0.01	-10.03 ± 0.02	-8.15 ± 0.02	-8.91 ± 0.04
m ⁷ GpppG	-9.11 ± 0.01	-8.19 ± 0.02	-8.16 ± 0.02	-9.27 ± 0.02	-7.48 ± 0.01	-8.91 ± 0.03	-8.302 ± 0.004	-7.02 ± 0.02	-7.94 ± 0.02
m ₂ ^{7,2′O} GpppG	-9.08 ± 0.01	-8.18 ± 0.01	-8.07 ± 0.02	-9.32 ± 0.04	-7.43 ± 0.01	-9.03 ± 0.01	-8.133 ± 0.004	-6.80 ± 0.03	-7.98 ± 0.01
m ⁷ GpppA	-8.70 ± 0.01	-7.90 ± 0.05	-7.86 ± 0.02	-9.07 ± 0.01	-7.20 ± 0.01	-8.57 ± 0.02	-7.78 ± 0.03	-6.85 ± 0.02	-7.84 ± 0.01
m ⁷ GpppC	-8.76 ± 0.03	-8.02 ± 0.01	-8.01 ± 0.01	-8.96 ± 0.04	-7.27 ± 0.01	-8.71 ± 0.02	-7.99 ± 0.02	-6.97 ± 0.07	-7.86 ± 0.04
m ⁷ GppppG	-10.32 ± 0.01	-9.17 ± 0.01	-9.14 ± 0.01	-10.12 ±	-8.36 ± 0.01	-9.86 ± 0.08	-9.03 ± 0.02	-7.86 ± 0.01	-8.58 ± 0.04
GTP	-6.23 ± 0.02	-6.11 ± 0.09	-6.30 ± 0.12	-6.08 ± 0.02	-6.48 ± 0.03	-6.37 ± 0.03	-6.52 ± 0.04	-6.02 ± 0.04	-5.59 ± 0.08

Table S3. Equilibrium association constants (K_{as}) for the formation of binary and ternary complexes of *Dm* eIF4E-4 and human eIF4E proteins with eIF4G peptides, and with the cap analogs m⁷GTP and m⁷GpppG.

	<i>Dm</i> eIF4G peptide	<i>Dm</i> eIF4G3A peptide	<i>Hs</i> eIF4G peptide
K_{as} for the peptides, (μM^{-1})			
<i>Dm</i> eIF4E -4 (apo)	7.39 ± 0.60	<0.03	4.64 ± 0.14
<i>Dm</i> eIF4E-4 (saturated with m ⁷ GTP)	10.95 ± 0.88	ND	9.32 ± 0.49
<i>Dm</i> eIF4E-4 (saturated with m ⁷ GpppG)	12.47 ± 0.79	ND	8.55 ± 0.35
<i>Hs</i> eIF4E (apo)	4.44 ± 0.62	<0.17	7.08 ± 0.30
<i>Hs</i> eIF4E (saturated with m ⁷ GTP)	ND	ND	7.75 ± 0.30
<i>Hs</i> eIF4E (saturated with m ⁷ GpppG)	ND	ND	7.52 ± 0.68
K_{as} for m ⁷ GTP, (μM^{-1})			
<i>Dm</i> eIF4E -4 (apo)	2.12 ± 0.06		
<i>Dm</i> eIF4E-4 (saturated with the peptide)	3.70 ± 0.05	2.24 ± 0.08	4.00 ± 0.06
<i>Hs</i> eIF4E (apo)	73.5 ± 2.4		
<i>Hs</i> eIF4E (saturated with the peptide)	70.3 ± 3.1	73.4 ± 3.2	68.4 ± 5.5
K_{as} for m ⁷ GpppG, (μM^{-1})			
<i>Dm</i> eIF4E -4 (apo)	0.40 ± 0.01		
<i>Dm</i> eIF4E-4 (saturated with the peptide)	0.58 ± 0.02	0.41 ± 0.01	0.69 ± 0.01
<i>Hs</i> eIF4E (apo)	6.17 ± 0.25		
<i>Hs</i> eIF4E (saturated with the peptide)	N.D	N.D	5.41 ± 0.18

Warszawa, dn. 20 lipca 2017 r.

dr Joanna Maria Żuberek

Zakład Biofizyki, Instytut Fizyki Doświadczalnej

Wydział Fizyki, Uniwersytet Warszawski

ul. Żwirki i Wigury 93

02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy: *Zuberek J., Kuchta K., Hernandez G., Sonenberg N., Ginalski K. Diverse cap-binding properties of Drosophila eIF4E isoforms. 2016 Biochimica et Biophysica Acta, Proteins and Proteomics 1864:1292-1303*

mój udział polegał na:

- na sformułowaniu problemu badawczego i stworzeniu koncepcji pracy;
- zaplanowaniu wszystkich badań doświadczalnych w zakresie: przygotowania konstruktów do otrzymania rekombinowanych białek eIF4E z *Drosophila melanogaster*, ich otrzymania w komórkach *E. coli* oraz procedur oczyszczenia;
- zaplanowaniu badań biofizycznych mających na celu scharakteryzowanie powinowactwa białek eIF4E z muszki owocowej do analogów kapu i peptydowego fragmentu białka eIF4G w kompleksach dwu i trójskładnikowych;
- wykonaniu wszystkich badań doświadczalnych związanych z otrzymywaniem białek oraz przeprowadzeniu wszystkich badań biofizycznych w zakresie spektroskopii;
- wyznaczeniu stałych równowagowych stałych asocjacji dla badanych kompleksów z przeprowadzonych miareczkowań fluorescencyjnych oraz ich analizie i interpretacji.
- napisaniu pierwszej wersji manuskryptu, z wyłączeniem metodologii homologicznego modelowania białek i opisu różnic sekwencyjno-strukturalnych pomiędzy białkami eIF4E z *A. thaliana*, oraz ostatecznej redakcji manuskryptu uwzględniającej uwagi i poprawki współautorów. Przygotowaniu wszystkich tabel, oraz rysunków: nr 2A, nr 3A, nr 5A i 5B, nr 6. Przygotowaniu Supelmentary do manuskryptu.

J. Żuberek

20 lipiec 2017

.....
Miejscowość i data

Krzysztof Kuchta
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

Zuberek J., Kuchta K., Hernandez G., Sonenberg N., Ginalski K.

Diverse cap-binding properties of Drosophila eIF4E isoforms.

2016 Biochimica et Biophysica Acta, Proteins and Proteomics 1864:1292-1303

mój udział polegał na:

- Analizie sekwencyjno-strukturalnej homologów eIF4E,
- Stworzeniu modeli struktury przestrzennej izoform eIF4E z *D. melanogaster*,
- Zidentyfikowaniu i opisaniu w manuskrypcie różnic sekwencyjno-strukturalnych pomiędzy białkami eIF4E z *D. melanogaster*, które potencjalnie tłumaczą otrzymane wyniki doświadczalne (różnice w powinowactwie wiązania izoformy eIF4E z analogiem czapeczki i peptydem eIF4G),
- Przygotowaniu rysunku 1, rysunku 2 (panel B), rysunku 3 (panele B i C), rysunku 4, rysunku 5 (panel C),

.....
Krzysztof Kuchta
Podpis

Mexico City, 31st July 2017

.....

Greco Hernández
Division of Basic Research,
National Institute of Cancer (INCan),
Tlalpan, Mexico City 14080,
Mexico

AUTHORSHIP STATEMENT

I hereby declare that in the publication

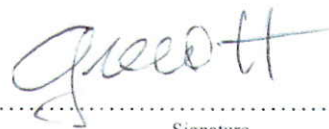
Zuberek J., Kuchta K., Hernandez G., Sonenberg N., Ginalski K.

Diverse cap-binding properties of Drosophila eIF4E isoforms.

2016 Biochimica et Biophysica Acta, Proteins and Proteomics 1864:1292-1303

I participated in drafting and revising the manuscript.

I revised the manuscript critically for important intellectual content.



.....

Signature

Montreal, July 28 2017
City and Date

Nahum Sonenberg
Gilman Cheney Professor
Department of Biochemistry and Goodman Cancer Research Centre
McGill University
1160 Pine Avenue West, Montreal QC H3A 1A3
Montreal, QC H3A 1A3
Canada

AUTHORSHIP STATEMENT

I hereby declare that in the publication

Zuberek J., Kuchta K., Hernandez G., Sonenberg N., Ginalski K.

Diverse cap-binding properties of Drosophila eIF4E isoforms.

2016 Biochimica et Biophysica Acta, Proteins and Proteomics 1864:1292-1303

I participated in drafting and/or revising the manuscript.

I performed critical revision of the article.

I revised the manuscript critically for important intellectual content.



.....
Signature

7 sierpień 2017

.....
Miejscowość i data

Krzysztof Ginalski
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

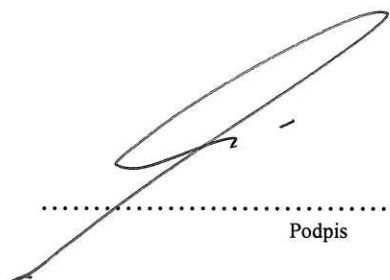
Oświadczam, że w pracy:

Zuberek J., Kuchta K., Hernandez G., Sonenberg N., Ginalski K.

Diverse cap-binding properties of Drosophila eIF4E isoforms.

2016 Biochimica et Biophysica Acta, Proteins and Proteomics 1864:1292-1303

mój udział polegał na zaplanowaniu badań bioinformatycznych, współuczestnictwie w analizie otrzymanych danych doświadczalnych oraz pomocy w pisaniu manuskryptu.


.....
Podpis

Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity toward CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails

KAMIL KOBYLECKI,^{1,2,4} KRZYSZTOF KUCHTA,^{3,4} ANDRZEJ DZIEMBOWSKI,^{1,2} KRZYSZTOF GINALSKI,³ and RAFAŁ TOMECKI^{1,2}

¹Laboratory of RNA Biology and Functional Genomics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland

²Department of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, 02-106 Warsaw, Poland

³Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, 02-089 Warsaw, Poland

ABSTRACT

Noncanonical RNA nucleotidyltransferases (NTases), including poly(A), poly(U) polymerases (PAPs/PUPs), and C/U-adding enzymes, modify 3'-ends of different transcripts affecting their functionality and stability. They contain PAP/OAS1 substrate-binding domain (SBD) with inserted NTase domain. *Aspergillus nidulans* CutA (AnCutA), synthesizes C/U-rich 3'-terminal extensions in vivo. Here, using high-throughput sequencing of the 3'-RACE products for tails generated by CutA proteins in vitro in the presence of all four NTPs, we show that even upon physiological ATP excess synthesized tails indeed contain an unprecedented number of cytidines interrupted by uridines and stretches of adenosines, and that the majority end with two cytidines. Strikingly, processivity assays documented that in the presence of CTP as a sole nucleotide, the enzyme terminates after adding two cytidines only. Comparison of our CutA 3D model to selected noncanonical NTases of known structures revealed substantial differences in the nucleotide recognition motif (NRM) within PAP/OAS1 SBD. We demonstrate that CutA specificity toward CTP can be partially changed to PAP or PUP by rational mutagenesis within NRM and, analogously, Cid1 PUP can be converted into a C/U-adding enzyme. Collectively, we suggest that a short cluster of amino acids within NRM is a determinant of NTases' substrate preference, which may allow us to predict their specificity.

Keywords: CutA; nucleotidyltransferase (NTase); noncanonical poly(A) polymerase; NTP specificity; 3'-RACE-seq

INTRODUCTION

3'-End processing and modification contribute to determination of stability of various RNA molecules. The most widespread post-transcriptional modification of the RNA 3'-end is a template-independent addition of nucleotides. This process is catalyzed by enzymes belonging to the DNA polymerase β -like nucleotidyltransferase (NTase) superfamily (Aravind and Koonin 1999; Martin and Keller 2007) (also known as NTase fold superfamily [Kuchta et al. 2009]), which display variable substrate specificity. The best characterized representatives of these NTases are (i) canonical poly(A) polymerases (PAPs), responsible for polyadenylation of mRNA and (ii) CCA-adding enzymes, synthesizing CCA triplet at the 3'-end of tRNAs (Yue et al. 1996; Martin and

Keller 2007). Furthermore, there is an expanding group of noncanonical NTases, encompassing not only atypical PAPs, but also 2'-5'-oligoadenylate synthetases (OAS), poly(U) polymerases (PUPs), and terminal uridylyltransferases (TUTases) (Aphasizhev 2005; Stevenson and Norbury 2006; Kwak and Wickens 2007; Martin and Keller 2007; Rissland and Norbury 2008; Torralba et al. 2008; Schmidt and Norbury 2010).

Shortening or removal of poly(A) tails from the 3'-ends of mRNAs has been long considered as a rate-limiting step of their degradation, which may then proceed in a 5'-3' direction (through concerted action of decapping enzymes and Xrn1 exoribonuclease) or in a 3'-5' direction (catalyzed by

⁴These authors contributed equally to this work.

Corresponding authors: rtom1916@gmail.com, kginal@cent.uw.edu.pl

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.061010.117>.

© 2017 Kobylecki et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

the RNA exosome complex) (for review, see Labno et al. 2016a). Studies in *Schizosaccharomyces pombe* revealed that deadenylation is not an absolute prerequisite for degradation, and uridylation by noncanonical cytoplasmic NTase Cid1 stimulates decapping instead (Rissland and Norbury 2009). Importantly, both oligo(A) tails arising after deadenylation and 3'-uridylation tags are recognized by the Lsm1-7/Pat1 complex, which recruits decapping machinery and is therefore crucial for both deadenylation-dependent and -independent decay (Chowdhury et al. 2007; Song and Kiledjian 2007).

mRNA uridylation and the responsible PUPs gained significant attention because they are conserved among eukaryotes (except *Saccharomyces cerevisiae*, which genome does not encode noncanonical NTase with PUP activity), although some species-specific differences exist (Scott and Norbury 2013; Munoz-Tello et al. 2015; Viegas et al. 2015; Scheer et al. 2016). For instance, uridylation of oligo(A) tails in *Arabidopsis thaliana* by URT1 PUP does not seem to influence the mRNA degradation rates globally, but rather protects 3'-ends from trimming to ensure a 5'-3' directionality of cotranslational mRNA decay (Sement et al. 2013; Scheer et al. 2016; Zuber et al. 2016). In mammals, uridylation catalyzed by Cid1 homologs—TUT4 (ZCCHC11) or TUT7 (ZCCHC6)—occurs on deadenylated mRNAs with oligo(A) tails shorter than 25 nt (Chang et al. 2014; Lim et al. 2014). Furthermore, all above-mentioned TUTases regulate small RNA biogenesis and activity. In human stem cells or certain cancer cells, TUT4/7 are recruited to pre-miRNAs by Lin28 protein, which leads to 3'-uridylation and degradation by a Perlman syndrome exonuclease DIS3L2 (Heo et al. 2009; Chang et al. 2013; Ustianenko et al. 2013; Faehnle et al. 2014; Kim et al. 2015). Notably, nucleolytic activity of the latter, stimulated by TUTase-mediated oligo(U) tail synthesis, has been recently documented to control quality of multiple noncoding cytoplasmic transcripts in different species (Labno et al. 2016b; Pirouz et al. 2016; Reimao-Pinto et al. 2016; Ustianenko et al. 2016).

In addition to uridylation, a substantial fraction of eukaryotic poly(A)⁺ mRNAs is subject to guanylation, performed by as yet unidentified NTases (Chang et al. 2014; Zuber et al. 2016). The function of guanosine-rich extensions remains unknown—it was proposed that they may protect poly(A) tails from deadenylation (Chang et al. 2014).

Another interesting example of posttranscriptional mRNA 3'-end modification was discovered in *Aspergillus nidulans*. It involves addition of short (1–7 nt) C/U-rich tails by a Cid1-related enzyme, CutA (Morozov et al. 2010). Similar to *S. pombe*, such heteropolymeric C/U extensions are added to poly(A) tails, circumventing the need for their shortening prior to decapping (Morozov et al. 2010).

Several crystal structures of Cid1 provided insights into the catalytic mechanism of uridylation (Lunde et al. 2012; Munoz-Tello et al. 2012, 2014; Yates et al. 2012, 2015). The structural basis for addition of other nucleotides, like cytidine in the case of CutA, has not been explained. Noncanonical

NTases generally comprise catalytic and central domains (the latter also known as PAP/OAS1 substrate binding domain [SBD]) (Fig. 1; Aravind and Koonin 1999; Stevenson and Norbury 2006; Martin and Keller 2007; Kuchta et al. 2009), and some of them may additionally contain other functionally important motifs, like zinc fingers (Hagan et al. 2009; Heo et al. 2009; Rajappa-Titu et al. 2016). The catalytic domain of Cid1 is built from a five-stranded β -sheet flanked by α -helices (Lunde et al. 2012; Yates et al. 2012). The second β -strand contains a DxD motif between positions 101–103, which, together with the third D in the fifth β -strand (D160) (Fig. 1), interacts with the incoming RNA molecule and two metal ions, thus creating the active site, located at the bottom of the cleft separating catalytic and central domains (Lunde et al. 2012; Munoz-Tello et al. 2012; Yates et al. 2012). The three catalytic aspartates also face β - and γ -phosphates of the UTP substrate (Munoz-Tello et al. 2012). The α -helical central domain contains a nucleotide recognition motif (NRM)—it corresponds to a $\sim$ 10–15 aa-long loop (located between positions 327–339 in Cid1) (Fig. 1), forming one end of the NTP-binding pocket (Lunde et al. 2012; Munoz-Tello et al. 2012). The residues in the NRM, together with amino acids from other parts of the protein (Fig. 1) stabilize the base of the NTP substrate via water-mediated and/or direct hydrogen bonds with their side chain atoms (Munoz-Tello et al. 2012). The hydrogen bond crucial for UTP selectivity of Cid1 is formed by the unique histidine (H336) in the NRM (Fig. 1), and the nucleobase (Lunde et al. 2012; Munoz-Tello et al. 2012; Yates et al. 2012). The affinity of Cid1 toward UTP is 30-fold higher than toward ATP and substituting H336 with asparagine or alanine significantly increases Cid1 ability to utilize ATP as a substrate (Lunde et al. 2012; Yates et al. 2012). UTP recognition by Cid1 is supported by water-mediated interactions between D330 and E333 (Fig. 1) and uridine (Lunde et al. 2012; Yates et al. 2012). The preference for UTP over CTP is partially due to the protonation state of H336, which disfavors formation of the hydrogen bond with the latter (Lunde et al. 2012). Cid1 interaction with UTP is further stabilized by several other hydrogen bonds, involving particularly N171 and T172 (Fig. 1), which contact the ribose directly or via water bridge, and K193/K197, interacting with phosphate groups with some contribution from S90, S211, and Y212 (Fig. 1; Lunde et al. 2012; Munoz-Tello et al. 2012; Yates et al. 2012).

In this study, we investigated the molecular basis for the ability of *A. nidulans* CutA and its homolog from *Thielavia terrestris*—a thermophilic fungus species—to add both cytidine and uridine to the 3'-ends of RNA. To assess preference of these enzymes toward different NTPs, products of in vitro nucleotidylation were subjected to the 3'-RACE analysis coupled with high-throughput sequencing (3'-RACE-seq), a methodology never previously used to study specificity of NTases. We noted that the wild-type versions of CutA proteins have an unprecedented preference toward cytidine,

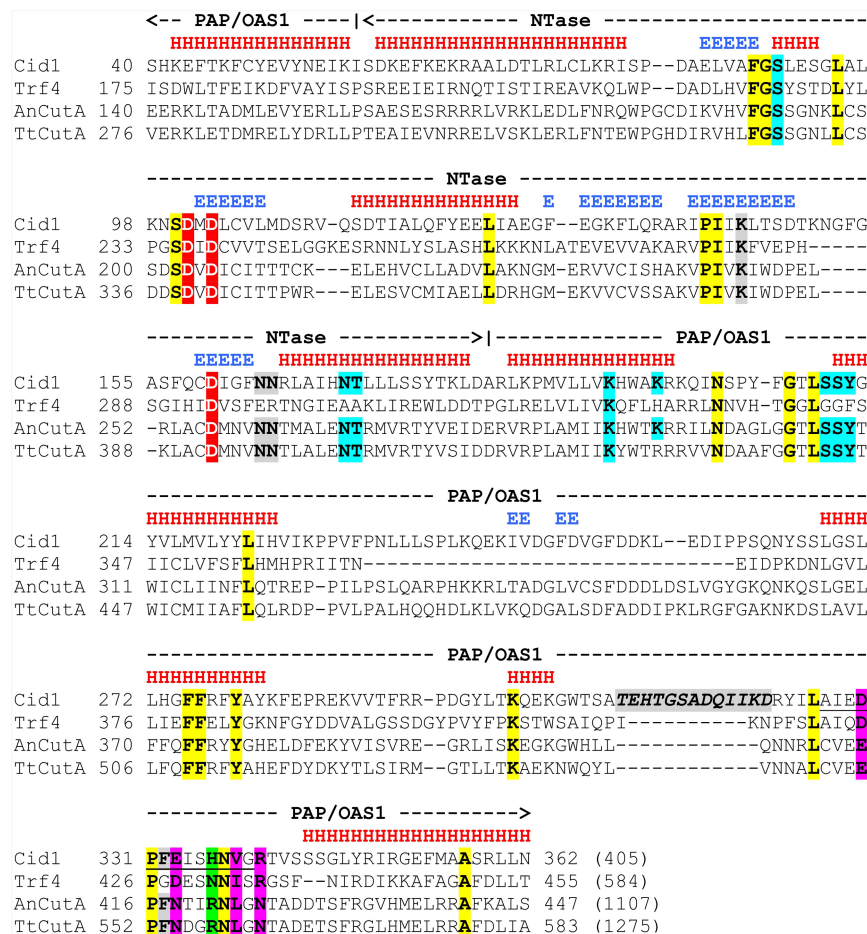


FIGURE 1. Multiple sequence alignment of the catalytic (NTase) and central (PAP/OAS1 SBD) domains from *S. pombe* Cid1 poly(U) polymerase, *S. cerevisiae* Trf4 poly(A) polymerase, and CutA proteins from *A. nidulans* (AnCutA) and *T. terrestris* (TtCutA). Domain ranges are shown above the alignment with dashed lines ending with arrowheads. Locations of observed secondary structure elements in Cid1: α -helices and β -strands are marked with red "H" and blue "E" letters, respectively, above its amino acid sequence. Three conserved aspartates in the active site are highlighted in red. H336, primarily responsible for UTP selectivity of Cid1, and corresponding amino acids in other proteins are marked with green. Other residues within or in the vicinity of the nucleotide recognition motif (NRM), including those that were subjected to mutations in our study, are highlighted in magenta. Amino acids from other parts of the protein, involved in NTP recognition, are marked with blue. Remaining residues important for nucleotidyltransferase activity are highlighted in gray (β -trapdoor loop, unique for Cid1, is additionally marked with italics). Cid1 NRM (residues 327–339) is underlined. Other highly evolutionary conserved amino acids are indicated in yellow.

even in the excess of ATP over the other nucleotides, mimicking physiological conditions. Furthermore, most of the tails ended with two cytidines. This was in agreement with results of biochemical assays, which demonstrated low CutA processivity in the presence of CTP as the sole nucleotide and termination of the reaction after addition of only two cytidine residues. In silico structure modelling revealed substantial differences in the NTP-binding site, which likely explain distinct substrate specificity of CutA in comparison to other noncanonical NTases. Corroborating our presumptions, we were able to partially convert CutA into PAP or PUP by changing several amino acids critical for NTP recognition.

Similarly, we successfully transformed Cid1 PUP into a variant with increased specificity toward CTP. Our data provide new mechanistic insight into the catalytic mechanism of RNA nucleotidylation.

RESULTS

Fungal CutA proteins are noncanonical NTases with enzymatic activity dependent on the intact conserved aspartate residues in the active site

In this study, we focused on CutA protein from *A. nidulans* (AnCutA), which was previously shown to synthesize C/U-rich 3'-terminal tails in vivo (Morozov et al. 2010), and on its close homolog from *T. terrestris* (TtCutA). CutA proteins contain both catalytic (NTase) and central (PAP/OAS1 SBD) domains and show significant sequence similarity to several NTases of known structure, including poly(U) polymerase Cid1 and noncanonical poly(A) polymerase Trf4 (Figs. 1, 2A; Lunde et al. 2012; Munoz-Tello et al. 2012, 2014; Yates et al. 2012, 2015). In CutA enzymes, NTase and PAP/OAS1 SBD domains are flanked with a relatively short upstream fragment and C-terminal region, considerably extended in comparison to Cid1 and Trf4 (Fig. 2A).

To analyze the specificity of CutA enzymes, we first produced recombinant AnCutA (Fig. 2B), which displayed a nucleotidyltransferase activity, but its efficiency varied depending on the NTP utilized (ATP > CTP $\approx$ UTP > GTP) (Fig. 2C). These results unveil differences in the enzyme processivity rather than specificity. Notably, the length of tails

synthesized in the presence of ATP + CTP + GTP + UTP (AGCU) combinations was comparable to those generated in the presence of individual CTP or UTP (Fig. 2C), suggesting that AnCutA may indeed have an unusual substrate specificity, which would agree with in vivo data (Morozov et al. 2010). Importantly, the AnCutA^{D203A/D205A} catalytic mutant (DADA variant) was inactive in vitro, proving that the observed activity was indeed attributable to the analyzed protein and demonstrating the importance of the conserved aspartic acid residues (Fig. 2C). Finally, the enzyme displayed a preference toward RNA substrates with terminal A or U, while it had low activity for terminal C (Fig. 3). The activity toward

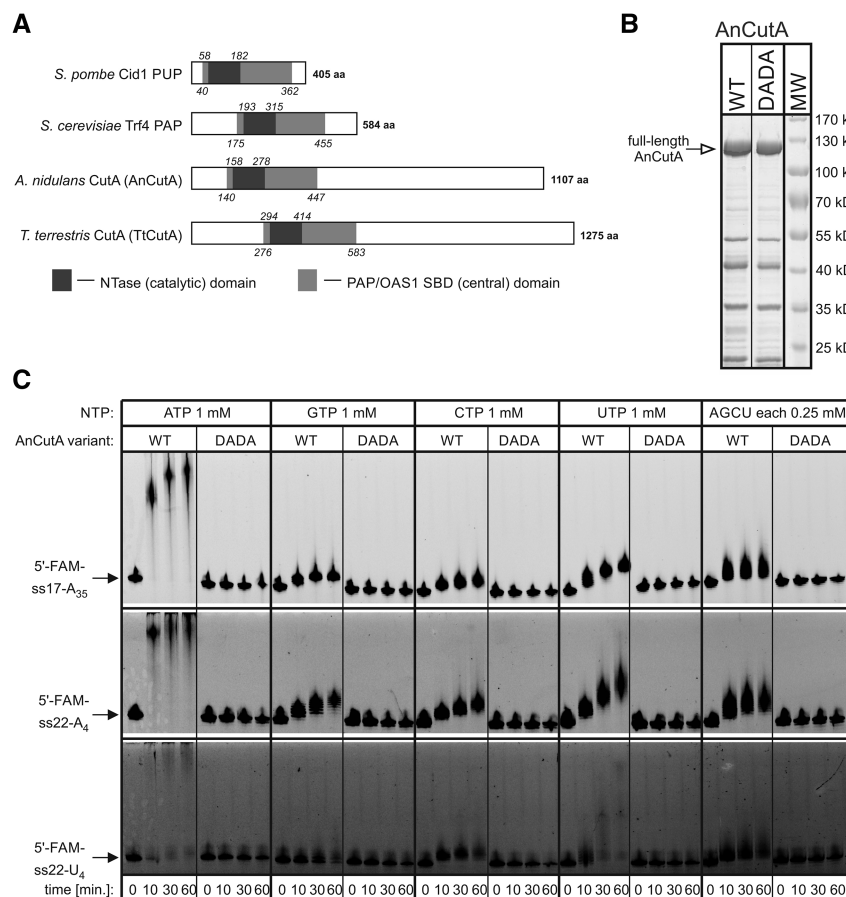


FIGURE 2. AnCutA displays a nucleotidyltransferase activity, dependent on the intact aspartates in the active site. (A) Domain organization of different fungal noncanonical nucleotidyltransferases: *S. pombe* Cid1 poly(U) polymerase, *S. cerevisiae* Trf4 poly(A) polymerase, and CutA proteins from *A. nidulans* and *T. terrestris* (AnCutA and TtCutA, respectively). All proteins contain Ntase domain (dark gray rectangle), which is inserted into PAP/OAS1 SBD (light gray rectangle), formed by protein fragments separated in the primary sequence, but located in proximity in the three-dimensional structure. Italicized numbering refers to amino acids located at the borders of Ntase (above) and PAP/OAS1 SBD (below) domains. (B) SDS-PAGE analysis of the recombinant full-length AnCutA WT and DADA mutant (with aspartates at positions 203 and 205 substituted with alanines) variants purified by nickel-column affinity chromatography. Protein preparations were analyzed together with PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) as a molecular weight marker (lane MW). Masses of individual bands are indicated on the right. Position of band corresponding to the full-length protein is marked with an open arrow. (C) Analysis of nucleotidyltransferase activity, performed for AnCutA WT and DADA variants with three different RNA oligonucleotides, labeled with fluorescein amidite (FAM) at the 5'-end. Reactions were programmed with individual NTPs or a mixture, as indicated at the top, and terminated after the time points indicated below the lanes.

substrate with terminal G could not be analyzed, because such RNA precipitated in the gel wells. Therefore, only substrates terminating with adenosine or uridine stretches were utilized in subsequent experiments.

Identification of the enzymatically active core of a fungal CutA enzyme

Recombinant full-length AnCutA and TtCutA proteins were prone to degradation and their enzymatic activity successively decreased after consecutive stages of purification (Figs. 2B,

4A and data not shown). Thus, in further experiments, we aimed at using more stable, highly active enzyme fragments. To this end, TtCutA was subjected to limited proteolysis (Fig. 4). Protein obtained after nickel affinity chromatography (Fig. 4A) was digested using different concentrations of trypsin, and the samples were analyzed in SDS-PAGE (Fig. 4B). We observed that the full-length TtCutA (lane ND) got degraded to ~38- to 44-kDa-long fragments (lanes B1–B7), of which the longest, relatively stable one was present in the lane B3 (Fig. 4B). Since all samples collected after limited proteolysis comprised some common contaminant migrating at ~60 kDa, we further separated proteins present in the sample B3 by gel filtration (Fig. 4C). The band corresponding to the TtCutA proteolytic digestion product was analyzed by Edman degradation, which revealed that the four most N-terminal residues of this fragment have the following sequence: SAYS (Fig. 4C). This allowed us to localize the N-terminal boundary of this fragment to position 216 in the TtCutA amino acid sequence.

In parallel to SDS-PAGE analyses, the nucleotidyltransferase activity of the fractions obtained in the course of limited proteolysis and gel filtration was monitored by *in vitro* assays. Notably, fractions resulting from limited proteolysis retained nucleotidyltransferase activity, but it was profoundly reduced in comparison with the full-length protein, particularly in the case of CTP as an NTP substrate (Fig. 4D,E). This was most likely related to significantly lower protein concentration (cf. lanes B3 and A1 in Fig. 4B and 4A, respectively). The decrease of the enzymatic activity correlated with the increased trypsin concentration and thus, with the reduced length of the proteolytic fragments (Fig. 4E).

Analysis of nucleotidyltransferase activity for fractions obtained after gel filtration revealed that the faster-migrating band (see fractions C4–C6 in Fig. 4C) indeed represented the proteolytic fragment of TtCutA, at least partially retaining its catalytic properties (Fig. 4F), while the slower-migrating protein (see fractions C1–C3 in Fig. 4C) was an unrelated contamination, lacking enzymatic activity (Fig. 4F).

To summarize, a limited proteolysis approach and Edman degradation analysis, in conjunction with biochemical

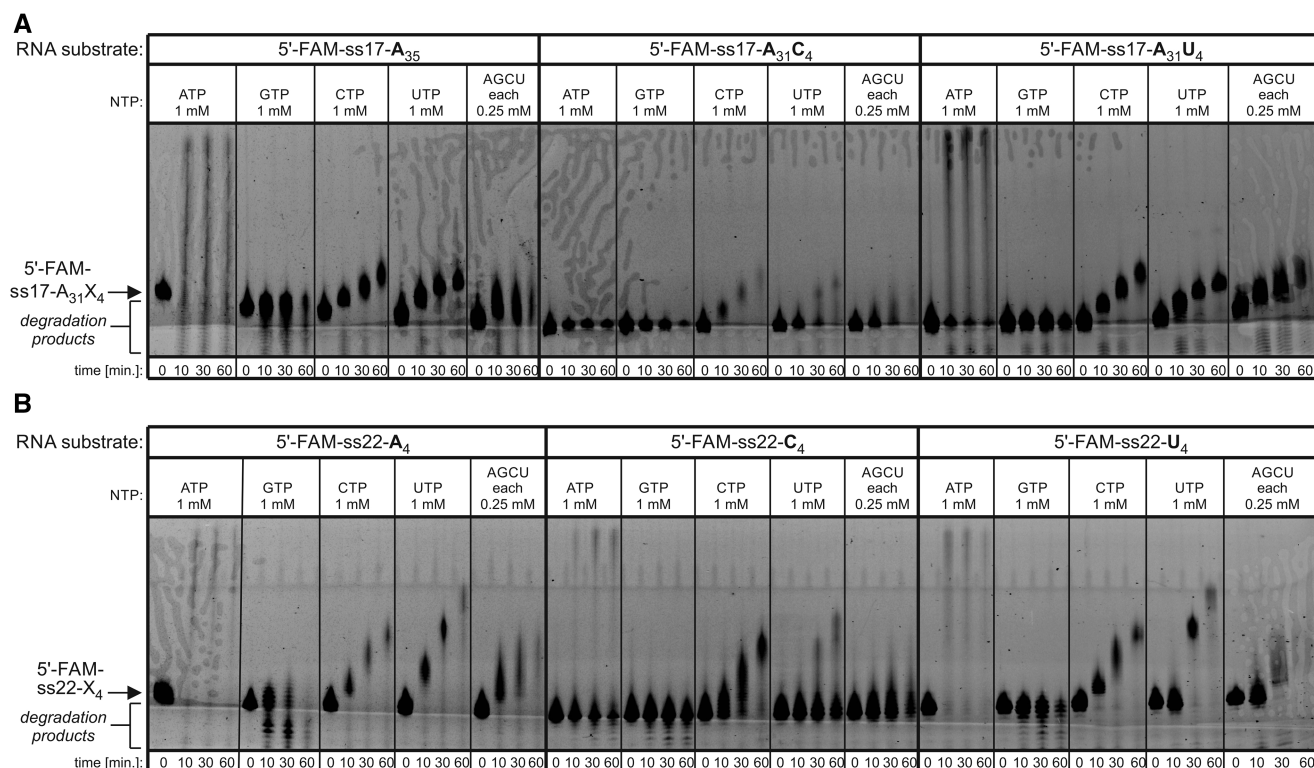


FIGURE 3. AnCutA efficiently extends 3'-ends of RNA substrates terminating with adenosines or uridines, but not cytidines. Nucleotidyltransferase activity assays were performed for the full-length AnCutA WT protein with two sets of synthetic oligoribonucleotides (ss17-A₃₁ or ss22 in panels A and B, respectively), labeled with fluorescein amidite (FAM) at the 5'-end, having 4-nucleotide-long stretches of adenosines, cytidines, or uridines at their 3'-ends. Reactions were programmed with individual NTPs or a mixture and terminated after the time points indicated below the lanes. Position of the substrates is marked with an arrow on the left. Products of unspecific degradation are indicated.

investigation of nucleotidyltransferase activity allowed us to identify a stable, enzymatically active fragment of TtCutA protein.

Based on the results of limited proteolysis and secondary structure predictions, we designed truncated variants of TtCutA and AnCutA: TtCutA²¹⁶⁻⁶¹⁷ and AnCutA⁸⁴⁻⁴⁸⁷, as well as a longer version of TtCutA: TtCutA¹²⁷⁻⁸²⁷, further referred to herein as TtCutA_tr1, AnCutA_tr, and TtCutA_tr3, respectively. All three truncated proteins were successfully purified (Fig. 5A), and biochemical analyses confirmed their ability to synthesize 3'-terminal extensions in the presence of each of the four individual NTPs as well as CU and AGCU mixtures, for each of the RNA substrates tested (Fig. 5B). 3'-Tails synthesized by TtCutA_tr3 were considerably longer than in the case of TtCutA_tr1, irrespective of the NTPs utilized. Furthermore, TtCutA_tr3 was able to generate 3'-terminal tails with individual NTPs in the following order of decreasing efficiency: ATP > CTP > UTP > GTP (Fig. 5B). On the contrary, for TtCutA_tr1, as well as its counterpart from *A. nidulans*—AnCutA_tr—reaction efficiency was comparable in the presence of CTP and UTP (Fig. 5B). This indicates that the lateral extensions present in TtCutA_tr3, when compared to TtCutA_tr1, may influence enzyme processivity. Importantly, there were no significant differences in

the pattern of nucleotidylation between substrates having stretches of either adenosine (ss17-A₃₅, ss22-A₄) or uridine (ss22-U₄) residues of different lengths at their 3'-termini for any of the examined protein variants (Fig. 5B).

Full-length and truncated wild-type CutA proteins display a strong preference toward CTP as an NTP substrate in vitro

In vitro assays for nucleotidyltransferase activity based on incubation of the recombinant proteins in the presence of labeled RNA substrates and individual NTPs, combined with an in-gel analysis of the products may be misleading in terms of the evaluation of the NTP preference. Therefore, to analyze NTP-specificity of the wild-type full-length and truncated AnCutA and TtCutA protein variants in a more reliable way, we devised a 3'-RACE procedure combined with high-throughput sequencing (3'-RACE-seq). Libraries were prepared for products of reactions catalyzed by AnCutA, AnCutA_tr, TtCutA, TtCutA_tr1, and TtCutA_tr3 in the presence of either ss22-A₄ or ss22-U₄ substrate and an equimolar AGCU mixture.

Comparison of the average tail lengths revealed that truncated protein variants usually produced shorter 3'-terminal

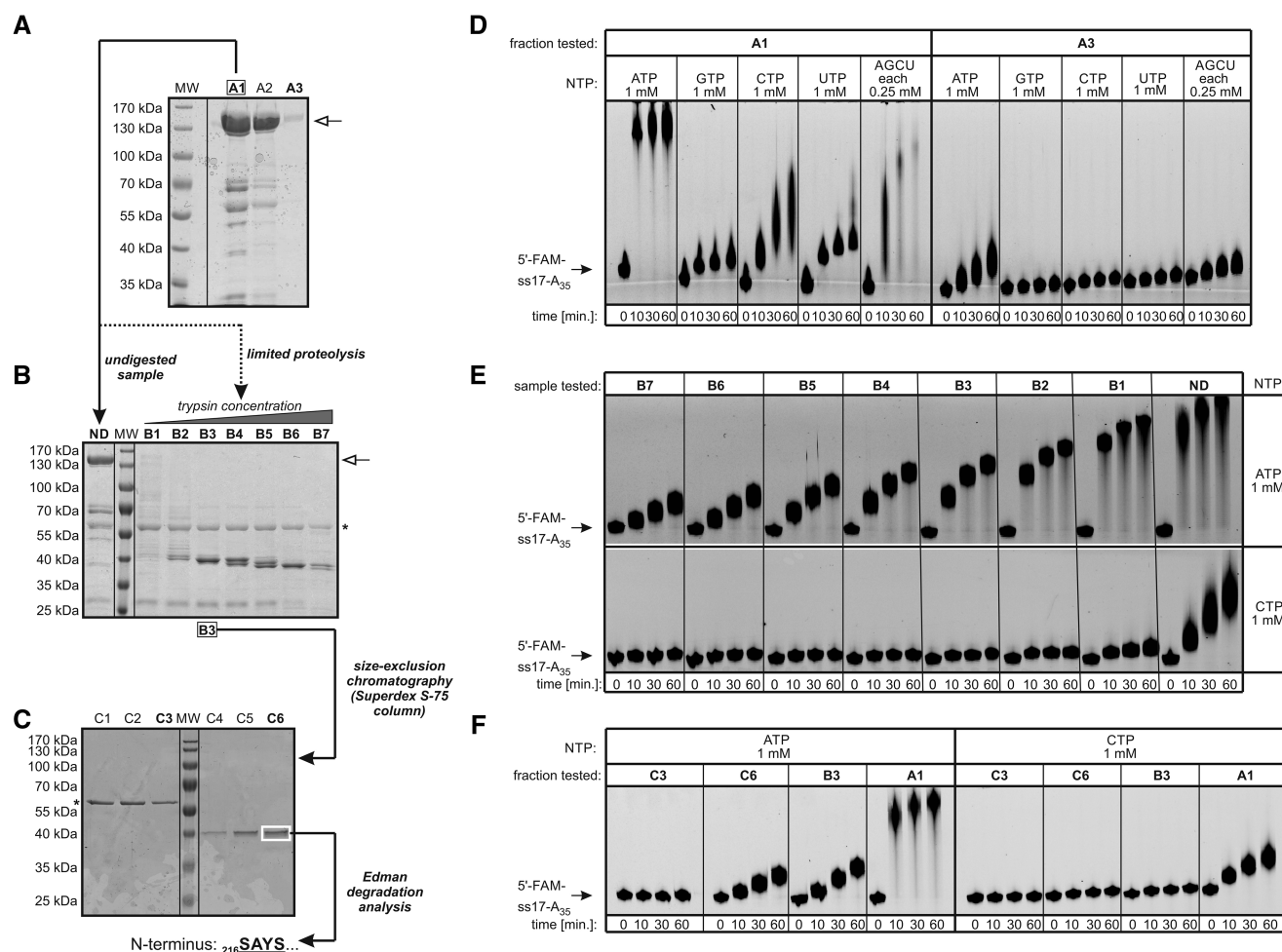


FIGURE 4. Identification of the N terminus of enzymatically active TtCutA proteolytic degradation fragment. (A) SDS-PAGE analysis of protein fractions obtained after purification of the full-length TtCutA by affinity chromatography on the nickel resin. Fraction A1 was subjected to limited proteolysis with trypsin. (B) SDS-PAGE analysis of TtCutA sample, which was not treated with trypsin (lane ND) and samples subjected to proteolytic digestion with different concentrations of trypsin (increasing from lane B1 to lane B7, as indicated above the lanes). Fraction B3 was further purified by gel filtration. (C) SDS-PAGE analysis of protein fractions obtained after separation of the TtCutA limited proteolysis products by gel filtration. Band marked with a white rectangle in lane C6 was subjected to Edman degradation analysis, which demonstrated that the N terminus of the proteolysis-derived fragment begins with S at position 216 of TtCutA, as indicated below the gel. Protein fractions were loaded on SDS-PAGE gels shown in A–C together with PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) as a molecular weight marker (lanes MW). Masses of individual bands are indicated on the left. Positions of bands corresponding to the full-length protein are marked with an open arrow in A and B. Fractions highlighted in bold were tested for nucleotidyltransferase activity in the corresponding assays presented in panels D–F. (D–F) Nucleotidyltransferase activity assays were performed for protein fractions highlighted in bold in panels A–C, using 5'-FAM-ss17-A₃₅ RNA oligo-nucleotide as a substrate (position indicated with an arrow). Reactions were programmed with individual NTPs or a mixture and terminated after the time points indicated below the lanes.

extensions than full-length CutA enzymes, possibly due to the lack of some additional interactions between the synthesized tail and the C-terminal unstructured CutA region, absent in shorter versions (Fig. 6A). Inspection of the percentage distribution of synthesized tails' lengths showed that most of the examined wild-type proteins and their variants synthesized shorter tails for the ss22-U₄ substrate than for the ss22-A₄ substrate (Fig. 6A). This might be the consequence of the fact that the frequency of cytidine addition to the ss22-U₄ substrate is high, and the resulting tail, ending with uridines and cytidines, interacts most probably very weakly with the substrate binding site (see below).

Importantly, for virtually all tested AnCutA and TtCutA protein variants, cytidine was the predominant incorporated nucleotide, usually followed by adenosine, while the incorporation of uridine and—particularly—guanosine, was significantly less frequent (Fig. 6B,C).

We also realized that the nucleotide incorporation is non-random. Thus, to analyze the distribution of nucleotides in the added tails, we have aligned all the sequences starting from the 3'-end. Most importantly, almost all tails synthesized by AnCutA and TtCutA ended with at least two cytidines (Fig. 6B,C), which seems to be a unique feature of CutA enzymes. Furthermore, we observed that the addition

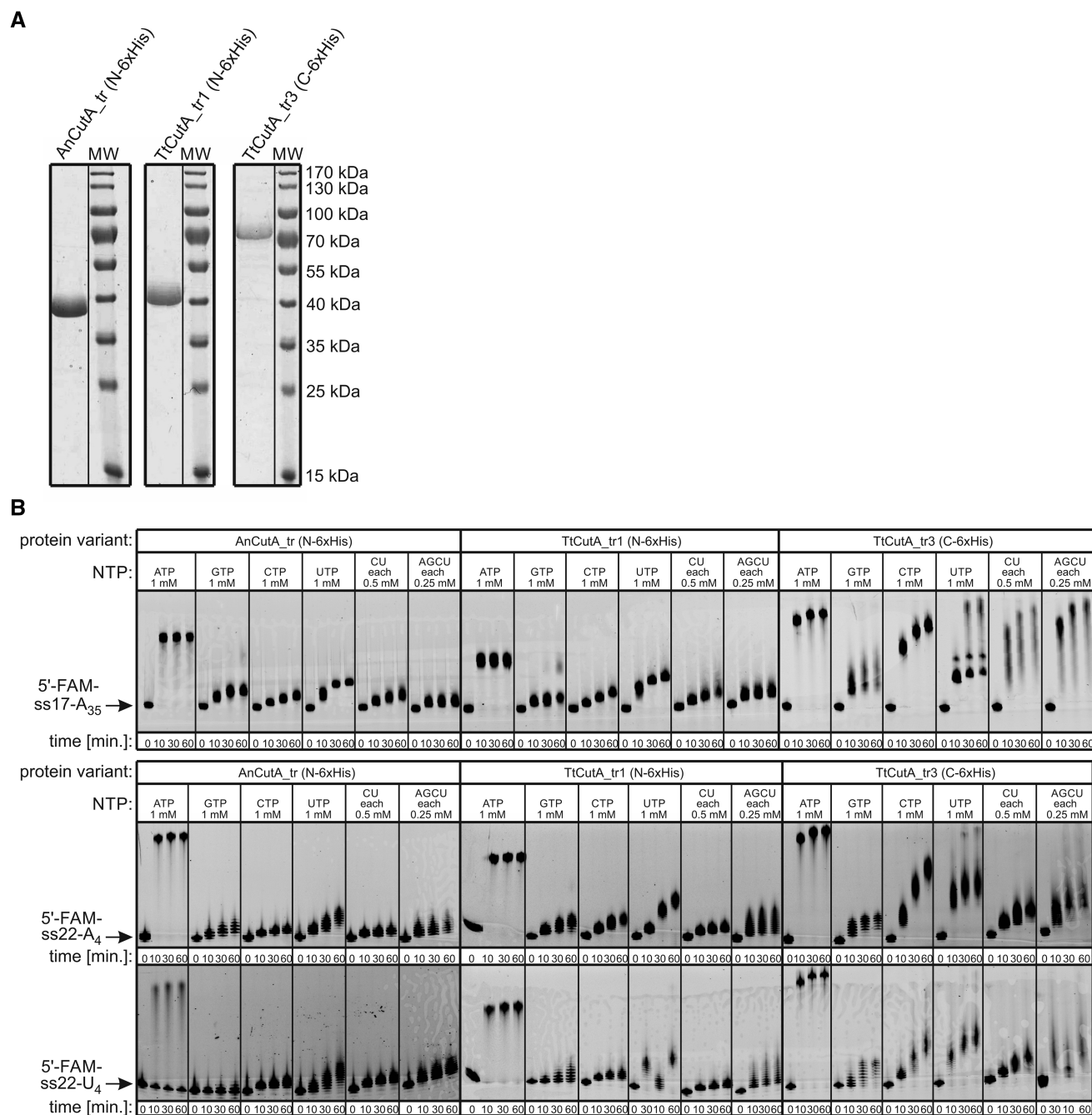


FIGURE 5. Truncated AnCutA and TtCutA variants are enzymatically active. (A) SDS-PAGE analysis of the recombinant (wild-type) AnCutA_{tr}, TtCutA_{tr1}, and TtCutA_{tr3} variants, purified by nickel-column affinity chromatography followed by gel filtration. Protein preparations were analyzed together with PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) as a molecular weight marker (lanes MW). Masses of individual bands are indicated on the right. (B) Analysis of nucleotidyltransferase activity, performed for truncated CutA protein variants from panel A with three different RNA oligonucleotides, labeled with FAM at the 5'-end, as substrates (positions indicated with arrows on the left). Reactions were programmed with individual NTPs or mixtures (CU or AGCU), as indicated at the top, and terminated after the time points indicated below the lanes.

of the particular nucleotide strongly depends on the previously added nucleotide. For instance, AnCutA and TtCutA apparently incorporated adenosine more frequently when the antecedent nucleotide of the substrate was also adenosine. For ss22-A₄ oligonucleotide, the percentage of added adenosines is increased in such case to 42% and 59%, compared

to the average frequency of adenosine incorporation, which is 31% and 48% for AnCutA and TtCutA, respectively (Supplemental Table S1). The situation was different when the last nucleotide of the primer was cytidine or uridine. In such an instance, for the same RNA substrate, we observed: (i) decrease of the adenosine incorporation, (ii) increase of

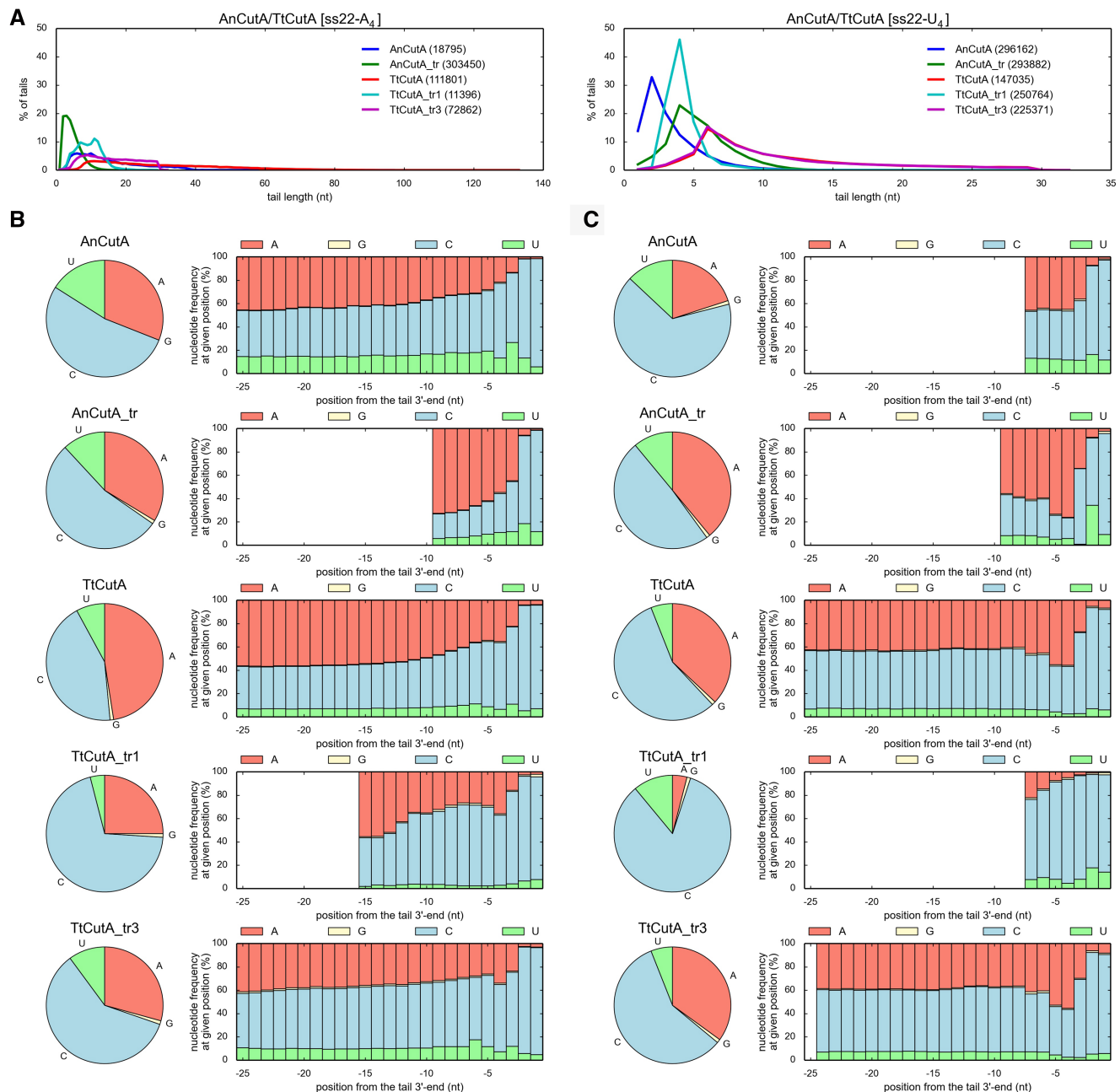


FIGURE 6. Full-length and truncated recombinant AnCutA and TtCutA proteins synthesize 3'-terminal tails rich in cytidines, most frequently ending with two C residues. (A) Length distribution for tails added by different AnCutA/TtCutA variants to ss22-A₄ (left) and ss22-U₄ (right) oligoribonucleotide substrates. Numbers of analyzed sequences, obtained after 3'-RACE-seq high-throughput analysis for each protein variant, are indicated in parentheses. (B,C) Analysis of nucleotide composition for tails added by different AnCutA/TtCutA variants to ss22-A₄ (panel B) and ss22-U₄ (panel C) RNA substrates. Pie charts demonstrate an overall frequency of incorporation of individual nucleotides into the 3'-terminal extensions. Histograms show the percentage of nucleotide types incorporated at the given position of the tail (the ultimate, 3'-terminal position is on the right). Only those positions, which were covered by at least 5% of the 3'-RACE-seq reads, are presented in the histograms. Furthermore, a limit of 25 nucleotides (nt) from the 3'-end of the tail was arbitrarily set for simplicity.

the cytidine incorporation (from 53% or 44% on average to 77% or 76% for AnCutA or TtCutA, respectively) when the preceding nucleotide was uridine, and (iii) increase of the uridine incorporation (from 16% and 8% on average to 22% and 11% for AnCutA and TtCutA, respectively) when

the previous nucleotide of the substrate was cytidine (Supplemental Table S1). It is worth noting that the described effect was slightly enhanced when the primer ended with the CU doublet (increasing the percentage of added cytidines to 79% and 78% for AnCutA and TtCutA,

respectively) or UC (increasing the percentage of added uridines to 14% for TtCutA) (Supplemental Table S1). This is important in the context of findings reported by Morozov et al. (2010), who suggested that AnCutA is a CUCU-adding enzyme that participates in mRNA modification, which promotes decapping and transcript degradation in *A. nidulans*. The results of our analyses are partially consistent with their observation, as we noticed that the presence of uridine at the 3'-end of the RNA substrate enhances the possibility of cytidine incorporation, and vice versa, by CutA proteins. Consequently, it is evident that nucleotide selectivity of these enzymes depends on preceding added nucleotides and is determined by protein–substrate interactions. It must be pointed out, however, that AnCutA and TtCutA protein variants do not terminate the nucleotides incorporation process after adding a single (CU) or double (CUCU) nucleotide doublet, in contrast to CCA-adding enzymes, which add a single (CCA) or double (CCACCA) nucleotide triplet to the 3'-end of tRNAs (Kuhn et al. 2015).

To sum up, high-throughput 3'-RACE-seq analysis unambiguously demonstrated that CutA enzymes from two fungal species display strong preference toward CTP as an NTP substrate in vitro, which has not been observed for noncanonical nucleotidyltransferases studied so far.

3D model of *T. terrestris* CutA proteins reveals the location of the potential NTP specificity determinant region

We next asked what could be the structural basis for the observed in vitro CutA enzymes' preference toward CTP. To answer this question, we generated a 3D homology model of TtCutA and compared it to the experimentally solved structures of closely related enzymes: wild-type and H336N mutant of poly(U) polymerase Cid1 from *S. pombe* (Lunde et al. 2012), and noncanonical poly(A) polymerase Trf4 from *S. cerevisiae* (Hamill et al. 2010) (Fig. 7; see Fig. 1 for alignment of amino acid sequences of these NTases).

In addition, we investigated our 3D model of TtCutA in complex with RNA [oligo(A)], which was taken from

S. cerevisiae canonical poly(A) polymerase (pdb|2q66) (Balbo and Bohm 2007) after superposition of the structures. We noticed that the analyzed CutA variants might interact mainly with the last three nucleotides of the RNA substrate. The interaction of the third nucleotide with the protein is the weakest, while the second nucleotide is most probably interacting mainly with the C-terminal, unstructured extension of the enzyme. As both AnCutA and TtCutA modify mRNA in vivo, they are evolutionarily designed to bind to RNAs ending with adenosines. Consequently, substrates with two pyrimidines at the 3'-end probably could not be strongly bound in their active sites. This observation, combined with the ability of these proteins to incorporate their preferred nucleotide—cytidine, may fully explain the phenomenon that the final product is almost always terminated with two cytidines (see above).

The Cid1 PUP incorporates mostly uridines, while Trf4 and H336N mutants of Cid1 are effective in adding poly(A) tails. TtCutA, similarly to Cid1 and Trf4, possesses the two following domains: a PAP/OAS1 SBD with an inserted NTase domain right after the first core PAP/OAS1 SBD α -helix (Figs. 1, 2A). All these proteins share highly conserved NTase domain motifs ([DE]h[DE]h, h[DE]h, and hG[GS]) (Kuchta et al. 2009) and, in addition, they also retain three residues critical for NTP binding (T404, K425, Y445 in TtCutA), which are conserved in proteins possessing both NTase domain and PAP/OAS1 SBD (Fig. 1; Kuchta et al. 2016). The PAP/OAS1 SBD specifically binds a nucleobase of the incoming NTP mainly by amino acids that provide Watson–Crick hydrogen bonds, either directly or indirectly (via water molecules). Those amino acids are located mostly in the NRM region, between the fourth and fifth core α -helices of PAP/OAS1 SBD (Kuchta et al. 2016). In Cid1, residues from this region, which are responsible for UTP selectivity (D330, E333, and H336; see Fig. 1), cluster around the N3 amine group of uridine and interact with uridine directly or indirectly via one or more water molecules (Fig. 7; Yates et al. 2012, 2015). Another residue, N171 (Fig. 1), binds the uridine O2 carbonyl group, while O4 carbonyl group forms a water-mediated contact to S210 (Figs. 1, 7).

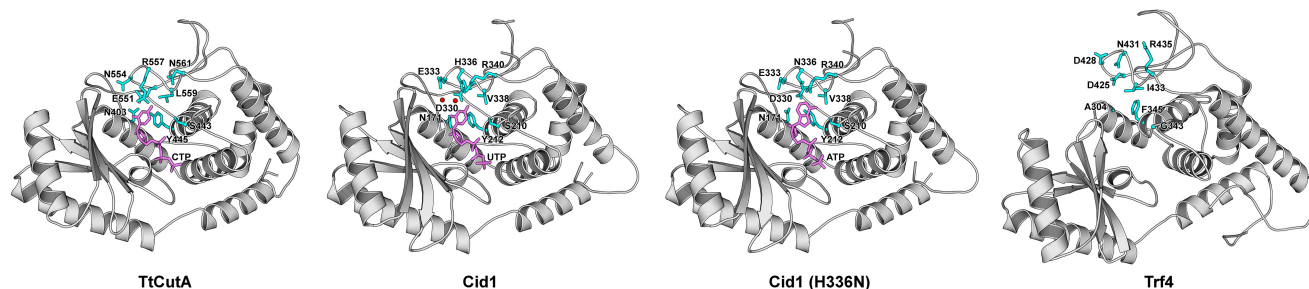


FIGURE 7. Comparison of the 3D model of TtCutA with CTP, and available crystal structures of poly(U) polymerase Cid1 with UTP (pdb|4fh5), Cid1 (H336N) mutant with ATP (pdb|4fhx), and noncanonical poly(A) polymerase Trf4 (pdb|3nyb). Key amino acids forming a nucleobase binding pocket are shown in blue. Nucleotides are shown in magenta. Position of CTP in TtCutA was taken from pdb|4fhv. Selected water molecules responsible for interactions between Cid1 and UTP are denoted by red spheres.

In addition, V338 makes a hydrophobic interaction with the nucleobase (Figs. 1, 7; Lunde et al. 2012). The H336N mutant of Cid1 is effective in adding poly(A) tails mainly through the direct interaction of arginine R340 with the N1 position of the incoming adenosine (Figs. 1, 7; Lunde et al. 2012). Position R340 is stabilized by N336, which also forms a hydrogen bond to the side chain of D330 (Figs. 1, 7), probably having a discriminating effect on the uridine selection in the H336N mutant (Lunde et al. 2012). On the contrary, larger H336 in wild-type Cid1 precludes such unique positioning of the R340 side chain, which adopts a different conformation (Fig. 7). It should be noted that the key residues (D330, N336, and R340) present in the active site of the H336N mutant of Cid1 are also conserved in the noncanonical poly(A) polymerase Trf4 from *S. cerevisiae* (D425, N431 and R435) (Figs. 1, 7; Hamill et al. 2010), which also utilizes ATP in the reaction. We thus concluded that the TtCutA NRM region located between the fourth and fifth core α -helices of the PAP/OAS1 SBD (mostly the fragment encompassing amino acids 551–559—see Fig. 1) remains likely the most important NTP-specificity determinant (NSD) of the analyzed enzymes.

Deciphering the mechanism of CutA preference toward CTP through site-directed mutagenesis in the NRM region

Having identified the NSD of CutA enzymes, we focused on verifying whether particular amino acids within NRM indeed play a role in the CTP specificity, and rationally designed mutations in this region of TtCutA.

We noticed that R557, substituted by H336 in Cid1, is unique for both TtCutA and AnCutA (Fig. 1). It allowed us to speculate that changing this residue may strongly influence substrate specificity. We thus constructed the R557A mutant (PAP1) following the H336A mutation in Cid1 proposed by Yates and coworkers (Yates et al. 2012, 2015) to strengthen the production of poly(A) tails. We also proposed R557Q (PAP2) mutation following the study of Lunde and coworkers, who demonstrated that the Cid1^{H336N} mutant catalyzes robust adenosine incorporation (Lunde et al. 2012). We chose the mutation R557Q over R557N, because the Q557 side chain in the PAP2 mutant could probably be better stabilized by longer E551, compared to a shorter D330 in the H336N mutant of Cid1. We also wondered how to enhance the poly(A) polymerase activity by additional changes in the NRM, apart from substituting a R557 residue. Therefore, we created a PAP3 mutant (E551D/N554E/R557A), in which we reconstructed other amino acids present in the active site of Cid1^{H336A}, which has proven PAP activity (Yates et al. 2012). Additionally, we constructed another mutant—PAP4 (E551T/R557A/L559S), with the aim of enlarging the TtCutA active site and eliminating some (in)direct interactions between the protein and its preferred incoming nucleotide—CTP. The construction of PAP4 variant was guided by

the idea that a canonical poly(A) polymerase (e.g., PAP from *S. cerevisiae*) binds the incoming NTP mainly through non-polar interactions above and below the plane of the adenosine base and to the sugar, and water mediated interactions to polar groups in the plane of the base, and to the sugar hydroxyl groups (Balbo and Bohm 2007).

We also proposed three mutants: PUP1 (R557H), PUP2 (E551D/R557H), and PUP3 (E551D/N554E/R557H), which we assumed should allow us to gradually construct a modified nucleobase recognition site in TtCutA, resembling the one observed in the Cid1 poly(U) polymerase.

To experimentally analyze the influence of the designed mutations on the NTP specificity, we expressed and purified TtCutA_{tr3} variants containing appropriate amino acid substitutions within NRM. All purified TtCutA_{tr3} versions still displayed nucleotidyltransferase activity, but based on the biochemical data it was impossible to assess whether the NTP substrate preference was definitely changed upon a given mutation (data not shown). Thus, to analyze the specificity, 3'-RACE-seq for both ss22-A₄ and ss22-U₄ substrates was performed (Figs. 8 and 9). In concordance with previously obtained results, the most predominant nucleotide added by the wild-type TtCutA_{tr3} in the presence of an equimolar AGCU mixture, was cytidine, followed by adenosine and uridine (Figs. 8 and 9). The frequency of guanosine incorporation fluctuated around 0.5%–2%, corroborating biochemical data indicating that it is definitely the least often utilized nucleotide (Figs. 8 and 9).

Importantly, in the case of the TtCutA_{tr3}^{R557A} (PAP1) mutant, the frequency of adenosine incorporation was increased, which was particularly visible for tails added to ss22-A₄ substrate (Fig. 8), indicating that R557A substitution partially converted the enzyme specificity from PCP to PAP. This implied that R557 is an important residue for selection of CTP over other nucleotides, but not as crucial as its Cid1 counterpart (H336) for the UTP selectivity of the latter (Yates et al. 2012). The R557A mutation weakened the CTP binding affinity, probably because of destroying one of the key contacts between the protein and this preferred nucleotide. Additionally, replacing arginine with much smaller alanine enlarged the TtCutA nucleobase binding site. As a result, the R557A mutation, by diminishing the cytidine-adding activity of the enzyme, increased the presence of adenosines in synthesized tails, as we expected.

The influence of PAP2, PAP3, and PAP4 mutations on the NTP specificity of TtCutA_{tr3} was less obvious (Fig. 8). Similar to other mutant protein variants, the number of cytidines built into the 3'-terminal tails by TtCutA_{tr3}^{R557Q} (PAP2), TtCutA_{tr3}^{E551D/N554E/R557A} (PAP3), and TtCutA_{tr3}^{E551T/R557A/L559S} (PAP4) was reduced in comparison to the wild-type enzyme (Fig. 8). The percentage of adenosines was also lower for these protein variants than in the wild-type control—such an effect was most remarkable for the PAP3 triple mutant (Fig. 8). On the other hand, the parallel increase in the frequency of uridine incorporation,

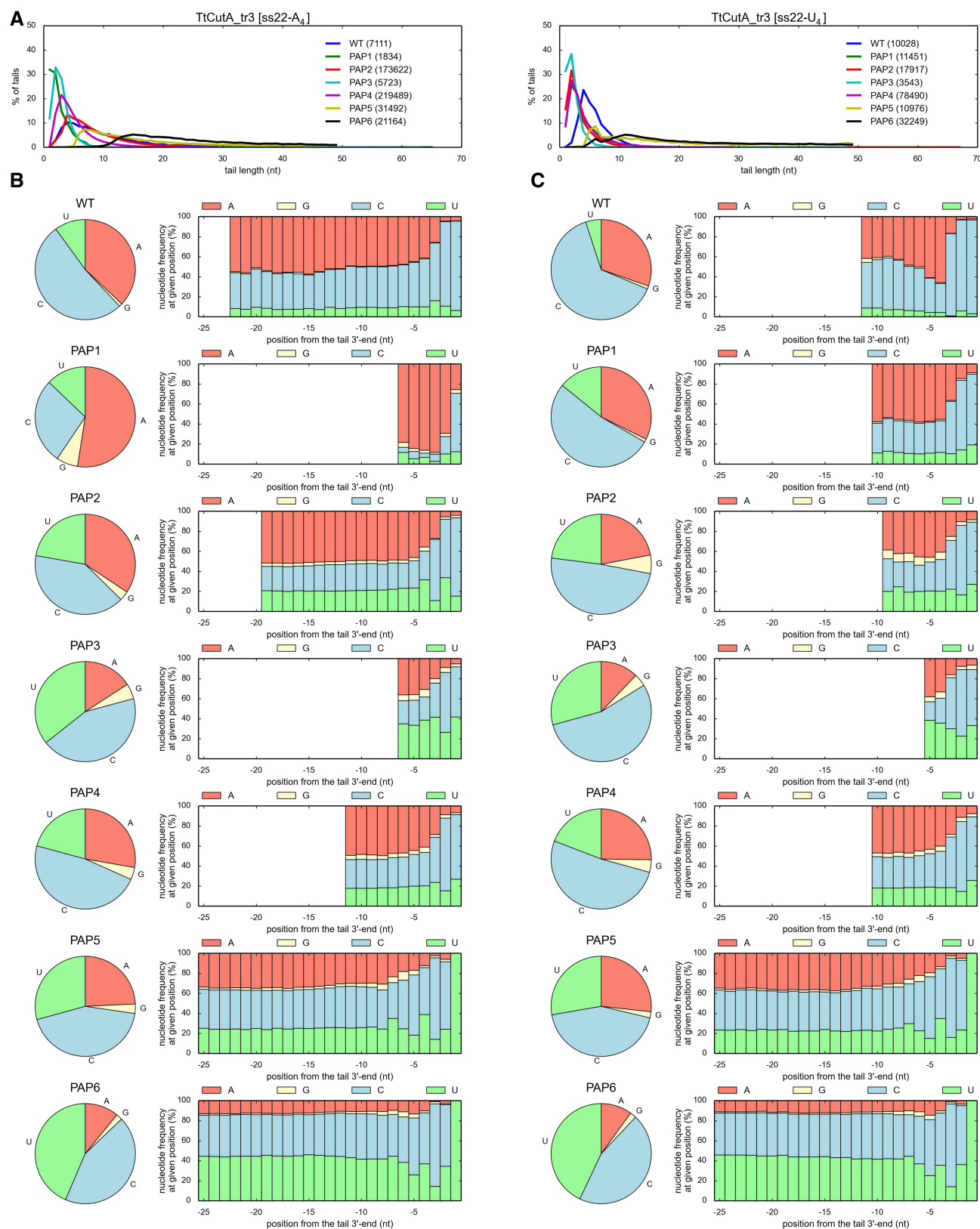


FIGURE 8. Effect of mutations designed to change TtCutA specificity into PAP. (A) Length distribution for tails added by wild-type (WT) and mutant (PAP1-PAP6) TtCutA_tr3 variants to ss22-A₄ (left) and ss22-U₄ (right) oligoribonucleotide substrates. Numbers of analyzed sequences, obtained after 3'-RACE-seq high-throughput analysis for each protein variant, are indicated in parentheses. (B,C) Analysis of nucleotide composition for tails added by different TtCutA_tr3 variants to ss22-A₄ (panel B) and ss22-U₄ (panel C) RNA substrates. Results are presented as in Figure 6.

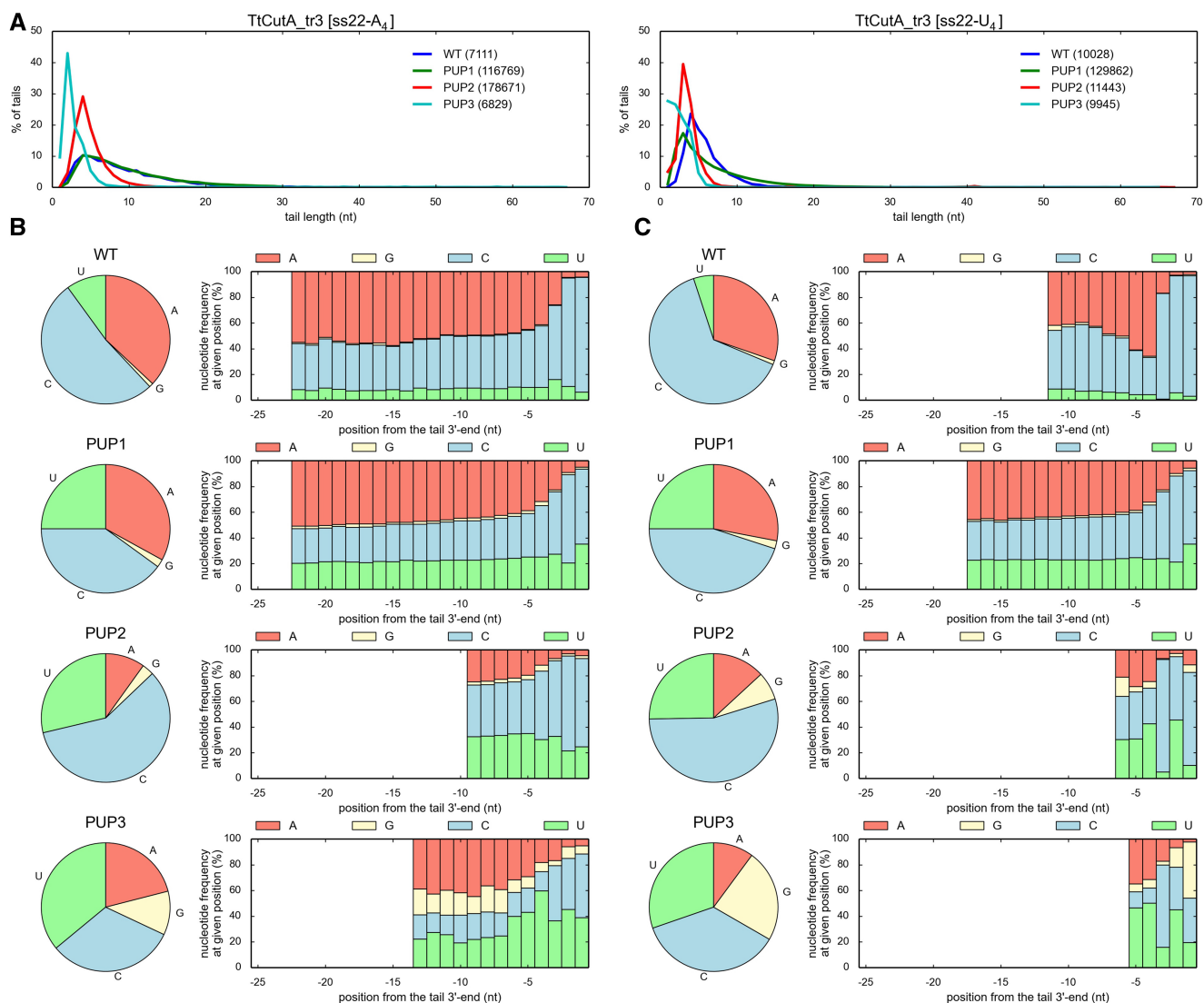


FIGURE 9. Effect of mutations designed to change TtCutA specificity into PUP. (A) Length distribution for tails added by wild-type (WT) and mutant (PUP1-PUP3) TtCutA_{tr3} variants to ss22-A₄ (left) and ss22-U₄ (right) oligoribonucleotide substrates. Numbers of analyzed sequences, obtained after 3'-RACE-seq high-throughput analysis for each protein variant, are indicated in parentheses. (B,C) Analysis of nucleotide composition for tails added by different TtCutA_{tr3} variants to ss22-A₄ (panel B) and ss22-U₄ (panel C) RNA substrates. Results are presented as in Figure 6.

observed for all these protein variants (again—most clearly in the case of PAP3 mutant) (Fig. 8), suggested that R557Q single mutation and E551D/N554E/R557A or E551T/R557A/L559S triple substitutions change NTP preference more toward UTP than ATP. It thus seems that PAP2, PAP3, and PAP4 mutants gained additional interactions with uridine, in the case of PAP2 and PAP3 probably through increase of similarity to Cid1 in the NRM region. Specifically, Q557 in PAP2 corresponds to H336 in Cid1 and thus seems to replace it functionally to perform a similar role in UTP recognition. In addition, we did not observe the increased amount of adenosines in tails synthesized by PAP2 (Fig. 8), probably because TtCutA possesses asparagine (N561) instead of key arginine, which can be found at this position in the non-canonical poly(A) polymerase Trf4 (R435) and Cid1^{H336N}

mutant, specific toward adenosine (R340). On the other hand, according to the 3D modeling, R557 may be a structural equivalent of this key arginine responsible for ATP recognition, which may explain substantial adenosine-adding activity of wild-type TtCutA. Consequently, we introduced the N561R mutation and gradually changed TtCutA to make it more similar to both the Cid1^{H336N} mutant and Trf4, by generation of PAP5 (R557N/N561R) and PAP6 (E551D/R557N/N561R) variants. Surprisingly, these mutants had significantly decreased frequency of adenosine incorporation and largely increased specificity toward uridines (particularly PAP6) (Fig. 8). This suggests that in TtCutA—probably due to slight differences in the geometry of the NTP binding site (e.g., NRM conformation)—the role of arginine at position 557 in adenosine recognition cannot

be substituted by moving this arginine to position 561 (where it is present in Trf4 poly(A) polymerase).

When trying to re-engineer poly(U) polymerase from TtCutA, we noticed other apparent changes in the substrate preference in the case of TtCutA_tr3^{R557H} (PUP1), TtCutA_tr3^{E551D/R557H} (PUP2), and TtCutA_tr3^{E551D/N554E/R557H} (PUP3) mutants. First of all, they all synthesized 3'-tails with fewer adenosines than the wild-type counterpart (Fig. 9). At the same time, the percentage of uridines increased considerably (three- to ninefold, depending on the substrate and the length of analyzed tails), which suggested that for these variants, the specificity toward uridines was increased (Fig. 9). Furthermore, the amount of uridines in the 3'-terminal extensions generated by TtCutA_tr3^{E551D/N554E/R557H} was higher than for TtCutA_tr3^{E551D/R557H}, and simultaneously the frequency of cytidine incorporation decreased by ~20%–25% (Fig. 9). This implicated that while E551D and R557H substitutions are partially responsible for PCP→PUP modification of specificity, N554E amino acid change may additionally increase enzyme preference toward UTP. These results collectively indicate that while replacing TtCutA E551 and R557 residues with their Cid1 counterparts (aspartic acid and histidine, respectively) apparently failed to reconstruct all water-mediated contacts between protein and UTP, an additional N554E substitution caused UTP to be the preferred nucleotide. This also implies that N554, together with R557 is a very important residue responsible for CTP selectivity of TtCutA.

In sum, results of our high-throughput 3'-RACE-seq analysis, combined with structural inspection, demonstrated the critical importance of arginine (R), glutamic acid (E), and asparagine (N) at positions 557, 551, and 554 of TtCutA (Fig. 1), located within NRM, for CTP-specificity of the examined enzyme. Furthermore, R557A substitution enabled partial conversion of TtCutA nucleotidyltransferase activity from PCP to PAP, whereas R557H, E551D, and N554E substitutions changed the enzyme specificity from PCP to PUP. Notably, however, while we successfully engineered proteins with significantly reduced or increased specific nucleotide selectivity, we were not able to construct the enzymes, which would use only one preferred nucleotide for the polymerization process. This may be caused by the character of the CutA enzymes' NRM region, which allows binding of all NTPs in the active site without a stringent preference for a single nucleotide.

Nucleotide specificity of TtCutA variants is preserved at a physiologically relevant NTP ratio

Since the 3'-RACE-seq analyses described above were performed for products of reactions with equimolar AGCU mixtures, the increased frequency of cytidine incorporation in vitro by TtCutA wild-type protein could potentially result from artificially high CTP concentration. To test this possibility, assays for TtCutA_tr3 WT, as well as for PAP1 and PUP3

variants, which displayed the most obvious changes in NTP specificity, were repeated at more physiologically accurate 10:1:1:1 ratio of ATP:GTP:CTP:UTP, assuming that ATP is in approximately 10-fold excess over other NTPs in vivo.

Analysis of the 3'-RACE-seq data derived from assays conducted in such modified conditions (Fig. 10) revealed for the TtCutA_tr3 wild-type version that although the overall percentage of adenosines increased to 60%–70% (from 30% to 40% for equimolar NTP ratio), depending on the substrate, the level of cytidine incorporation decreased at the same time from 50%–60% to 30%–40% (cf. Fig. 10B,C with Fig. 8B,C). This demonstrates that even when CTP is underrepresented, the enzyme retains high specificity toward this nucleotide. Furthermore, decreasing CTP concentration did not preclude predominant incorporation of cytidine at the 3'-terminal and penultimate positions of the extension (Fig. 10B,C and data not shown). In addition, we noticed that more than 99%, ~80%, and ~15% of tails contained at least one, two, and three cytidines, respectively (Fig. 11).

We then checked whether changes in the NTP specificity that we observed for TtCutA_tr3 PAP1 and PUP3 mutant variants could be due to the excess of ATP. Surprisingly, the frequency of overall ATP incorporation increased very modestly for PAP1 variant, when compared to WT (Fig. 10B,C), but this could be partly due to a shorter average length of the tails generated by the PAP1 mutant relative to the wild-type protein (Fig. 10A and data not shown). Nevertheless, the R557A mutation present in TtCutA_tr3 PAP1 variant led to other apparent changes, e.g., the increased presence of adenosine at two ultimate tail positions (more tails contained AA, AC, or CA 3'-terminal dinucleotides) (Fig. 10B,C and data not shown), appearance of a fraction of tails (~10%) lacking cytidine, and concomitant reduction in the percentage of tails containing at least two (~20%) consecutive cytidines (Fig. 11). This confirms that the partial change of specificity from CTP toward ATP has been achieved in the case of the TtCutA_tr3 PAP1 variant.

More pronounced changes were noted at the physiologically relevant NTP ratio for the TtCutA_tr3 PUP3 mutant (Fig. 10B,C). First, despite significant ATP excess, uridine incorporation into tails increased to ~30% (compared to ~2% for the WT variant). Second, numerous tails contained CU, UC, UU, UA, and AU dinucleotides at the 3'-end, instead of CC (Fig. 10B,C and data not shown). Finally, fractions of tails lacking cytidines and containing two or more cytidines in a row were more and less abundant, respectively, than for TtCutA_tr3 WT (Fig. 11). At the same time, the percentage of extensions encompassing at least one uridine was as high as ~85%, compared to only 10%–20% for WT and PAP1 versions (Fig. 11). Furthermore, longer (2–5 nt) stretches of uridines were also evident in some tails generated by the PUP3 mutant (Fig. 11 and data not shown), which virtually did not occur in the extensions synthesized by TtCutA_tr3 WT and PAP1.

In summary, the relatively high in vitro preference of TtCutA_tr3 WT toward CTP is maintained even in

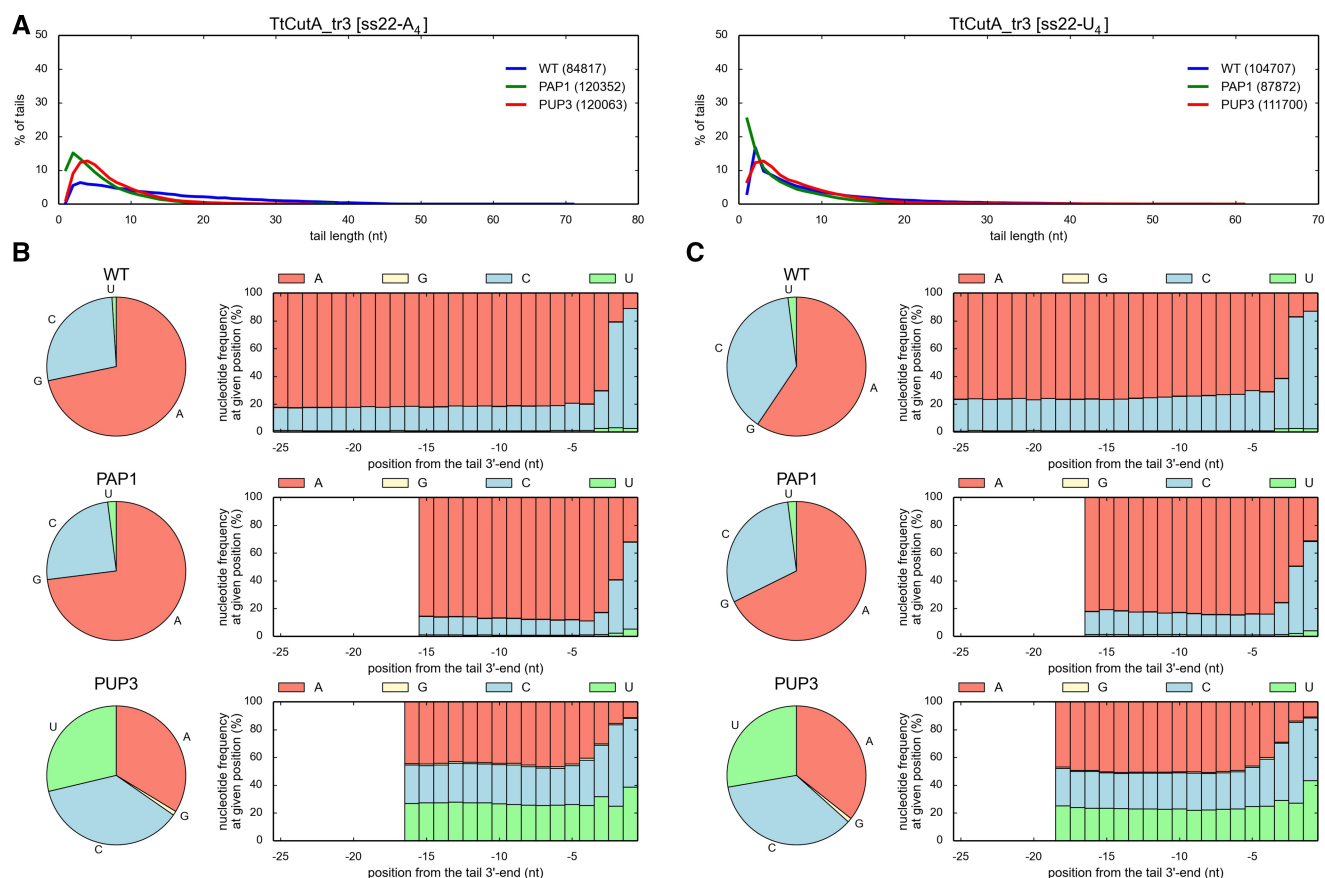


FIGURE 10. Specificity of TtCutA with an excess of ATP over other nucleotides. (A) Length distribution for tails added by wild-type (WT) and mutant (PAP1 and PUP3) TtCutA_{tr3} variants to ss22-A₄ (left) and ss22-U₄ (right) oligoribonucleotide substrates in the presence of 10:1:1:1 ATP:GTP:CTP:UTP molar ratio. Numbers of analyzed sequences, obtained after 3'-RACE-seq high-throughput analysis for each protein variant, are indicated in parentheses. (B,C) Analysis of nucleotide composition for tails added by different TtCutA_{tr3} variants to ss22-A₄ (panel B) and ss22-U₄ (panel C) RNA substrates at physiologically relevant NTP ratio. Results are presented as in Figure 6.

conditions where the availability of this nucleotide is considerably reduced with respect to ATP, which likely reflects the situation in vivo. Moreover, changes of NTP specificity toward adenosine or uridine due to mutations within NRM introduced in PAP1 or PUP3 variant, respectively, could be observed under the same conditions, which further supports our conclusion about the importance of E, N, and R amino acids at TtCutA positions 551, 554, and 557, respectively, for CTP selectivity.

Distinct affinities and variable enzyme processivity in the presence of various NTPs account for differences in the nucleotide composition of extensions generated by different TtCutA variants

We next sought to determine whether differential nucleotide composition of tails synthesized by different variants of TtCutA_{tr3} could result from distinct affinities toward various NTPs. To this end, we first attempted to monitor polymerization efficiency in the presence of individual NTPs by fluorescence-based coupled activity assays, which rely on

indirect measuring of pyrophosphate release with the use of inorganic pyrophosphatase (PPase) and purine nucleoside phosphorylase (PNP), able to cleave a fluorescent 7-methylguanosine derivative in the presence of generated phosphate. Similar approach proved previously to be successful for calculation of kinetic parameters for *S. pombe* Cid1 (Lunde et al. 2012). However, we found it inappropriate in the case of TtCutA, since the reaction conditions optimized for PPase and PNP were totally incompatible with requirements of TtCutA polymerization activity.

Therefore, to roughly compare affinities of TtCutA_{tr3} WT, PAP1, and PUP3 variants toward different NTPs, a set of gel-based assays was performed under conditions where the RNA:protein molar ratio was 40:1 and individual ATP, CTP, or UTP were utilized at various concentrations, ranging from 0.1 to 300 μ M and reactions were stopped after much shorter time-points than in previous biochemical analyses. Results of these experiments are presented in Figure 12.

While basing on data obtained in this manner we were unable to provide quantitative measurements of kinetic constants (which is also due to substantial differences in enzyme

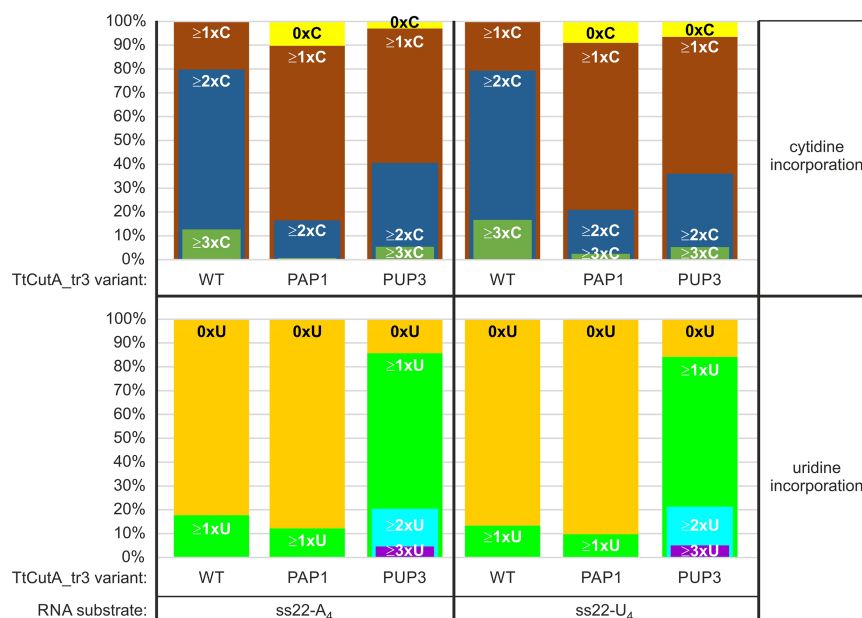


FIGURE 11. Different fractions of tails synthesized by WT, PAP1, and PUP3 variants of TtCutA_tr3 contain a given number of cytidines or uridines. Upper graph shows percentage of tails derived from the 3'-RACE-seq data presented in Figure 10, which entirely lack cytidine (0xC), or contain at least one single cytidine or at least one 2- or 3-nt-long string of consecutive cytidines. Lower graph demonstrates analogous results for uridines.

processivity in the presence of ATP relative to CTP or UTP, as discussed below), some relative comparisons could be made. First, both WT and PAP1 variants displayed a similar affinity toward ATP, starting polymerization at 1 μ M ATP concentration (Fig. 12A). In contrast, the PUP3 mutant was unable to begin ATP addition unless its concentration was higher than 10 μ M. Moreover, the reaction progressed much more slowly and resulting tails were significantly shorter (Fig. 12A). This, together with other observations discussed below, may partially explain why the stretches of adenosines present in tails generated by TtCutA_tr3 PUP3 variant are more often interrupted with other residues, particularly uridine.

In the case of CTP, PUP3 and PAP1 were able to start polymerization at a much higher concentration of this nucleotide (10 and 30 μ M, respectively), when compared to the wild-type TtCutA_tr3 version (Fig. 12B). This indicates that the affinity of PAP1 and PUP3 mutants toward CTP might be indeed somewhat lower, which is reflected by a lower incidence of cytidine in tails generated by these protein variants in comparison to TtCutA_tr3 WT (see Fig. 11). On the contrary, affinity of all three tested TtCutA_tr3 versions toward UTP did not seem to differ dramatically, since all of them started adding uridine already at 1 μ M UTP concentration.

In addition, we noted that when the RNA substrate was in excess over protein, reflecting an *in vivo* situation, all TtCutA_tr3 variants synthesized very short tails with CTP and UTP, when compared to ATP, indicating lower enzyme processivity in the presence of pyrimidines. This difference

was less obvious in previous gel-based assays, where a slight excess of protein over RNA (2.5:1) was applied.

To investigate differences in processivity more systematically, reactions were carried out individually for each of the three TtCutA_tr3 variants with either ATP, CTP, or UTP at a fixed, nonlimiting (100 μ M) concentration in a time-course extended to 60 min, and the products were analyzed in a sequencing gel to obtain single-nucleotide resolution (Fig. 13). Initial analysis revealed that at 1 μ M CTP concentration, TtCutA_tr3 WT was able to add only one cytidine, whereas at higher concentrations of this nucleotide maximum length of the extension achieved only two residues, demonstrating that the enzyme tends to terminate after adding two cytidines (at least by the end of a short reaction time-course) (Fig. 13A). Further experiments demonstrated that all protein variants were processive with ATP and acted distributively in the presence of CTP or UTP (Fig. 13B–D). Processivity

of TtCutA_tr3 PUP3 variant with ATP was a bit lower than that observed for WT or PAP1 versions, but this conclusion may be obscured due to RNA degradation at later time points. On the contrary, both WT and PUP3 variants stopped after adding maximally two cytidines or uridines, whereas in the case of the PAP1 mutant termination occurred even after the addition of one pyrimidine.

These observations corroborate our 3'-RACE-seq data, demonstrating for instance that in the presence of AGCU mixture, TtCutA_tr3 wild-type protein usually processively adds a stretch of adenosines, before switching at some point to cytidine (or sometimes uridine), but is able to distributively add only one or two such residues. Then, depending on the nature of the next incoming nucleotide, the enzyme likely dissociates from the substrate, resulting in the frequent generation of tails terminating with two cytidines, or continues synthesis of longer adenosine stretches if the subsequent nucleotide is ATP. Furthermore, the majority of tails generated by PAP1 variant contain only one cytidine, since addition of the second one seems to be inefficient in this case.

In conclusion, nucleotide composition of tails synthesized by different TtCutA variants is determined by a combination of several interdependent factors, such as NTP specificity determined by the character of the NRM (affinity toward particular NTPs), enzyme processivity and the nature of preceding nucleotide in the RNA substrate. It should be also noted that the analyzed mutations (e.g., substitution of arginine) in the NRM region may also have some effect on the electrostatic environment of the nucleotide binding

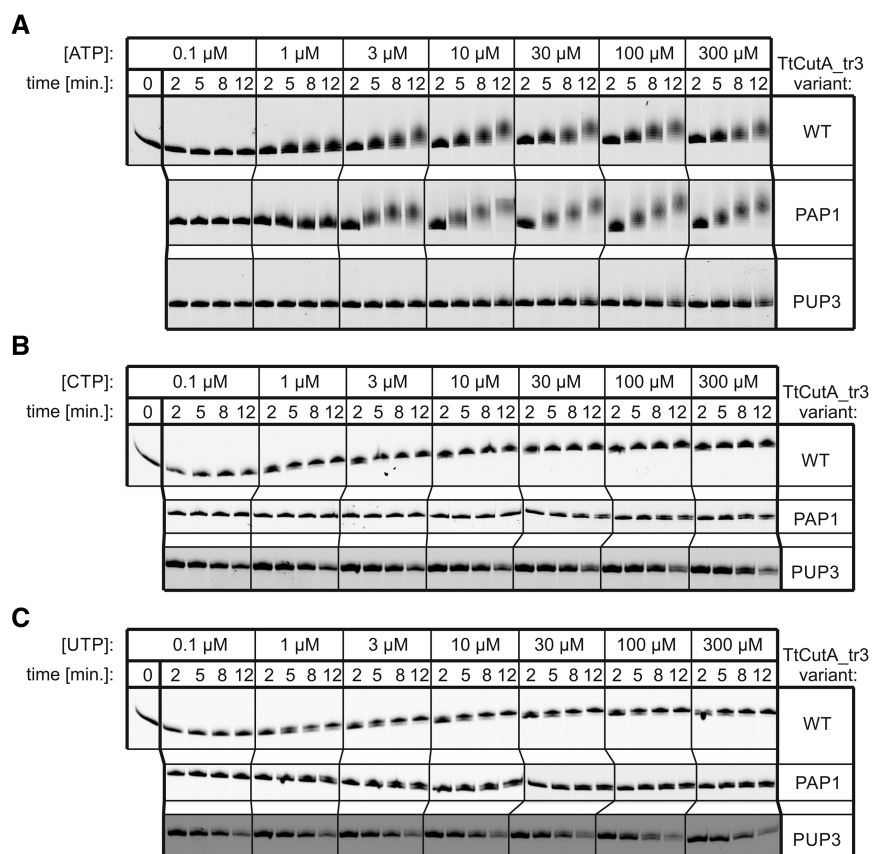


FIGURE 12. Different affinity of TtCutA_tr3 variants toward individual NTPs. (A–C) Nucleotidyltransferase activity assays were performed for WT, PAP1, and PUP3 TtCutA_tr3 protein variants, with 40-fold molar excess of the 5'-FAM-labeled ss22-A₄ RNA substrate. Reactions were programmed with individual ATP (A), CTP (B), or UTP (C) at different concentrations, as indicated above the lanes and terminated after short time points (2–12 min) following start of incubation.

pocket, which could influence interaction with the RNA 3'-end near the active site.

Mutations in the Cid1 NRM region enable its partial conversion from PUP to PCP

To further confirm our conclusions about the importance of particular amino acids within the NRM of CutA enzymes, and to extend our findings to other fungal noncanonical NTases, we attempted to construct the mutant of *S. pombe* Cid1 PUP with increased preference for cytidine. To this end, we expressed in *E. coli* a full-length wild-type Cid1 (Cid1^{WT}), as well as its mutated variant, in which aspartic acid, glutamic acid, and histidine, at positions 330, 333, and 336, equivalent to critical TtCutA positions 551, 554, and 557 (Fig. 1), were changed into arginine, glutamate, and asparagine present in the latter protein, respectively. Such amino acid changes, designed to convert Cid1 PUP into PCP were rationalized not only by 3D structure modeling, but also justified by data obtained for TtCutA, demonstrating that the reciprocal substitutions within its NRM

partially alter NTP specificity of this enzyme from CTP toward UTP (see above). We will further refer to this Cid1 triple mutant as Cid1^{D330E/E333N/H336R} or Cid1^{PCP} variant. Both Cid1^{WT} and Cid1^{D330E/E333N/H336R} proteins were successfully purified following heterologous expression, and biochemical assays confirmed that they have nucleotidyltransferase activity (data not shown).

To verify whether Cid1^{D330E/E333N/H336R} triple mutant indeed acquired specificity toward CTP, reactions for Cid1^{WT} and Cid1^{D330E/E333N/H336R} proteins were carried out for both ss22-A₄ and ss22-U₄ substrates in the presence of equimolar AGCU combination, followed by 3'-RACE-seq. Results of this experiment are summarized in Figure 14. The 3'-terminal tails synthesized by Cid1^{WT} were composed almost exclusively of uridines (specificity >97%), as expected. On the contrary, Cid1^{D330E/E333N/H336R} variant generated 3' extensions containing only ~50% of uridines. In parallel, incorporation of remaining nucleotides increased significantly, in the following order: C > A > G. Importantly, fraction of cytidines built into the tails achieved 30%–40%, indicating that the mutations that we designed indeed significantly increased cytidine and decreased uridine incorporation by Cid1 enzyme.

In conclusion, investigation of Cid1 specificity allowed us to complement data obtained for TtCutA, demonstrating that the ExxNxxR and DxxExxH clusters present at the equivalent positions of NRM, are primarily responsible for the preference of noncanonical NTases toward CTP or UTP, respectively. Unfortunately, we were unable (via proposed mutations) to convert TtCutA into highly effective poly(A) or poly(U) polymerase, as we probably failed to reconstruct all Watson–Crick hydrogen bonds between protein and the desired nucleotide. It seems that TtCutA has a unique geometry and environment of the nucleotide binding site, which makes it impossible to achieve specificity toward a single nucleotide by only a few mutations.

DISCUSSION

It has been more than 10 years from the discovery of noncanonical PAPs (Saitoh et al. 2002; Wang et al. 2002). Multiple follow-up studies have demonstrated that several enzymes from this group synthesize polyuridine, rather than polyadenosine tails, which classified them as PUPs or TUTases (Kwak and Wickens 2007; Rissland et al. 2007). Since then,

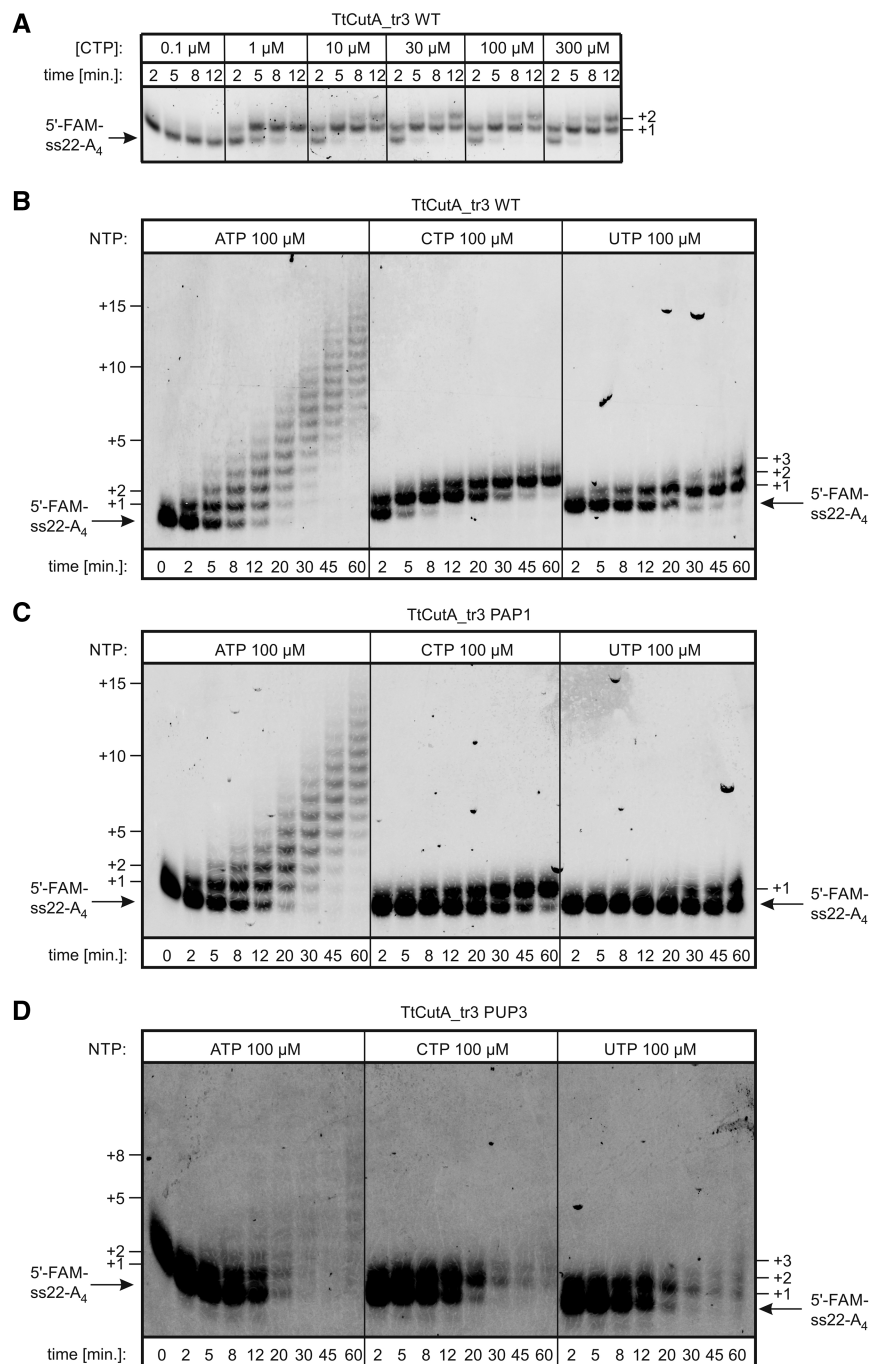


FIGURE 13. TtCutA_tr3 variants act processively with ATP and distributively in the presence of CTP or UTP. (A) Products of reactions shown in Figure 12B for TtCutA_tr3 WT protein were resolved in 10% high-resolution sequencing gel, showing that the substrate 3'-end can be extended by not more than two cytidines, as indicated on the right (B–D). Nucleotidyltransferase activity assays were performed for WT (B), PAP1 (C), and PUP3 (D) TtCutA_tr3 protein variants, with 40-fold molar excess of the 5'-FAM-labeled ss22-A₄ RNA substrate (position marked with arrows). Reactions were programmed with 100 μM ATP, CTP, or UTP, as indicated above the lanes and terminated after 2–60 min following start of incubation. Lengths of the synthesized tails are indicated on both sides of the gel images.

cytoplasmic uridylation was demonstrated to play a crucial role in the maintenance of cellular homeostasis by regulating stability of both coding and various noncoding transcripts in

virtually all model organisms analyzed (Scott and Norbury 2013; Munoz-Tello et al. 2015; Viegas et al. 2015; Scheer et al. 2016). *S. pombe* Cid1 has been the most extensively investigated PUP, for which several crystal structures were solved (both apo- and NTP-bound) (Lunde et al. 2012; Munoz-Tello et al. 2012, 2014; Yates et al. 2012, 2015), which laid the ground for the attempts to explain differences in the substrate specificity between noncanonical PAPs and PUPs. What made the situation even more complicated was the discovery that Cid1 homolog in *A. nidulans*—CutA—does not act as a PUP in vivo, but rather adds C/U-rich tails to the 3'-ends of mRNAs in this fungal species (Morozov et al. 2010). Subsequent global transcriptomic analyses demonstrated that untemplated incorporation of cytidines (and also guanosines) into the 3'-terminal extensions is not an uncommon event (Chang et al. 2014; Zuber et al. 2016); however, the noncanonical NTases responsible for such modifications occurring in mammalian or plant cells have not been identified yet. Nevertheless, these findings unequivocally show that the repertoire of substrate specificity among this class of enzymes is not limited to ATP/UTP and some of the group members display distinct NTP preferences. This warrants further investigations of the molecular basis for the observed variability in the substrate specificities of noncanonical NTases.

Previous structural inspections, confirmed by biochemical studies, revealed that H336 in the Cid1 NRM is a major determinant of its specificity toward UTP (Lunde et al. 2012; Munoz-Tello et al. 2012; Yates et al. 2012). UTP recognition by Cid1 is supported by interactions involving other amino acids from the NRM, particularly D330 and E333, which together with H336 constitute DxxExxH cluster (Table 1). Importantly, very similar clusters are present in the human TUTases ZCCHC6 and ZCCHC11 and proteins from other species, for which PUP activity has been confirmed (Ren et al. 2012; Zhao et al. 2012; Sement et al. 2013; Chang et al. 2014; Lim et al. 2014; Tu et al. 2015; Zuber et al. 2016), although *A. thaliana* HESO-1 TUTase

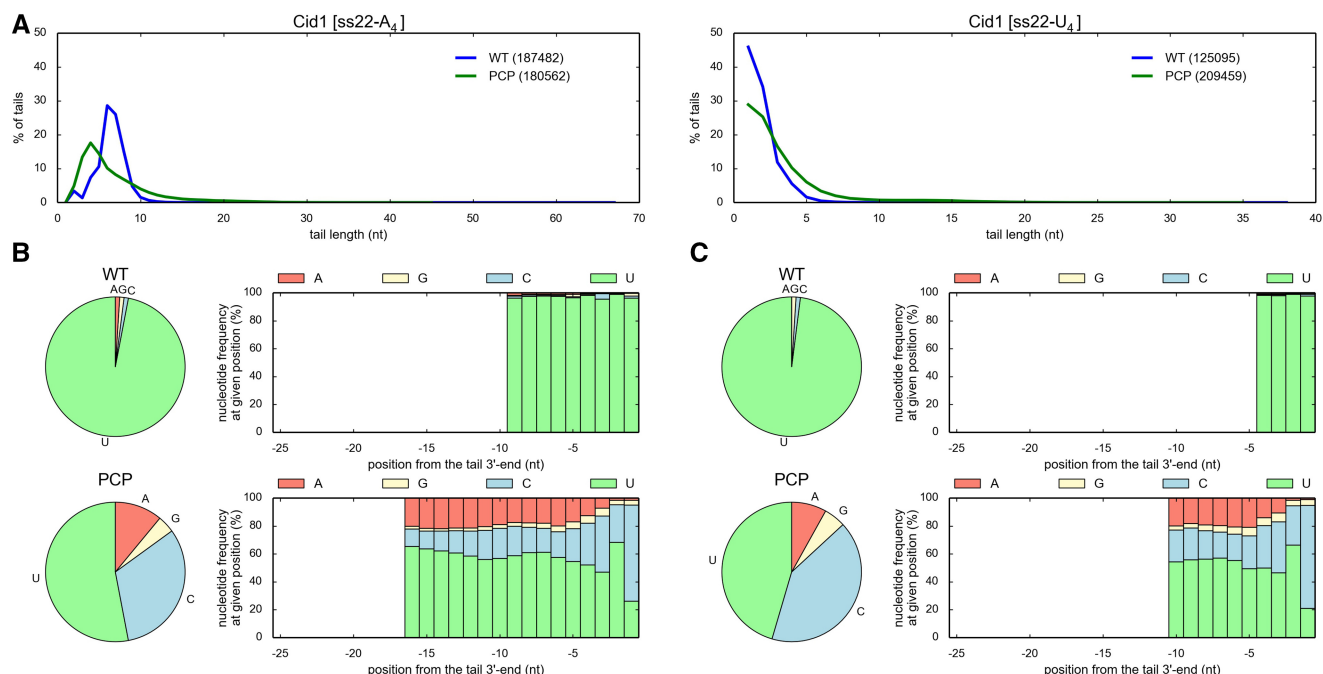


FIGURE 14. Substitution of several amino acids within Cid1 NRM increases its specificity toward CTP. (A) Length distribution for tails added by wild-type (WT) and D330E/E333N/H336R triple mutant (PCP) Cid1 variants to ss22-A₄ (left) and ss22-U₄ (right) oligoribonucleotide substrates. Numbers of analyzed sequences, obtained after 3'-RACE-seq high-throughput analysis for each protein variant, are indicated in parentheses. (B, C) Analysis of nucleotide composition for tails added by Cid1 WT and PCP variants to ss22-A₄ (panel B) and ss22-U₄ (panel C) RNA substrates. Results are presented as in Figure 6.

possesses valine instead of otherwise conserved histidine (Table 1). It would be interesting to check whether differences between these NRM clusters present in *A. thaliana* HESO-1 and URT1 TUTases (Table 1) contribute somehow to variable preference of these two enzymes toward small RNA substrates terminating with adenosines or uridines, as has been demonstrated recently (Tu et al. 2015).

On the contrary, *S. cerevisiae* Trf4 and Trf5 proteins, displaying PAP activity (LaCava et al. 2005; Vanacova et al. 2005; Wyers et al. 2005; Houseley and Tollervey 2006; Hamill et al. 2010), contain **DxxDxxN** cluster at the equivalent position and the same is true for *S. pombe* nuclear non-canonical PAP, Cid14 (Table 1; Win et al. 2006). Importantly, asparagine seems to be invariantly present in the last position of the NRM cluster in all noncanonical NTases, for which PAP activity has been experimentally demonstrated, including recently studied human GLD-2 (TUT2) which contains **ExxDxxN** cluster (Table 1; Chung et al. 2016). In concordance, the Cid1 H336N mutation (creating **DxxExxN** cluster) changed Cid1 specificity from PUP into PAP (Lunde et al. 2012). The identity of the remaining residues within the cluster seems to be less critical—the first one is always acidic (D or E), whereas the middle one is most often aspartate, but other amino acids also occur at this position, as exemplified by well-characterized *C. elegans* PAPs: Gld-2 and Gld-4 and *Drosophila melanogaster* Trf4 homolog, DmTrf4-1 (Table 1; Kwak et al. 2004; Nakamura et al. 2008; Schmid et al. 2009).

These examples indicate that the presence of H or N at the last position of the cluster largely determines whether the noncanonical NTase acts as PUP or PAP, respectively. The only significant exception from this rule regards mitochondrial PAPs (mtPAPs) in animals (Tomecki et al. 2004), which contain a completely different, but highly conserved, **NxxExxL** cluster (Table 1). Notably, substitution of the asparagine with aspartate results in the compromised PAP activity, both in vitro and in vivo, which leads to development of progressive neurodegenerative disorder—spastic ataxia 4 (Wilson et al. 2014; Lapkouski and Hallberg 2015). Structural studies revealed that this asparagine may play a role in a correct spatial positioning of the mRNA 3'-end and the bound ATP substrate, rather than in discrimination of mtPAP specificity toward NTP (Lapkouski and Hallberg 2015). Therefore, the molecular mechanisms ensuring the same NTP preference of nuclear/cytoplasmic and mitochondrial noncanonical PAPs may be distinct, and further structural insights are clearly required to understand these differences. It would be particularly interesting to determine whether there is any link between the divergence of NRM clusters present in nuclear/cytoplasmic and mitochondrial PAPs and the fact that the former act as homodimers, which is a unique feature among noncanonical NTases, obligatory for enzymatic activity of mtPAPs (Bai et al. 2011; Lapkouski and Hallberg 2015).

In this paper, we investigated the basis of unusual properties of the fungal CutA enzymes. Unlike other authors, we

TABLE 1. Comparison of cluster sequences present in NRMs of noncanonical NTases with distinct NTP specificities

Poly(U) polymerases (confirmed)	
<i>S. pombe</i> Cid1 (SPAC19D5.03)	DPF E ISHNVGR
<i>Homo sapiens</i> ZCCHC11/TUT4 (Uniprot: Q5TAX3.3)	DPF D LNHNLGA
<i>H. sapiens</i> ZCCHC6/TUT7 (Uniprot: Q5VYS8)	DPF D LNHNLGA
<i>H. sapiens</i> U6 TUTase/TUT1/Star-PAP (Uniprot: Q9H6E5)	DPF D LSHNVAA
<i>A. thaliana</i> URT-1 (At2g45620)	DPF E TS H DLGR
<i>D. melanogaster</i> Tailor (CG1091)	DPI Q LNHNVTK
<i>A. thaliana</i> HESO-1 (At2g39740)	DPF E Q P VNAAR
Poly(U) polymerases (predicted)	
<i>C. elegans</i> PUP-1/Cid1 homolog (CELE_K10D2.3)	DPF D LSHNLSS
<i>C. elegans</i> PUP-2 (CELE_K10D2.2)	DPF E TDHNLAQ
<i>C. elegans</i> MUT-2/RDE-3 (CELE_K04F10.6)	DPF D -E H NPGR
Poly(A) polymerases (confirmed)	
<i>S. cerevisiae</i> Trf4/Pap2 (Uniprot: P53632)	D P G D ES N IS R
<i>S. cerevisiae</i> Trf5/Pap2 (Uniprot: P48561)	D P G D PN N IS R
<i>S. pombe</i> Cid14 (SPAC12G12.13c)	DPV D FQ N DVSK
<i>D. melanogaster</i> Wispy/GLD-2 (CG15737)	EPF D QT N TARS
<i>H. sapiens</i> hGLD-2/TUT2 (Uniprot: Q6PIY7)	EPF D GT N TARA
<i>C. elegans</i> GLD-2 (CELE_ZC308.1)	EPF T NS N TAHS
<i>C. elegans</i> GLD-4 (CELE_ZK858.1)	DPL L TAN N DVGR
<i>D. melanogaster</i> Trf4-1 (CG11265)	DPL L PG N DVGR
Poly(A) polymerases (predicted)	
<i>S. pombe</i> Cid12 (SPCC663.12)	DPI D R N NDIGK
<i>C. elegans</i> CELE_F31C3.2	EPF D AIN N TART
<i>D. melanogaster</i> Trf4-2 (CG17462)	DPV T PT N DIG R
<i>H. sapiens</i> PAPD7/POLS/TRF4-1 TUT5 (Uniprot: Q5XG87)	DPL L PG N DVGR
<i>H. sapiens</i> PAPD5/TRF4-2/TUT3 (Uniprot: Q8NDF8)	DPL Q PG N DVGR
<i>A. thaliana</i> At2g40520	DPL K HS N NLGR
<i>A. nidulans</i> CutB (AN5694)	D P NNPD N DISC
Mitochondrial poly(A) polymerases	
<i>H. sapiens</i> hmtPAP/PAPD1 (Uniprot: Q9NVV4)	NPF E TS L NISK
<i>Gallus gallus</i> mtPAP (Uniprot: F1NBW0)	NPF E Q D L N ISK
<i>D. melanogaster</i> MTPAP/PAPD1 (CG11418)	NPL E Q L LNVSK
<i>Danio rerio</i> mtPAP (Uniprot: E7FEW2)	NPF E AT L NVSK
<i>Strongylocentrotus purpuratus</i> mtPAP (Uniprot: W4Y7K4)	NPL E P D L N SSK
C/U-adding NTases (confirmed)	
<i>A. nidulans</i> CutA (AN7748)	EPF N T I R N LGN
<i>T. terrestris</i> CutA (Uniprot: G2R014)	EPF N D G R N LGN
C/U-adding NTases (predicted)	
<i>S. pombe</i> Cid11 (SPBC1685.06)	EPF N T A R N LAN
<i>S. pombe</i> Cid13 (SPAC821.04c)	EPF H T S R N LAN
<i>D. melanogaster</i> Mkg-p (CG7163)	DPF E LS R NVAK
<i>C. elegans</i> PUP-3 (CELE_F59A3.9)	DPL L I P R N CAG
Uncertain	
<i>A. thaliana</i> At3g45750	DFT N IS Q NVAR
<i>S. pombe</i> Cid16 (SPAC17H9.01)	DPF L KK K NLTK

Amino acids located at positions equivalent to *S. pombe* Cid1 330–340 are shown. Highly conserved residues at positions 1, 4, and 7 of NRM clusters characteristic for PUPs/TUTases, PAPs, mtPAPs and C/U-adding enzymes are highlighted in green, yellow, blue, or magenta, respectively. Other residues occurring at these positions, or those present in enzymes with unclear specificity are highlighted in boldface black. Arginine, located four amino acids downstream from the NRM cluster characteristic for PAPs, which plays an important role in ATP binding, is indicated in red. Where indicated, amino acid sequences were retrieved from the Uniprot database (www.uniprot.org). In remaining cases, databases dedicated to given model organisms were utilized as sequence sources (*S. pombe*—www.pombase.org; *D. melanogaster*—flybase.org; *C. elegans*—www.wormbase.org; *A. nidulans*—www.aspergillusgenome.org), and numbers in parentheses refer to IDs from these databases.

refrained from drawing conclusions based on gel analysis of the in vitro reaction products generated by recombinant proteins in the presence of individual NTPs. We reasoned that such, *nota bene* commonly used, approach provides information about differences in enzyme processivity toward particular NTPs (see Fig. 13 herein as an example), rather than illustrates genuine NTase specificity. Instead, we took advan-

tage of the high-throughput methodology and subjected 3'-RACE products generated for tails synthesized by recombinant proteins in the presence of AGCU mixture to next-generation sequencing and data analysis. This strategy allowed us to quantify percentage of individual nucleotides in tails generated by analyzed NTases and to analyze distribution of nucleotides at every position of the synthesized extensions.

Owing to this approach, we were able to confirm that *S. pombe* Cid1 is extremely specific toward uridines, whereas CutA enzymes from two different fungal species display less strict specificity, but tend to predominantly incorporate cytidines. Importantly, the most striking feature of these enzymes is a frequent incorporation of two cytidine residues as the last added nucleotides, evident even at a physiologically relevant NTP ratio. Moreover, biochemical experiments with CTP as the only nucleotide present in the reaction mixture revealed that if RNA is in excess over protein, mimicking the *in vivo* situation, CutA is not able to efficiently extend the RNA substrate with more than two cytidines.

As expected, a 10-fold increase in the ATP concentration over CTP resulted in the appearance of longer adenosine stretches in the tails synthesized by the wild-type CutA, but the overall incorporation of cytidines did not decrease dramatically, indicating that the enzyme retains a high degree of specificity toward CTP, even if the latter is underrepresented. The pattern of tail composition, wherein strings of adenosines are interrupted by one or two cytidines and frequently terminate with two C residues, is due to several factors, as demonstrated herein, namely: character of the nucleotide recognition motif (NRM) and variable affinity of the enzyme toward different NTP substrates, much higher processivity of CutA with ATP than with CTP or UTP, as well as a tendency to poorly bind RNA molecules with two cytidines at their 3'-ends. It should be emphasized that these findings may be relevant to CutA activity *in vivo*. In *A. nidulans*, the enzyme was implicated in addition of C/U-rich extensions to polyadenylated mRNAs. However, it was not directly demonstrated that these unique tails are added to nascent poly(A) tails. It could be envisioned that after normal mRNA deadenylation, CutA synthesizes mixed extensions, containing several processively added adenosines, followed by distributively added cytidines and uridines. Interestingly, in addition to CutA's previously described CU-tagging activity (Morozov et al. 2010), it has recently been found that CutA is able to adenylate 3'-ends of some mRNAs *in vivo* (N Dube, IY Morozov, and MX Caddick, pers. comm.) Thus, although our work was performed on recombinant proteins, concordance with *in vivo* observations suggests that the nucleotide specificity of CutA was preserved.

Comparison of the 3D model of CutA with solved crystal structures of Cid1 and Trf4, combined with site-directed mutagenesis led to the conclusion that CutA NTases have an unusual nucleotide-binding site, in which the unique ExxNxxR cluster, at position equivalent to NRM clusters present in the noncanonical PAPs and PUPs, is mainly responsible for the observed preference of CutA toward cytidine. To support this conclusion, we demonstrated that introduction of ExxNxxR cluster into Cid1 NRM significantly increased cytidine incorporation at the expense of uridine inclusion into the 3'-terminal tails. Importantly, replacing CutA ExxNxxR with Cid1-like DxxExxH resulted in generation of the tails containing less cytidines and more uridines, but such a triple mutation

did not lead to the complete conversion of CutA into PUP enzyme. Similarly, we were not able to convert CutA into highly efficient PAP, probably due to a unique geometry and environment of the nucleotide binding site, which makes it impossible to achieve specificity toward a single nucleotide by only a few mutations. Further structural studies—particularly analysis of CutA cocrystals with different NTP substrates, supported by mutational and biochemical studies are clearly needed to fully understand the complete network of protein-nucleotide interactions responsible for the specificity of NTP selection for these noncanonical NTases.

We believe that based on the NRM cluster sequence one may now try to predict the specificity of the putative noncanonical NTases in the majority of the cases. For instance, there are three yet-uncharacterized enzymes from *C. elegans*, which all contain DxxD/ExxH NRM cluster and several other proteins from different species with D/ExxXxxN cluster, which are most likely PUPs and PAPs, respectively (Table 1). Most of the NTases from the latter group also possess arginine located four amino acids downstream from the NRM cluster, which is known to stabilize interaction with ATP, provided that the NRM cluster terminates with asparagine (Table 1).

Furthermore, inspection of NRM clusters present in enzymes from several model species suggests that there are enzymes which may display an NTP preference similar to CutA, namely: Cid11 and Cid13 from *S. pombe*, Mkg-p from *D. melanogaster* and PUP-3 from *C. elegans* (Table 1). Neither of these proteins has been thoroughly characterized in terms of NTP substrate specificity, however it was shown that Cid13 utilizes CTP more readily than UTP or GTP (Saitoh et al. 2002). Therefore, it would be worth checking if the enzymes mentioned above are able to synthesize 3'-terminal tails containing mainly cytidines *in vitro* using the 3'-RACE-seq approach described herein, as well as to perform in parallel global TAIL-seq experiments to analyze whether RNA molecules with C/U-rich tails exist *in vivo* in fission yeast, fruitfly and worm. Strikingly, it seems that mammalian genomes do not encode a noncanonical PAP with NRM cluster that would suggest such specificity toward CTP, which is in agreement with TAIL-seq analysis of human mRNA 3'-ends, in which tails containing cytidine were very rarely detected, when compared to uridine-rich extensions (Chang et al. 2014). All of that suggests that various species may utilize different enzymes and mechanisms to modify mRNA fate by addition of nontemplated nucleotides to the 3' ends of RNA molecules, however further research is needed to uncover physiological functions of various noncanonical PAPs with distinct nucleotide specificities.

MATERIALS AND METHODS

Bacterial strains

The following *Escherichia coli* strains were used in this study:

- MH1 (*E. coli* *araD lacX74 galU hsdR hsdM rpsL*),

- BL21-CodonPlus-RIL [Stratagene; *E. coli* B⁺ *ompT hsdS*(r_B⁺ m_B⁺) dcm⁺ Tet^r gal endA Hte (argU ileY leuW Cam^r)]

Oligonucleotides, plasmids, and cloning

Plasmids and oligonucleotides used in this study are listed in Supplemental Tables S2 and S3, respectively. All restriction enzymes and Phusion DNA polymerase for amplification of inserts were from Thermo Fisher Scientific. DNA purification kits (DNA Plasmid Mini and Gel-Out) were from A&A Biotechnology.

All plasmids for heterologous expression of fungal proteins in *E. coli* were generated by a sequence and ligation independent cloning (SLIC) approach (Li and Elledge 2007).

Plasmids pUC57-AnCutA and pUC57-TtCutA contained recoded sequences (see Supplemental Fig. S1) of CutA enzymes from *A. nidulans* and *T. terrestris*, which were synthesized by Genscript with the aim of codon optimization for heterologous expression in *E. coli*, and precloned into pUC57 vector.

Plasmids pCC1 and pCC4 were generated by inserting open-reading frames corresponding to the full-length AnCutA and TtCutA, respectively, amplified using respective primer pairs (278F–278R and 280F–280R) and templates (pUC57-AnCutA and pUC57-TtCutA), into NcoI–XhoI sites of pET28 C-6xHisTag vector.

Plasmid pCC2 was generated by site-directed mutagenesis with oligonucleotide pair CUCUmutF–CUCUmutR, using pCC1 as a template.

Plasmid pCC3 was constructed by inserting AnCutA ORF fragment corresponding to amino acids 84–487, amplified with primer pair 278SumoF–278Sumo2 and pCC1 as a template, into BamHI–XhoI sites of pET28 N-6xHis-SUMOTag vector. Plasmid pCC5, containing TtCutA ORF fragment corresponding to amino acids 216–617, was generated similarly, but using primer pair 280SumoF–280Sumo2 and pCC4 as a template in the amplification step.

Plasmids pCC6 and pCC7, containing TtCutA ORF fragments corresponding to amino acids 216–557 or 127–827, respectively, were constructed by inserting amplicons obtained with the respective primer pairs (12.280 S_F–12.280 sI_R or 280sIIIF–280sIIIR), into NcoI–XhoI sites of pET28 C-6xHisTag vector.

Plasmids pCC8, pCC9, pCC12, pCC14, and pCC15 were generated by site-directed mutagenesis with oligonucleotide pairs 280PAP1F–280PAP1R, 280PAP2F–280PAP2R, 280PAP5F–280PAP5R, 280PUP1F–280PUP1R, and 280PUP2F–280PUP2R, respectively, using pCC7 as a template. Plasmids pCC10 and pCC16 were constructed by site-directed mutagenesis with oligonucleotide pairs 280PAP3F–280PAP3R or 280PUP3F–280PUP3R, using pCC8 or pCC15 as templates, respectively. Plasmid pCC11 was generated by site-directed mutagenesis with oligonucleotide pair 280PAP4F–280PAP4R, using pCC8 as a template. Plasmid pCC13 was constructed by site-directed mutagenesis with oligonucleotide pair 280PAP6F–280PAP6R, using pCC15 as a template.

Plasmid pCC17 was constructed by inserting Cid1 ORF, amplified with primer pair Cid1SumoF–Cid1SumoR, using cDNA obtained following oligo(dT)-primed reverse transcription on total *S. pombe* RNA as a template, into BamHI–XhoI sites of pET28 N-6xHis-SUMOTag vector. Plasmid pCC18 was generated by site-directed mutagenesis with oligonucleotide pair Cid1PCPF–Cid1PCPR, using pCC17 as a template.

The presence of introduced mutations was verified by digestion with relevant restriction enzymes and correctness of sequences

was checked by sequencing the inserts with appropriate primers, listed in Supplemental Table S3.

Heterologous expression of proteins in *E. coli*

For expression of different variants of fungal nucleotidyltransferases, *Escherichia coli* BL21-CodonPlus-RIL strain was transformed with appropriate plasmids (pCC1–pCC18). Transformants were grown at 18°C for 48 h in 1 L or 50 mL (for large-scale and small-scale purification, respectively) of Auto Induction Media Super Broth Base Including Trace Elements (Formedium), supplemented with 2% glycerol, kanamycin (50 µg/mL), and chloramphenicol (34 µg/mL), following inoculation from a preculture, grown in standard Luria-Broth (LB) medium containing both antibiotics.

Small-scale protein purification

A bacterial pellet from 50 mL of liquid culture was resuspended in 5 mL of the lysis buffer (20 mM Na-phosphate, pH = 4.5 or 8.0/ 20 mM Tris–HCl, pH = 8.0; 300 mM NaCl; 10 mM imidazole; 1 mM PMSF; 10 mM β-mercaptoethanol; 50 µg/mL lysozyme) and subjected to 60 cycles of sonication (30 sec of full power sonication : 30 sec pause) in a Bioruptor XL device (Diagenode). Homogenate was centrifuged at 10,000g for 45 min at 4°C. The supernatant was then loaded onto a Protino 96 Ni-IDA plate (Macherey-Nagel), pre-equilibrated with the lysis buffer. Following protein binding, the resin was washed sequentially with 2 mL of low-salt (LS) buffer (20 mM Na-phosphate/Tris–HCl, pH = 8.0; 150–300 mM NaCl; 10 mM imidazole), 1 mL of high-salt (HS) buffer (20 mM Na-phosphate/Tris–HCl, pH = 8.0; 1 mM NaCl; 10 mM imidazole) and 2 mL of LS buffer. Bound proteins were recovered by three consecutive rounds of elution with 0.25 mL of buffer E (20 mM Na-phosphate/Tris–HCl, pH = 8.0; 150–300 mM NaCl; 300 mM imidazole). Purified proteins were analyzed by standard SDS-PAGE.

Large-scale protein purification

Following centrifugation of 1 L of the liquid culture at 4500 rpm in a Sorvall H6000A/HBB6 swinging bucket rotor for 15 min at 4°C, a bacterial pellet was resuspended in 70 mL of the lysis buffer, incubated with lysozyme (50 µg/mL; Roth) for 20 min in a cold room, and then broken in the EmulsiFlex-C3 High Pressure homogenizer at 1500 Bar. Homogenate was centrifuged in a Sorvall WX ULTRA SERIES ultracentrifuge (F37L rotor) at 32,000 rpm for 45 min at 4°C. The supernatant (extract) was then used for protein purification with the ÄKTA Xpress system (GE Healthcare).

Depending on the proteins' stability and contamination level, several various methods were used in purification procedures. The basic step of purification was nickel affinity chromatography on the 5 mL column compatible with ÄKTA, filled with Ni-NTA Superflow resin (Qiagen). The column was equilibrated with 25 mL of LS buffer prior to the extract loading. After protein binding, the resin was washed sequentially with 35–60 mL of LS buffer, 25 mL of HS buffer (to reduce contaminations), and 20 mL of LS buffer. Proteins with a C-terminal HisTag were simply recovered by elution with 25 mL of buffer E. For proteins fused with an N-terminal HisTag-SUMO, purification involving SUMO protease cleavage

(see Supplemental Table S4) included additional on-column digestion with 100 µg of the home-made SUMO protease resuspended in 5 mL of buffer E (8 h), followed by buffer exchange to LS on the desalting column and final separation of the desired protein from the cut out HisTag-SUMO epitope and SUMO protease on the Ni-NTA column.

Furthermore, size-exclusion chromatography on a Hiload 16/60 Superdex S200 column (GE Healthcare), was carried out in the majority of cases by applying 5 mL of eluate from the affinity chromatography step, with the use of 1.2 column volumes of gel-filtration (GF) buffer (20 mM Na-phosphate/Tris-HCl, pH = 8.0; 150 mM NaCl), to increase protein purity. Purified proteins were analyzed by standard SDS-PAGE. See Supplemental Table S4 for more details on purification protocols applied to particular CutA variants.

Protein concentration

Following purification, recombinant proteins were concentrated with the use of Vivaspin (500, 2, 4, or 15, depending on the initial volume of protein solution) ultrafiltration devices (Sartorius Stedim), according to the manufacturer's instructions.

Limited proteolysis

After small-scale purification of the full-length TtCutA, 15 µg of the protein was subjected to digestion with various trypsin concentrations (ranging from 0.0005 to 0.125 µg/µL) in 20 µL of the buffer containing 10 mM Tris-HCl, pH = 7.4 and 100 mM NaCl for 30 min at room temperature. Of note, 1 mM PMSF was then added to terminate reactions and standard SDS-PAGE was used for products' analysis.

Edman degradation analysis

Three hundred micrograms of TtCutA was digested as above in a total volume of 0.8 mL with trypsin (0.001 µg/mL) for Edman degradation analysis. Reaction products were separated by gel filtration on a Hiload 16/60 Superdex 75 column (GE Healthcare) in a buffer containing 10 mM Na-phosphate, pH = 8.0 and 100 mM NaCl. The fraction containing properly sized protein (about 40 kDa, as verified by SDS-PAGE) was concentrated approximately 10-fold. Fifteen microliters of the concentrated protein preparation was then mixed with 5 µL of denaturation buffer (150 mM Tris-HCl, pH = 6.8; 12% SDS; 30% glycerol; 0.05% Coomassie blue G-250; 6% β-mercaptoethanol) and separated in Tricine-SDS-PAGE with the use of anode buffer (0.2 M Tris-HCl, pH = 8.9) and cathode buffer (0.1 M Tris; 0.1 M tricine; 0.1% SDS). The protein was then subjected to wet electrotransfer onto methanol preactivated PVDF membrane (Thermo Fisher Scientific) in CAPS buffer (0.22% CAPS, pH = 11.0; 10% methanol). Following transfer, the membrane was stained with 0.1% Coomassie Brilliant Blue R-250 in 40% methanol; 1% acetic acid, then washed four times in 50% methanol and air-dried. The PVDF membrane fragment containing a protein of interest was cut out and an N-terminal sequence was analyzed by chemical Edman degradation with a fully automated Procise 491 Sequencer (Applied Biosystems) at the Selvita company.

Protein secondary structure analysis

CFSP (Chou&Fasman Secondary Structure Prediction Server, available at <https://sourceforge.net/projects/cfpred/files/latest/download>) was used to predict AnCutA and TtCutA secondary structures.

3D protein model

Homologs of TtCutA and AnCutA proteins were identified with BLAST (Altschul et al. 1997) searches performed against the NCBI nonredundant protein sequence database using sequences of TtCutA and AnCutA as queries. Multiple sequence alignment was derived using PCMA program (Pei et al. 2003) followed by some manual adjustments. Secondary structures were predicted with PSIPRED (McGuffin et al. 2000). A 3D model of TtCutA protein was constructed with MODELLER (Fiser and Sali 2003) using *S. pombe* Cid1 poly(U) polymerase (pdb|4fh5) (Lunde et al. 2012) as a template (Ginalski and Rychlewski 2003). The structures were visualized with Pymol (pymol.org).

In vitro biochemical assays

Nucleotidyltransferase activity of different AnCutA, TtCutA, and Cid1 protein variants was assessed by in vitro assays. Reactions were performed in the volume of 20 µL, containing 10 mM Tris-HCl, pH = 8.0; 50 mM NaCl; 1 mM MnCl₂; 1 mM DTT; and either individual NTP (ATP/GTP/CTP/UTP; 1 mM) or mixtures of two (0.5 mM each) or four (0.25 mM each, or 0.77 mM ATP:0.077 mM GTP:0.077 CTP:0.077 UTP, when a physiologically relevant NTP ration was desirable) different NTPs. In most experiments, each reaction was programmed with 5 µL of the purified protein (0.5 µM), and started by adding 4 µL of 1 µM 5'-FAM-labeled RNA substrate solution (see Supplemental Table S3 for sequences). For estimation of enzyme affinity towards various nucleotides and analysis of processivity, reactions were performed similarly, but using an excess of RNA substrate (1 µM) over the enzyme (0.025 µM), and individual NTPs (ATP, CTP, or UTP) at final concentrations ranging from 0.1 to 300 µM. Reactions were carried out at 37°C and terminated at various time-points after the start by collecting 4 µL of reaction mixture and adding it to 5 µL of formamide loading dye (90% formamide; 20 mM EDTA; 0.03% bromophenol blue; 0.03% xylene cyanol in 1× Tris-borate-EDTA [TBE]), followed by snap-freezing in liquid nitrogen. After thermal denaturation at 90°C for 5 min, reaction products were resolved either in denaturing 20% polyacrylamide/8 M urea/1× TBE gels, or—if single-nucleotide resolution was required—in 10% sequencing gels. A Typhoon FLA 9000 laser scanner (GE Healthcare) was used to visualize fluorescently labeled RNA molecules.

3'-RACE-seq

Nucleotide composition of the 3'-extensions added by different variants of fungal nucleotidyltransferases in the in vitro assays was analyzed by 3'-RACE and sequencing. Sequences of oligonucleotides used in the 3'-RACE are listed in the Supplemental Table S3.

RNA purification

Following RNA visualization, a gel fragment containing reaction products was cut out, frozen in liquid nitrogen and crushed with a pipette tip. RNA was then eluted from acrylamide with 350 μ L of the gel elution buffer (100 mM Tris-HCl, pH = 8.0; 150 mM NaCl; 12.5 mM EDTA; 1% SDS) in the presence of 800 μ L of phenol:chloroform (1:1, v:v) on a rotating device at room temperature for 16 h. Following centrifugation at 13,500g for 15 min at room temperature, the aqueous layer was transferred to a new tube and RNA was precipitated with isopropanol in the presence of 1 μ L of glycogen (20 mg/mL) at -20°C for 16 h. RNA was pelleted by centrifugation at 21,000g for 30 min at 4°C and washed twice with 80% ethanol. Eventually, the RNA pellet was suspended in 11 μ L of RNase-free water.

Adapter ligation

Of note, 10.7 μ L of purified RNA was mixed with 2 μ L of T4 RNA Ligase Reaction buffer (New England Biolabs), 1 μ L of 30 μ M RA3 adapter, 4.8 μ L of 50% PEG 8000 (New England Biolabs) and 0.5 μ L of RiboLock RNase Inhibitor (40 U/ μ L; Thermo Fisher Scientific). After 2 min of incubation at 70°C , the tube was placed on ice, and 1 μ L of T4 RNA Ligase 2, Truncated (200 U/ μ L; New England Biolabs) was added. The reaction was carried out for 1 h at 25°C .

Reverse transcription

Following ligation of the 3'-adapter, 1.5 μ L of 20 μ M reverse transcription (RTP) primer was annealed by heating the mixture at 75°C for 5 min, followed by incubation at 37°C for 30 min and at 25°C for 15 min. Ten microliters aliquot of the mixture was transferred to a new tube, mixed with 4 μ L of the 5 $\times$ First Strand Buffer (Invitrogen), 1 μ L of 10 mM dNTP Mix, 0.5 μ L of RiboLock RNase Inhibitor, 2 μ L of 100 mM DTT, 1.5 μ L of RNase-free H_2O and heated at 42°C for 2 min. Afterwards, 1 μ L of the SuperScript III Reverse Transcriptase (Invitrogen) was added, and RT-PCR was carried out at 42°C for 1 h. Enzyme inactivation was performed by heating the reaction mixture at 70°C for 30 min.

Amplification

Two-step amplification was performed for high-throughput RNA-seq of the 3'-RACE products. The first round PCR reaction mixture contained 20 μ L of the first-strand cDNA, 0.2 mM dNTP Mix, 20 μ M primer pair RA5_ss22-RPI_X (with Illumina index for RNA-seq), 5 $\times$ Phusion HF Buffer, Phusion DNA polymerase, and sterile water to 40 μ L. The following PCR profile was utilized: initial denaturation at 98°C for 30 sec, 25 cycles of amplification (with denaturation at 98°C , annealing at 55°C , and elongation at 72°C carried out for 10, 10, and 30 sec, respectively) and final elongation at 72°C for 5 min. The second round PCR reaction mixture comprised 1 μ L of the first-round PCR products, 0.2 mM dNTP Mix, 20 μ M primer pair RPI1-RPI_X, 5 $\times$ Phusion HF Buffer, Phusion DNA polymerase, and sterile water to 50 μ L. The PCR profile described above was essentially the same as in the first round, but with only five cycles of amplification. PCR products were analyzed by electrophoresis in 2% agarose gels.

Purification of the 3'-RACE-seq libraries

PCR products obtained following amplification were purified with the use of Agencourt AMPure XP magnetic beads (Beckman Coulter). Briefly, 48 μ L of PCR products were mixed with 48 μ L of the beads, incubated for 15 min at room temperature, and then placed on a magnetic stand. After 5 min, supernatant was discarded, and the beads were washed twice with 100 μ L of 80% ethanol and air-dried for 15 min. Elution was done by suspending the beads in 20 μ L of sterile water and 5 min of incubation at room temperature. Magnetic stand was used to separate eluted DNA from the beads.

High-throughput sequencing of the 3'-RACE products (3'-RACE-seq)

Concentration of the 3'-RACE-seq libraries was measured with the KAPA Library Quantification Universal Kit (Kapa Biosystems). Sequencing was performed using an Illumina NextSeq 500 System and NextSeq 500 v2 High Output Kit (2×75 cycles), applying standard procedures of a pair-end run.

Analysis of the 3'-RACE-seq results

Initially, the strands encompassing sequences of RNA substrates in the sense orientation were subjected to filtering, consisting in a selection of sequences containing 26 nt of the respective RNA substrate (ss22-A₄ or ss22-U₄) at the 5'-end, with no more than two erroneous bases, and an entirely correct sequence of the RA3 oligonucleotide (21 nt) at the 3'-end. Detailed analyses for all filtered sequences were carried out with in-house scripts using Python, NumPy, SciPy, and Matplotlib.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was mainly supported by a grant from The National Centre for Research and Development (NCBR LIDER/35/46/L-3/11/NCBR/2012 to R.T.) and co-supported by the European Research Council (ERC Starting Grant 309419-PAP&PUPs to A.D.), the Polish National Science Centre (2011/02/A/NZ2/00014, 2014/15/B/NZ1/03357 to K.G.), and the Foundation for Polish Science (TEAM/2010-6, TEAM/2016-2 to K.G.). Experiments were carried out with the use of CePT infrastructure that was financed by the European Union via the European Regional Development Fund (Innovative economy 2007–13, Agreement POIG.02.02.00-14-024/08-00). We acknowledge Dorota Adamska and Paweł Krawczyk for assistance with 3'-RACE-seq and data analysis. We thank Karolina Drazkowska for reading the manuscript.

Author contributions: K. Kobyłecki prepared plasmid constructs, expressed and purified all proteins, performed the majority of biochemical experiments, including generation of the 3'-RACE-seq libraries, and carried out initial analysis of high-throughput sequencing data under the supervision of R.T. and A.D. K. Kuchta generated the 3D TtCutA model, performed structural analyses, and thoroughly analyzed 3'-RACE-seq results under the supervision of K.G., with contributions from R.T. A.D. participated in the design of the experiments. R.T. conceived and directed the studies with the

help of K.G., performed biochemical experiments at the stage of revision, and wrote the manuscript, with contributions from K.G., K. Kuchta, and A.D.

Received February 20, 2017; accepted September 12, 2017.

REFERENCES

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Aphasizhev R. 2005. RNA uridylyltransferases. *Cell Mol Life Sci* **62**: 2194–2203.
- Aravind L, Koonin EV. 1999. DNA polymerase β -like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res* **27**: 1609–1618.
- Bai Y, Srivastava SK, Chang JH, Manley JL, Tong L. 2011. Structural basis for dimerization and activity of human PAPD1, a noncanonical poly(A) polymerase. *Mol Cell* **41**: 311–320.
- Balbo PB, Bohm A. 2007. Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure* **15**: 1117–1131.
- Chang HM, Triboulet R, Thornton JE, Gregory RI. 2013. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature* **497**: 244–248.
- Chang H, Lim J, Ha M, Kim VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* **53**: 1044–1052.
- Chowdhury A, Mukhopadhyay J, Tharun S. 2007. The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* **13**: 998–1016.
- Chung CZ, Jo DH, Heinemann IU. 2016. Nucleotide specificity of the human terminal nucleotidyltransferase Gld2 (TUT2). *RNA* **22**: 1239–1249.
- Faehnle CR, Walleshauser J, Joshua-Tor L. 2014. Mechanism of Dis3l2 substrate recognition in the Lin28-let-7 pathway. *Nature* **514**: 252–256.
- Fiser A, Sali A. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**: 461–491.
- Ginalski K, Rychlewski L. 2003. Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment. *Proteins* **53** (Suppl 6): 410–417.
- Hagan JP, Piskounova E, Gregory RI. 2009. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* **16**: 1021–1025.
- Hamill S, Wolin SL, Reinisch KM. 2010. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc Natl Acad Sci* **107**: 15045–15050.
- Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, Yeom KH, Han J, Kim VN. 2009. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**: 696–708.
- Houseley J, Tollervey D. 2006. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep* **7**: 205–211.
- Kim B, Ha M, Loeff L, Chang H, Simanshu DK, Li S, Fareh M, Patel DJ, Joo C, Kim VN. 2015. TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *EMBO J* **34**: 1801–1815.
- Kuchta K, Knizewski L, Wyrwicz LS, Rychlewski L, Ginalski K. 2009. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res* **37**: 7701–7714.
- Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K, Rychlewski L, Ginalski K. 2016. FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res* **44**: 3534–3548.
- Kuhn CD, Wilusz JE, Zheng Y, Beal PA, Joshua-Tor L. 2015. On-enzyme refolding permits small RNA and tRNA surveillance by the CCA-adding enzyme. *Cell* **160**: 644–658.
- Kwak JE, Wickens M. 2007. A family of poly(U) polymerases. *RNA* **13**: 860–867.
- Kwak JE, Wang L, Ballantyne S, Kimble J, Wickens M. 2004. Mammalian GLD-2 homologs are poly(A) polymerases. *Proc Natl Acad Sci* **101**: 4407–4412.
- Labno A, Tomecki R, Dziembowski A. 2016a. Cytoplasmic RNA decay pathways - enzymes and mechanisms. *Biochim Biophys Acta* **1863**: 3125–3147.
- Labno A, Warkocki Z, Kulinski T, Krawczyk PS, Bijata K, Tomecki R, Dziembowski A. 2016b. Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs. *Nucleic Acids Res* **44**: 10437–10453.
- LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**: 713–724.
- Lapkouski M, Hallberg BM. 2015. Structure of mitochondrial poly(A) RNA polymerase reveals the structural basis for dimerization, ATP selectivity and the SPAX4 disease phenotype. *Nucleic Acids Res* **43**: 9065–9075.
- Li MZ, Elledge SJ. 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **4**: 251–256.
- Lim J, Ha M, Chang H, Kwon SC, Simanshu DK, Patel DJ, Kim VN. 2014. Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* **159**: 1365–1376.
- Lunde BM, Magler I, Meinhart A. 2012. Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res* **40**: 9815–9824.
- Martin G, Keller W. 2007. RNA-specific ribonucleotidyl transferases. *RNA* **13**: 1834–1849.
- McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404–405.
- Morozov IY, Jones MG, Razak AA, Rigden DJ, Caddick MX. 2010. CUCU modification of mRNA promotes decapping and transcript degradation in *Aspergillus nidulans*. *Mol Cell Biol* **30**: 460–469.
- Munoz-Tello P, Gabus C, Thore S. 2012. Functional implications from the Cid1 poly(U) polymerase crystal structure. *Structure* **20**: 977–986.
- Munoz-Tello P, Gabus C, Thore S. 2014. A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure. *Nucleic Acids Res* **42**: 3372–3380.
- Munoz-Tello P, Rajappa L, Coquille S, Thore S. 2015. Polyuridylation in eukaryotes: a 3'-end modification regulating RNA life. *Biomed Res Int* **2015**: 968127.
- Nakamura R, Takeuchi R, Takata K, Shimanouchi K, Abe Y, Kanai Y, Ruike T, Ihara A, Sakaguchi K. 2008. TRF4 is involved in polyadenylation of snRNAs in *Drosophila melanogaster*. *Mol Cell Biol* **28**: 6620–6631.
- Pei J, Sadreyev R, Grishin NV. 2003. PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* **19**: 427–428.
- Pirouz M, Du P, Munafo M, Gregory RI. 2016. Dis3l2-mediated decay is a quality control pathway for noncoding RNAs. *Cell Rep* **16**: 1861–1873.
- Rajappa-Titu L, Suematsu T, Munoz-Tello P, Long M, Demir O, Cheng KJ, Stagno JR, Luecke H, Amaro RE, Aphasizheva I, et al. 2016. RNA editing TUTase 1: structural foundation of substrate recognition, complex interactions and drug targeting. *Nucleic Acids Res* **44**: 10862–10878.
- Reimao-Pinto MM, Manzenreither RA, Burkard TR, Sledz P, Jinek M, Mechtler K, Ameres SL. 2016. Molecular basis for cytoplasmic RNA surveillance by uridylation-triggered decay in *Drosophila*. *EMBO J* **35**: 2417–2434.
- Ren G, Chen X, Yu B. 2012. Uridylation of miRNAs by HEN1 SUPPRESSOR1 in *Arabidopsis*. *Curr Biol* **22**: 695–700.

- Rissland OS, Norbury CJ. 2008. The Cid1 poly(U) polymerase. *Biochim Biophys Acta* **1779**: 286–294.
- Rissland OS, Norbury CJ. 2009. Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. *Nat Struct Mol Biol* **16**: 616–623.
- Rissland OS, Mikulasova A, Norbury CJ. 2007. Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Mol Cell Biol* **27**: 3612–3624.
- Saitoh S, Chabes A, McDonald WH, Thelander L, Yates JR, Russell P. 2002. Cid13 is a cytoplasmic poly(A) polymerase that regulates ribonucleotide reductase mRNA. *Cell* **109**: 563–573.
- Scheer H, Zuber H, De Almeida C, Gagliardi D. 2016. Uridylation earmarks mRNAs for degradation... and more. *Trends Genet* **32**: 607–619.
- Schmidt MJ, Norbury CJ. 2010. Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *Wiley Interdiscip Rev RNA* **1**: 142–151.
- Schmid M, Kuchler B, Eckmann CR. 2009. Two conserved regulatory cytoplasmic poly(A) polymerases, GLD-4 and GLD-2, regulate meiotic progression in *C. elegans*. *Genes Dev* **23**: 824–836.
- Scott DD, Norbury CJ. 2013. RNA decay via 3' uridylation. *Biochim Biophys Acta* **1829**: 654–665.
- Sement FM, Ferrier E, Zuber H, Merret R, Alioua M, Deragon JM, Bousquet-Antonelli C, Lange H, Gagliardi D. 2013. Uridylation prevents 3' trimming of oligoadenylated mRNAs. *Nucleic Acids Res* **41**: 7115–7127.
- Song MG, Kiledjian M. 2007. 3' Terminal oligo U-tract-mediated stimulation of decapping. *RNA* **13**: 2356–2365.
- Stevenson AL, Norbury CJ. 2006. The Cid1 family of non-canonical poly(A) polymerases. *Yeast* **23**: 991–1000.
- Tomecki R, Dmochowska A, Gewartowski K, Dziembowski A, Stepien PP. 2004. Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Res* **32**: 6001–6014.
- Torralba S, Sojat J, Hartmann R. 2008. 2'-5' oligoadenylate synthetase shares active site architecture with the archaeal CCA-adding enzyme. *Cell Mol Life Sci* **65**: 2613–2620.
- Tu B, Liu L, Xu C, Zhai J, Li S, Lopez MA, Zhao Y, Yu Y, Ramachandran V, Ren G, et al. 2015. Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in *Arabidopsis*. *PLoS Genet* **11**: e1005119.
- Ustianenko D, Hrossova D, Potesil D, Chalupnikova K, Hrazdilova K, Pachernik J, Cetkovska K, Uldrijan S, Zdrahal Z, Vanacova S. 2013. Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. *RNA* **19**: 1632–1638.
- Ustianenko D, Pasulka J, Feketova Z, Bednarik L, Zigackova D, Fortova A, Zavolan M, Vanacova S. 2016. TUT-DIS3L2 is a mammalian surveillance pathway for aberrant structured non-coding RNAs. *EMBO J* **35**: 2179–2191.
- Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W. 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* **3**: e189.
- Viegas SC, Silva IJ, Apura P, Matos RG, Arraiano CM. 2015. Surprises in the 3'-end: 'U' can decide too! *FEBS J* **282**: 3489–3499.
- Wang L, Eckmann CR, Kadyk LC, Wickens M, Kimble J. 2002. A regulatory cytoplasmic poly(A) polymerase in *Caenorhabditis elegans*. *Nature* **419**: 312–316.
- Wilson WC, Hornig-Do HT, Bruni F, Chang JH, Jourdain AA, Martinou JC, Falkenberg M, Spahr H, Larsson NG, Lewis RJ, et al. 2014. A human mitochondrial poly(A) polymerase mutation reveals the complexities of post-transcriptional mitochondrial gene expression. *Hum Mol Genet* **23**: 6345–6355.
- Win TZ, Draper S, Read RL, Pearce J, Norbury CJ, Wang SW. 2006. Requirement of fission yeast Cid14 in polyadenylation of rRNAs. *Mol Cell Biol* **26**: 1710–1721.
- Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B, et al. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**: 725–737.
- Yates LA, Fleurdepine S, Rissland OS, De Colibus L, Harlos K, Norbury CJ, Gilbert RJ. 2012. Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nat Struct Mol Biol* **19**: 782–787.
- Yates LA, Durrant BP, Fleurdepine S, Harlos K, Norbury CJ, Gilbert RJ. 2015. Structural plasticity of Cid1 provides a basis for its distributive RNA terminal uridylyl transferase activity. *Nucleic Acids Res* **43**: 2968–2979.
- Yue D, Maizels N, Weiner AM. 1996. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA* **2**: 895–908.
- Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, Meyers BC, Mo B, Chen X. 2012. The *Arabidopsis* nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr Biol* **22**: 689–694.
- Zuber H, Scheer H, Ferrier E, Sement FM, Mercier P, Stupfler B, Gagliardi D. 2016. Uridylation and PABP cooperate to repair mRNA deadenylated ends in *Arabidopsis*. *Cell Rep* **14**: 2707–2717.



RNA

A PUBLICATION OF THE RNA SOCIETY

Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity toward CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails

Kamil Kobylecki, Krzysztof Kuchta, Andrzej Dziembowski, et al.

RNA 2017 23: 1902-1926 originally published online September 25, 2017

Access the most recent version at doi:[10.1261/rna.061010.117](https://doi.org/10.1261/rna.061010.117)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2017/09/25/rna.061010.117.DC1>

References

This article cites 68 articles, 18 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/23/12/1902.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Quantify the percentage of m6A modification of LncRNAs, CircRNAs & mRNAs



To subscribe to RNA go to:

<http://rnajournal.cshlp.org/subscriptions>

Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity towards CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails

Kamil Kobyłecki, Krzysztof Kuchta, Andrzej Dziembowski, Krzysztof Ginalski & Rafał Tomecki

SUPPLEMENTAL MATERIAL

Supplemental Figure.....	2
Supplemental Tables.....	12

SUPPLEMENTAL FIGURE

AnCutA	ATGGTGACCGTGACCGATGGCAACGATCGCCTGAGCAGCCGCCCGGGCGAACAGAGCCCG
AnCutA_opt	ATGGTCACCGTCACCGATGGTAACGACCGCCTGTCCAGCCGCCCGGGCGAACAAAGTCCG
protein	M V T V T D G N D R L S S R P G E Q S P
AnCutA	ATTGCGAGCGCGAAACTGAGCACCATGCCGTGGCGCAGCAGCAACGAAGAAAGCAGCGAA
AnCutA_opt	ATTGCCTCAGCCAAACTGTCCACGATGCCGTGGCGTAGCTCTAACGAAGAAAGTTCGGAA
protein	I A S A K L S T M P W R S S N E E S S E
AnCutA	ACCCCGAACCGCCCGGCGCTGACCAGCCAGCAGAGCAACAGCCTGCCGAGCACCCTGAT
AnCutA_opt	ACCCCGAATCGCCCGGCACTGACGTACAGCAATCGAACAGCCTGCCGTGACCCCGTAT
protein	T P N R P A L T S Q Q S N S L P S T P Y
AnCutA	CAGCATGCGCGCAAAATTAGCTTTTCATAGCCGAGCCCGAGCCCGCGCCATGGCAGCACC
AnCutA_opt	CAGCATGCTCGTAAAATTAGCTTTTCACTCTCGTAGTCCGTCCCCGCGTCATGGTTCAACG
protein	Q H A R K I S F H S R S P S P R H G S T
AnCutA	AGCCCGCGCAGCACCCATAGCGAAAGCACCCATATTCCGCGAGCGCGCGCAAACAGTTT
AnCutA_opt	TCGCCGCGTAGCACCCATAGCGAATCTACGCACATCCCGCCGTCTGCACGCAAACAGTTC
protein	S P R S T H S E S T H I P P S A R K Q F
AnCutA	ACCGGCTGCAAATATGAAACCGCGATGAGCTTTTTTCGCCGCGCATGCCGTATAGCATT
AnCutA_opt	ACCGGCTGCAAGTATGAAACCGGCTATGTCAATTTTTCCGTGCGCGTATGCCGTACTCGATT
protein	T G C K Y E T A M S F F R R R M P Y S I
AnCutA	GGCGCGGATGTGCTGCCGGAAGAAAAAGAAGGCCTGAAACAGCGCCTGCGCCCGGAAGAA
AnCutA_opt	GGTGCGGATGTGCTGCCGGAAGAAAAAGAAGGCCTGAAGCAGCGTCTGCGCCCGGAAGAA
protein	G A D V L P E E K E G L K Q R L R P E E
AnCutA	GAACGCAAACGACCGCGGATATGCTGGAAGTGTATGAACGCCTGCTGCCGAGCGCGGAA
AnCutA_opt	GAACGTAAACTGACCGCGGATATGCTGGAAGTTTATGAACGCCTGCTGCCGAGCGCGGAA
protein	E R K L T A D M L E V Y E R L L P S A E
AnCutA	AGCGAAAGCCGCCCGCCGCGCCTGGTGCGCAAACTGGAAGATCTGTTTAACCGCCAGTGG
AnCutA_opt	AGTGAATCCCGCGTCCGCGTCTGGTCCGTAACTGGAAGACCTGTTTAACCGCCAGTGG
protein	S E S R R R R L V R K L E D L F N R Q W
AnCutA	CCGGGCTGCGATATTAAAGTGCATGTGTTTGGCAGCAGCGGCAACAACTGTGCAGCAGC
AnCutA_opt	CCGGGTGCGACATTAAAGTGCATGTTTTCGGTTCATCGGGCAATAAGCTGTGTAGCTCT
protein	P G C D I K V H V F G S S G N K L C S S
AnCutA	GATAGCGATGTGGATATTTGCATTACCACCACCTGCAAAGAACTGGAACATGTGTGCCTG
AnCutA_opt	GACAGTGATGTGGACATTTGCATCACCACGACCTGTAAAGAACTGGAACATGTTTGCCTG
protein	D S D V D I C I T T T C K E L E H V C L
AnCutA	CTGGCGGATGTGCTGGCGAAAAACGGCATGGAACGCGTGGTGTGCATTAGCCATGCGAAA
AnCutA_opt	CTGGCGGATGTCTGGCCAAGAACGGCATGGAACGTGTGGTTTGTATCTCTCACGCGAAA
protein	L A D V L A K N G M E R V V C I S H A K
AnCutA	GTGCCGATTGTGAAAATTTGGGATCCGGAACGCGCCTGGCGTGCGATATGAACGTGAAC
AnCutA_opt	GTGCCGATTGTTAAGATCTGGGATCCGGAACGCGCCTGGCGTGTGACATGAACGTGAAC
protein	V P I V K I W D P E L R L A C D M N V N

AnCutA	AACACCATGGCGCTGGAAAACACCCGCATGGTGCGCACCTATGTGGAAATTGATGAACGC
AnCutA_opt	AATACCATGGCCCTGGAAAATACCCGTATGGTGCGCACGTATGTTGAAATTGATGAACGT
protein	N T M A L E N T R M V R T Y V E I D E R
AnCutA	GTGCGCCCGCTGGCGATGATTATTAAACATTGGACCAAACGCCGATCTCTGAACGATGCG
AnCutA_opt	GTTCGCCCGCTGGCGATGATTATCAAACATTGGACCAAGCGCCGTATCCTGAACGACGCC
protein	V R P L A M I I K H W T K R R I L N D A
AnCutA	GGCCTGGGCGGCACCCTGAGCAGCTATACCTGGATTGCTGATTATTAACTTTCTGCAG
AnCutA_opt	GGTCTGGGCGGTACCCTGAGTTCCTACACGTGGATTGCTGATTATCAATTTTCTGCAG
protein	G L G G T L S S Y T W I C L I I N F L Q
AnCutA	ACCCGCGAACCGCCGATTCTGCCGAGCCTGCAGGCGCGCCGCATAAAAAACGCCTGACC
AnCutA_opt	ACCCGTGAACCGCCGATCCTGCCGAGCCTGCAAGCACGTCCGCACAAAAAGCGCCTGACG
protein	T R E P P I L P S L Q A R P H K K R L T
AnCutA	GCGGATGGCCTGGTGTGCAGCTTTGATGATGATCTGGATAGCCTGGTGGGCTATGGCAAA
AnCutA_opt	GCCGATGGCCTGGTCTGTAGTTTCGATGACGATCTGGACTCCCTGGTGGGTTATGGCAAA
protein	A D G L V C S F D D D L D S L V G Y G K
AnCutA	CAGAACAAACAGAGCCTGGGCGAACTGTTTTTTCAGTTTTTTCGCTATTATGGCCATGAA
AnCutA_opt	CAGAATAAGCAATCTCTGGGTGAACTGTTTTTCCAGTTTTTCCGCTATTACGGCCATGAA
protein	Q N K Q S L G E L F F Q F F R Y Y G H E
AnCutA	CTGGATTTTGAAAAATATGTGATTAGCGTGCGCGAAGGCCGCCTGATTAGCAAAGAAGGC
AnCutA_opt	CTGGATTTCGAAAAGTACGTCATTAGCGTGCGTGAAGGTCGCCTGATCTCTAAAGAAGGT
protein	L D F E K Y V I S V R E G R L I S K E G
AnCutA	AAAGGCTGGCATCTGCTGCAGAACAACCGCCTGTGCGTGGAAGAACCGTTTAAACACCATT
AnCutA_opt	AAAGGTTGGCACCTGCTGCAGAACAATCGTCTGTGCGTGGAAGAACCGTTTAAACACCATT
protein	K G W H L L Q N N R L C V E E P F N T I
AnCutA	CGCAACCTGGGCAACACCGCGGATGATACCAGCTTTCGCGGCGTGCCATATGGAAGTGC
AnCutA_opt	CGCAACCTGGGTAATACGGCGGACGATACCAGTTTCCGTGGCGTTCATATGGAAGTGC
protein	R N L G N T A D D T S F R G V H M E L R
AnCutA	CGCGCGTTTAAAGCGCTGAGCGATGGCAACTTTGAACTGTGCTGCGAACAGTATGAATAT
AnCutA_opt	CGTGCAATTTAAAGCTCTGTCCGATGGCAATTTGAACTGTGCTGTGAACAGTATGAATAC
protein	R A F K A L S D G N F E L C C E Q Y E Y
AnCutA	CCGCCGGAAGCGGAACGCAGCTGGGAACGCCCGCGCCGAGCCGCGCCCGGTGGTGACC
AnCutA_opt	CCGCCGGAAGCAGAACGTTTCATGGGAACGCCCGGCTCCGCAACCGCGCCCGGTCTGTACG
protein	P P E A E R S W E R P A P Q P R P V V T
AnCutA	ACCGCGCCGCCCGGAGCCGCGGCGCGCGCGCGCGCGCGCGCGCGCCATAGCAAC
AnCutA_opt	ACCGCACCGCCCGCGTCCCGTGGCGGTGCGGCGGTGGCCGTGGTGGCCGCCATAGTAAC
protein	T A P P P S R G G R G G G R G G R H S N
AnCutA	CAGTATAGCCGCGCGGCCATAGCGGCGCGCCCGCAGCAGCAACCGCGCGAACAACAAAC
AnCutA_opt	CAGTATTCGCGTGGTGCCACAGTGGTGGCCCGGTTTCATCGAATGCGGCCAACAAAAT
protein	Q Y S R G G H S G G R R S S N A A N K N

AnCutA	CATTTTCGCCAGGCGAACAACGGCATGAGCGCGAGCGAACTGAGCCTGCAGGCGCAGCAT
AnCutA_opt	CATTTTCGCCAGGCGAACAATGGCATGTACAGCCTCGGAACTGAGCCTGCAGGCGCAACAC
protein	H F R Q A N N G M S A S E L S L Q A Q H
AnCutA	GCGCAGTATCTGCTGCATGATCATCTGTATCAGCAGATTGAGATTCTGCAGGCGCAGGAA
AnCutA_opt	GCCCAGTATCTGCTGCATGATCACCTGTACCAGCAAATTCAAATCCTGCAGGCACAAGAA
protein	A Q Y L L H D H L Y Q Q I Q I L Q A Q E
AnCutA	CAGGAACTGCGCCTGCAGCTGCAGAACCAGGCGCTGATTACCGGCCGCCGCCGCCGGTG
AnCutA_opt	CAGGAACTGCGTCTGCAACTGCAGAACCAAGCTCTGATTACCGGTCTCGGCCGCCGGTT
protein	Q E L R L Q L Q N Q A L I T G R P P P V
AnCutA	CTGATTGCCAGCCGTTTATTTCAGTTTCCGGTGCCGCGCAGGAAACCAGCGCGATGAT
AnCutA_opt	CTGATTGCTCAGCCGTTTATCCAATTCCCGGTCCCGGCACAGGAAACCAGCGGTGACGAT
protein	L I R Q P F I Q F P V P A Q E T S G D D
AnCutA	AGCAGCCGCGAGCCGCGAGCGCGAGCCGCGAGCGCGGTGAACCATCAGCCGATTAGCGTG
AnCutA_opt	AGCTCTCGTAGCCGCTCTGGTAGTCGTTCCGGCGCGGTGAATCATCAGCCGATCAGTGTG
protein	S S R S R S G S R S G A V N H Q P I S V
AnCutA	CCGAGCCGCCAGAACGTGTATTATAACAGCACCTATCTGCCGGTGGCGGTGCCGGGCGGC
AnCutA_opt	CCGTCCCGCCAAAACGTTTATTACAATTCAACCTATCTGCCGGTTCGAGTGCCGGGTGGC
protein	P S R Q N V Y Y N S T Y L P V A V P G G
AnCutA	CCGGGCGACAAACCAACCCGCCGAGCCCGAGCGCGGCGGCGGCGATGCCGGATCTGCGC
AnCutA_opt	CCGGGTTCGAACACGAATCCGCCGTACCGTTCGGCAGCTGCAGCAATGCCGGATCTGCGC
protein	P G S N T N P P S P S A A A A M P D L R
AnCutA	CGAACCCGCGCCGCGAGCAGCGTGCGGAACGGCAGCCGAACGGCAGCCTGCGCGCGCAT
AnCutA_opt	CGTAACCCGCGCCGTAGTTCCGTGGCCAACGGTTACCGAATGGCTCGCTGCGTGACAT
protein	R N P R R S S V A N G S P N G S L R A H
AnCutA	AGCCAGCCGGCGCGCCCGGTGAACAGCCTGCCGAACCTTTGCGCCGCTGTATAGCATTAGC
AnCutA_opt	AGTCAGCCGGCACGTCCGGTTAACTCCCTGCCGAATTTGCGCCGCTGTACAGCATTCTCT
protein	S Q P A R P V N S L P N F A P L Y S I S
AnCutA	CAGCCGACCGAAACCAGCCAGAGCAGCAAACAGCGCCATCCGAGCGGCGGAGCGATGAT
AnCutA_opt	CAACCGACGGAACCAGCCAGTCATCGAAACAACGCCACCCGTCAGGTGCCTCGGACGAT
protein	Q P T E T S Q S S K Q R H P S G A S D D
AnCutA	AGCCTGAGCGAAGAAGATGGCACCAGCATGCCGAACAGCCTGCCGAACCCGACCCGTGTTT
AnCutA_opt	AGTCTGTCCGAAGAAGATGGCACCAGCATGCCGAACCTCTCTGCCGAATCGTACGCTGTTT
protein	S L S E E D G T S M P N S L P N R T L F
AnCutA	ATTGATGAAGCGCGCAGCAACGATTATATGACCTATTATTTTGGCGGCAGCCGCGAGCTG
AnCutA_opt	ATCGATGAAGCGCGCAGTAACGACTATATGACCTATTACTTCGGTGGCTCCCCGCGAGCTG
protein	I D E A R S N D Y M T Y Y F G G S P Q L
AnCutA	CATGCGTATCATCAGAACCGCATGCTGAGCCCGATGAGCGCGAGCCTGGGCCTGGCGCTG
AnCutA_opt	CATGCATACCACAAAATGCTATGTGTCTCCGATGTCAGCATCGTGGGTCTGGCACTG
protein	H A Y H Q N A M L S P M S A S L G L A L

AnCutA	CCGAACGGCGGCATTGTGCCGTGGGTGGGCAACCCGAGGAATATATGAGCGCGGCGCCG
AnCutA_opt	CCGAACGGTGGCATTGTTCCGTGGGTCCGCAATCCGAGGAATATATGAGTGCAGTCCG
protein	P N G G I V P W V G N P Q E Y M S A A P
AnCutA	GGCGCGGAAAGCTATGGCCATAGCTTTGATAACGGCAGCATTACCAGCCGAAAAGCGGC
AnCutA_opt	GGTGCAGGAAAGTTACGGCCATTCTTTGATAATGGCAGCATCACCTCTCCGAAAAGTGGC
protein	G A E S Y G H S F D N G S I T S P K S G
AnCutA	CAGAGCCAGCAGCGCATGGGCAGCCGACCGCGGCGAGCAGCGATCGCGGCGGCCCGCTG
AnCutA_opt	CAGTCCCAGCAACGTATGGGTAGCCGTACGGCAGCAAGCTCTGATCGTGGTGGCCCGCTG
protein	Q S Q Q R M G S R T A A S S D R G G P L
AnCutA	ATTGTGGATGGCAGCGTGCCGCCGAGCGAACAGCGCATGAGCATTCCGGCGGATCATTAT
AnCutA_opt	ATTGTGACGGTTCAGTGCCGCCGTCGGAACAGCGCATGAGCATCCCGGCAGATCATTAT
protein	I V D G S V P P S E Q R M S I P A D H Y
AnCutA	GAACATTATGGCCTGATGAGCCATTGCCCGAGCAACAGCGATGATCAGAACACCGGATACC
AnCutA_opt	GAACACTACGGCCTGATGTCTCACTGTCCGAGCAACTCTGACGATCAGAATACGGATACC
protein	E H Y G L M S H C P S N S D D Q N T D T
AnCutA	CCGCGAGCGTGAGCGATAGCTTTAGCCAGGATTTTCAGGATAGCAGCAGCGTGGAATTT
AnCutA_opt	CCGGCCAGTGTCTCCGACTCATTTTCGCGAGGATTTCCAAGACAGTTCCTCAGTGGAATTT
protein	P A S V S D S F S Q D F Q D S S S V E I
AnCutA	GATCCGAGCCCGTTTTTTAGCCGCCAGGGCTTTGATAGCCATGCGATGGGCGGCCCGCGC
AnCutA_opt	GATCCGAGCCCGTTTTTCTCTCGTCAGGGTTTTGATAGCCATGCAATGGGTGGCCCGGCT
protein	D P S P F F S R Q G F D S H A M G G P A
AnCutA	GATGTGTTTGTGAACGGCCATAGCGGCAAAACAGAGCCTGCTGAGCAGCCGCCTGCAGGGC
AnCutA_opt	GACGTTTTCTGCAACGGTCACAGTGGCAAAACAGTCCCTGCTGTGCGAGCCGCCTGCAAGGT
protein	D V F V N G H S G K Q S L L S S R L Q G
AnCutA	CTGCATCTGAGCAACGGCGAAAAAGTGGATATTGCGAGCAAAAGCGCGCAGGATAAACCG
AnCutA_opt	CTGCATCTGTCTAATGGCGAAAAAGTGGATATTGCAAGCAAGTCTGCTCAGGACAAACCG
protein	L H L S N G E K V D I A S K S A Q D K P
AnCutA	AAAGCGGTGGTGCCGCATAACGATGCGGTGGAACGCGAAGCGCCGCGCGCGAAAAACCAT
AnCutA_opt	AAGGCAGTTGTCCCGCACACGATGCTGTGGAACGTGAAGCACCGCGCGCTAAAAATCAT
protein	K A V V P H N D A V E R E A P R A K N H
AnCutA	CATGCGGTGGATAAAGGCCCGAGCGCGAGCAGCTTTAGCCATGCGGCGGATAGCCATATT
AnCutA_opt	CACGCGGTTGATAAGGGCCCGAGCGCCTCTAGTTTTTCTCATGCAGCTGACTCACACATC
protein	H A V D K G P S A S S F S H A A D S H I
AnCutA	GCGAGCCCGAACGGCGCCGCCGCCATCATGGCAGCGATACCAGCGAAAAAGTGAACGGC
AnCutA_opt	GCATCGCCGAACGGTCGCCGTGCCATCACGGCTCAGATACCTCGGAAAAAGTTAACGGT
protein	A S P N G R R R H H G S D T S E K V N G
AnCutA	ACCGCGCATAAAGGCAAAACCGCGCAACCAGACCAGCAGCACCGCGAGCAACGATCAGAAA
AnCutA_opt	ACGGCACATAAAGGCAAGCCGCGTAATCAGACGTCTCAACCGCAAGCAACGATCAAAAA
protein	T A H K G K P R N Q T S S T A S N D Q K

AnCutA
AnCutA_opt
protein
GAACGCAACGCGGGCGGCCGCGCAAACCGAACGCGGCGGCCTGACCCCGCAGGATAGC
GAACGTAATGCGGGTGGCCCCGCTAAGCCGAATGCAGCAGGTCTGACCCCGCAGGACAGC
***** ** ***** ***** ** ***** ** ** ***** ***** ***
E R N A G G P R K P N A A G L T P Q D S

AnCutA
AnCutA_opt
protein
AACCATAGCCATGCGAACAGCGGCTGGCAGACCACCAAACGCAAAAAACAAAAAACAAC
AACCATTCTCACGCGAATTCAGGCTGGCAAACGACCAAGCGTAAGAACAAAAAGAACAAC
***** ** ***** ***** ** ***** ** ** ***** *****
N H S H A N S G W Q T T K R K N K K N N

AnCutA
AnCutA_opt
protein
AAACTGTTTAGCGATCATCGCCATGGCCTGCAGAACGGCGCGGAACCGCTGCCGATTGAT
AAGCTGTTCTCGGATCATCGCCACGGTCTGCAGAATGGCGCCGAACCGCTGCCGATTGAC
** ***** ***** ** ***** ***** *****
K L F S D H R H G L Q N G A E P L P I D

AnCutA
AnCutA_opt
protein
GAAAGCCTGCGCAAAGGCGGC
GAAAGCCTGCGCAAAGGTGGC
***** ***** ***
E S L R K G G

TtCutA
TtCutA_opt
protein
ATGGATAGCCAGCCGCGACCGCGGAAAGCAAACATCATTATCAGAGCTATCCGTGGCCG
ATGGACTCGCAACCGGCAACCGCAGAATCAAAGCATCACTACCAATCATATCCGTGGCCG
***** ** ***** ***** ** ***** ** ** ***** *****
M D S Q P A T A E S K H H Y Q S Y P W P

TtCutA
TtCutA_opt
protein
GGCGCGCCGAGCGCGGGCACCACCAGCCGCGTGAAACCGGCGCCGGTGCCGGAAGCAGC
GGCGCTCCGTCAGCAGGCACCACCAGCCGCGTGAAACCGGCGCCGGTCCGGAAGCTCT
***** ** ***** ***** ***** ***** *****
G A P S A G T T S R V K P A P V P E S S

TtCutA
TtCutA_opt
protein
CCGGGCTATTTTCCGGCGCAGCTGGATCTGCTGCCGATTCTGTTTCCGCAGCAGCTGTAT
CCGGGTATTTTCCGGCCCGAGCTGGATCTGCTGCCGATTCTGTTTCCGCAGCAACTGTAT
***** ***** ***** ***** ***** *****
P G Y F P A Q L D L L P I L F P Q Q L Y

TtCutA
TtCutA_opt
protein
CAGGATAAACTGCTGCAGTATAACAACTGGTGAGCGCGGGCCGCGGCGGCGGCGGCGGC
CAAGACAAGCTGCTGCAGTACAACAACTGGTCAGCGCAGGTCGTGGCGGTGGCGGTGGC
** ** ***** ***** ***** ** ** ***** ***** ***
Q D K L L Q Y N K L V S A G R G G G G

TtCutA
TtCutA_opt
protein
CTGCAGACCCCGCCGATCCGCCGTGCTGAGCAGCCCGCAGAGCAAAAGCAGCACCCGC
CTGCAGACGCCCGCCGACCCGCCGTTCTGAGTTCCTCGCAGAGTAAATCATCGACCCGC
***** ***** ***** ***** ***** *****
L Q T P P H P P V L S S P Q S K S S T R

TtCutA
TtCutA_opt
protein
AGCCGCAGCAGCAGCAAAGTGAGCAACAAAGATAACAGCGCGCAGGCGAAAAGCGCGCCG
TCCCGTAGCTCTAGTAAGGTGTCTAACAAAGATAATTCGCCCCAAGCAAAAAGCGCGCCG
** ** ***** ***** ***** ***** *****
S R S S S K V S N K D N S A Q A K S A P

TtCutA
TtCutA_opt
protein
CAGAGCTGCCATGGCAACAAACCGGCGAAAAACAGCGAACCGGCGGAACGCGCGCTGATG
CAGTCTTGCCATGGTAACAAGCCGCGCAAAAATTCAGAACCGGCAGAACGCGCTCTGATG
*** ***** ***** ***** ***** *****
Q S C H G N K P A K N S E P A E R A L M

TtCutA
TtCutA_opt
protein
AAAAGCGGCAAAGAACCGAGCAGCAAAATGAGCGCGAAAAACCGAGCGAACAGGCGGTG
AAGTCGGGCAAAGAACCGTCCTCAAAAATGTGAGCAAAAAGCCGTCGGAACAGGCTGTC
** ***** * ***** ** ***** ** *****
K S G K E P S S K M S A K K P S E Q A V

TtCutA	GATGGCGATGCGGCGGTGCGCCAGCCGCCGGCGAGCCGCCAGGGCGCGCCGGCGACCGCG
TtCutA_opt	GATGGTGACGCGGCGGTGCGTCAACCGCGGCATCACGTGAGGTGCTCCGGCAACCGCC
protein	D G D A A V R Q P P A S R Q G A P A T A
TtCutA	ACCCAGAGCAGCAGCGTGCCGAGCACCCCGCATCAGCATCCGCGCAAATTTAGCTTTGAA
TtCutA_opt	ACGCAATCGAGCTCTGTTCCGTCGACCCCGCATCAGCACCCGCGCAAATTTTCCTTCGAA
protein	T Q S S S V P S T P H Q H P R K F S F E
TtCutA	AGCCGCGAACAGAGCCCGGGCGGCGCAGAACCATAGCCGCGCAGCGGTATAGCGAA
TtCutA_opt	TCAGTGAACAAAGTCCGGGTGCAGCTCAGAACCATTCGCCGCGCAGCGCTATTCTGAA
protein	S R E Q S P G A A Q N H S P R S A Y S E
TtCutA	ACCAACGGCAACGTGCCGAGCCTGCGCCCGCTGCCGCCGCGCTGGGCGGCTGCCGCTTT
TtCutA_opt	ACCAACGGTAATGTGCCGAGCCTGCGTCCGCTGCCGCCGCGTCTGGGTGGCTGTCGCTTC
protein	T N G N V P S L R P L P P R L G G C R F
TtCutA	GAAACCGGATTCGGTTTAGCCGCCGCGCATGCCGTATAGCCTGGGCACCGATAAACTG
TtCutA_opt	GAAACGGCAATTCCGTTTAGCCGTGCGCGTATGCCGTACTCTCTGGGCACCGATAAGCTG
protein	E T A I P F S R R R M P Y S L G T D K L
TtCutA	GAAAAAGTGGATCCGGATAAAATTAAAAGCAAACCTGAGCGAAGATGTGGAACGCAAACCTG
TtCutA_opt	GAAAAAGTGGATCCGGACAAAATCAAGAGTAAACCTGCCGAAGACGTGGAACGTAAACCTG
protein	E K V D P D K I K S K L S E D V E R K L
TtCutA	GAAACCGATATGCGCGAACTGTATGATGCGCTGCTGCCGACCGAAGCGATTGAAGTGAAC
TtCutA_opt	GAAACGGATATGCGTGAACCTGTATGACCGCTGCTGCCGACCGAAGCGATTGAAGTTAAC
protein	E T D M R E L Y D R L L P T E A I E V N
TtCutA	CGCCGCGAACTGGTGAGCAAACCTGGAACGCCTGTTTAAACACCGAATGGCCGGGCCATGAT
TtCutA_opt	CGCCGTGAACCTGGTCAGCAAACCTGGAACGCCTGTTCAATACCGAATGGCCGGGTATGAT
protein	R R E L V S K L E R L F N T E W P G H D
TtCutA	ATTCGCGTGATCTGTTTGGCAGCAGCGGCAACCTGCTGTGTCAGCGATGATAGCGATGTG
TtCutA_opt	ATCCGTGTTACCTGTTTGGTAGTTCCGGCAATCTGCTGTGCTCAGATGACTCGGATGTT
protein	I R V H L F G S S G N L L C S D D S D V
TtCutA	GATATTTGCATTACCACCCCGTGGCGCGAACTGGAAAGCGTGTGCATGATTGCGGAACTG
TtCutA_opt	GACATTTGTATCACCACGCCGTGGCGCGAACTGGAAAGCGTGTGCATGATTGCAGAACTG
protein	D I C I T T P W R E L E S V C M I A E L
TtCutA	CTGGATCGCCATGGCATGGAAAAAGTGGTGTGCGTGAGCAGCGCGAAAGTGCCGATTGTG
TtCutA_opt	CTGGATCGTCATGGCATGGAAAAGTGGTTTGTGTGTCATCGGCGAAGGTGCCGATTGTT
protein	L D R H G M E K V V C V S S A K V P I V
TtCutA	AAAATTTGGGATCCGGAACCTGAAACTGGCGTGCGATATGAACGTGAACAACACCCCTGGCG
TtCutA_opt	AAAATCTGGGATCCGGAACCTGAAACTGGCATGCGACATGAACGTTAACAATACCCCTGGCT
protein	K I W D P E L K L A C D M N V N N T L A
TtCutA	CTGGAAAACACCCGCATGGTGCACCTATGTGAGCATTGATGATCGCGTGCGCCCGCTG
TtCutA_opt	CTGGAAAATACCCGCATGGTCCGTACGTATGTGAGCATTGATGACCGCGTGCCTCCGCTG
protein	L E N T R M V R T Y V S I D D R V R P L

TtCutA	GCGATGATTATTAAATATTGGACCCGCCGCCGCGTGGTGAACGATGCGGCGTTTGGCGGC
TtCutA_opt	GCGATGATTATCAAATACTGGACCCGCCGTCGCGTCGTGAACGATGCAGCCTTCGGTGCC
protein	A M I I K Y W T R R R V V N D A A F G G
TtCutA	ACCCTGAGCAGCTATACCTGGATTTGCATGATTATTGCGTTTCTGCAGCTGCGCGATCCG
TtCutA_opt	ACCCTGAGCTCTTACACGTGGATCTGTATGATTATCGCATTTCTGCAGCTGCGTGATCCG
protein	T L S S Y T W I C M I I A F L Q L R D P
TtCutA	CCGGTGCTGCCGGCGCTGCATCAGCAGCATGATCTGAAACTGGTGAACAGGATGGCGCG
TtCutA_opt	CCGGTCCTGCCGGCTCTGCATCAGCAACACGACCTGAAGCTGGTGAACAGGATGGTGCG
protein	P V L P A L H Q Q H D L K L V K Q D G A
TtCutA	CTGAGCGATTTTGGCGATGATATTCGAAACTGCGCGGCTTTGGCGCGAAAAACAAAGAT
TtCutA_opt	CTGTCTGACTTCGCCGATGACATTCGAAACTGCGTGGTTTTGGCGCGAAGAATAAAGAT
protein	L S D F A D D I P K L R G F G A K N K D
TtCutA	AGCCTGGCGGTGCTGCTGTTTCAGTTTTTTCGCTTTTATGCGCATGAATTTGATTATGAT
TtCutA_opt	AGTCTGGCCGTGCTGCTGTTTCAGTTTTTTCGCTTCTATGCACACGAATTTGATTATGAC
protein	S L A V L L F Q F F R F Y A H E F D Y D
TtCutA	AAATATACCCTGAGCATTCGCATGGGCACCCTGCTGACCAAAGCGGAAAAAACTGGCAG
TtCutA_opt	AAATACACCCTGAGCATCCGTATGGGCACCCTGCTGACGAAGGCAGAAAAAGAACTGGCAG
protein	K Y T L S I R M G T L L T K A E K N W Q
TtCutA	TATCTGGTGAACAACGCGCTGTGCGTGGAAGAACCGTTTAACGATGGCCGCAACCTGGGC
TtCutA_opt	TACCTGGTGAACAATGCTCTGTGCGTTGAAGAACCGTTTAATGATGGTGCACACCTGGGC
protein	Y L V N N A L C V E E P F N D G R N L G
TtCutA	AACACCGCGATGAAACAGCTTTTCGCGCCTGCATATGGAACGCGCCGCGCGTTTGAT
TtCutA_opt	AATACCGCCGACGAAACGAGCTTCCGTGGTCTGCATATGGAACGCGTCGCGCATTTGAT
protein	N T A D E T S F R G L H M E L R R A F D
TtCutA	CTGATTGCGGAAGGCAAACTGGAAGAATGCTGCGAACAGTATGTGTTCCGAAAGAAGAA
TtCutA_opt	CTGATTGCTGAAGGCAAACTGGAAGAATGCTGTGAACAGTATGTGTTCCGAAAGAAGAA
protein	L I A E G K L E E C C E Q Y V F P K E E
TtCutA	GAACGCGTGTGGCAGAAACCGCCGCCGCCGCCGCGCCGATTCCTGCTGCGCAGCAGCAGC
TtCutA_opt	GAACGCGTTTGGCAGAAACCGCCGCCGCCGCCGCCGCGTCCGATCCTGCTGCGCAGTTCCCTCA
protein	E R V W Q K P P P A P R P I L L R S S S
TtCutA	CAGCAGCATGGCAACAGCGGCAACCGCCTGCGCGGCGGCGGCTTTGCGGCAACCGCCAG
TtCutA_opt	CAGCAACACGGTAACAGCGGCAATCGCCTGCGTGGTGGCGGTTTCCGCGGTAACCGTCAG
protein	Q Q H G N S G N R L R G G G F R G N R Q
TtCutA	TTTCATCGCAACGCGGCGCGGGGCCGCCGAGCAGCAGCAGCGTGACCTATGATCCGAAC
TtCutA_opt	TTTCATCGTAATGGCGGTGCAGGTGCGTTCGCTCGAGCTCTAGTGTACGTATGATCCGAAC
protein	F H R N G G A G R R S S S S V T Y D P N
TtCutA	AGCGTGTATAGCCAGGCGGGCATGCCGCCGACCATGAGCCCGCAGGATCTGCTGTGGTAT
TtCutA_opt	AGCGTGTACTCTCAAGCCGGTATGCCGCCGACCATGAGCCCGCAGGATCTGCTGTGGTAT
protein	S V Y S Q A G M P P T M S P Q D L L W Y

TtCutA CAGGCGCAGAGCCCGCAGCTGAGCGTGCCGCGAGGAAATTCTGACCACCAGCCTGAACGCG
TtCutA_opt CAAGCTCAGAGTCCGCAACTGTCCGTGCCGCGAGGAAATTCTGACCACGTCACTGAACGCG
** ** ***** ** *****
protein Q A Q S P Q L S V P Q E I L T T S L N A

TtCutA CTGGCGGCGCAGGAAAGCATGCGCTTTCAGCTGTATACCCAGCTGAACCAGCAGCAGGTG
TtCutA_opt CTGGCAGCTCAGGAATCGATGCGCTTTCAGCTGTACACCCAGCTGAATCAGCAACAGGTT
***** ** *****
protein L A A Q E S M R F Q L Y T Q L N Q Q Q V

TtCutA CTGGCGCATGCGCAGCGCGTGCGAGCGGGCGCGAGCACCACCGATCGCAGCCGCACCAAC
TtCutA_opt CTGGCACACGCTCAGCGTGTCCAACAGGGCGCCAGCACCACGGATCGCTCACGTACGAAC
***** ** ** ***** ** *****
protein L A H A Q R V Q Q G A S T T D R S R T N

TtCutA AGCCTGGATAACCCGCGCTGACCGCGCCGGTGCGCCCGGAACCTGATGTATGGCATTGGC
TtCutA_opt TCAGCTGGACAATCCGCGCTGACCGCACCGGTTCTGCCGAACCTGATGTACGGTATTGGC
***** ** *****
protein S L D N P P L T A P V R P E L M Y G I G

TtCutA TTTCCGATTTCAGCATGCGCCGTATTTTCATCCGGGCTTTACCACCTATCCGAGCAGCCCG
TtCutA_opt TTCCCGATCCAGCATGCGCCGTATTTTCACCCGGGTTTACCACGTACCCGTCCTCACCG
** *****
protein F P I Q H A P Y F H P G F T T Y P S S P

TtCutA GCGACCGCGCGCCGCCGAGCAACGCCAGCCGGAATTTGCGCCGAGCCTGCATCGCACC
TtCutA_opt GCAACCGCAGCACCGCCGAGCAACGCCAGCCGGAATTTGCTGCTCTCTGCATCGTACC
** *****
protein A T A A P P S N G Q P E F R R S L H R T

TtCutA GCGGCGGCGAGCGATAGCGGCGTGAGCCCGGCGAGCGGCGGCTGCGCAGCCAGAGCCAG
TtCutA_opt GCAGCTGCGTCAGATTGCGGTGTGAGTCCGGGTTCGGGTGCACTGCGTAGCCAATCTCAG
** ** ** ** **
protein A A A S D S G V S P G S G A L R S Q S Q

TtCutA CCGGCGAGCCGCACCACCCTGCCGCCGATAGCATGGCGGTGTATAGCACCAGCAGCCAG
TtCutA_opt CCGGCTTCTCGTACCACGCTGCCGCCGATAGCATGGCGGTTTATTCTACCTCGAGCCAG
***** ** *****
protein P A S R T T L P P H S M A V Y S T S S Q

TtCutA AGCCATACCGGCATTGGCGCGAGCCAGAGCCGCCAGGTGCATGATGTGGCGGTGCCGAGC
TtCutA_opt AGTCACACGGGTATCGGCGCGAGTCAATCCCGCCAGGTGCATGATGTGCGGTCCCGAGC
** ** ** **
protein S H T G I G A S Q S R Q V H D V A V P S

TtCutA TTTATGCCGATGAACGCGGCGATCTGGGCTTTGATGATGTGCCGACAAACCGATTAGC
TtCutA_opt TTCATGCCGATGAACGTGGTGACCTGGGCTTTGATGACGTTCCGACAAACCGATTTC
** *****
protein F M P D E R G D L G F D D V P T K P I S

TtCutA GATAGCCCGCCGAAAACGATGGCCCGCGCTTTATTGGCTATTATGTGAACGCCATAGC
TtCutA_opt GATTCGCCCGCCGAAAATGACGGTCCGCGCTTCATCGGCTATTACGTGAACGGTCACAGC
*** *****
protein D S P P E N D G P R F I G Y Y V N G H S

TtCutA GGCAGCAGCCGCCCGGCGAACGGCTTTCGAGCGTGGGCGTGCGGAGCGCGGAAAGCAGC
TtCutA_opt GGCTCTAGTCGTCCGGCCAATGGTTTCCGTCCGTGGGCGTTGCGTCAGCCGAATCTCTCA
*** ** ** *****
protein G S S R P A N G F P S V G V A S A E S S

TtCutA CAGAGCGGCCAGGGCGGCGGCCCGCCCGCTGAGCACCGATCAGCTGCCGCGAGCGGTG
TtCutA_opt CAATCTGGTCAGGGCGGTGGCCGTCCGCGTCTGAGTACGGATCACTGCCGCGAGTCCGTG
** *****
protein Q S G Q G G G R R R L S T D Q L P Q S V

TtCutA CTGGATCGCCGCATGAAACGCACCAGCCGAGCCCGATGGGCCATGCGCGCGCG
TtCutA_opt CTGGACCGCCGTATGAAACGCACCAGCCGTTCTCCGAGTCCGATGGGTACATGCACGTGCT

protein L D R R M K R T S R S P S P M G H A R A

TtCutA CTGAGCGTGGGCGTGAACAGCGCGCCGCTGCTGAGCGCGCCGTTTGCAGAGCAACGGC
TtCutA_opt CTGTCCGGTCGGCGTGAACAGCGCACCGCTGCTGTCTGCACCGTTTGCACAGTCAAACGGC
*** **
protein L S V G V N S A P L L S A P F A Q S N G

TtCutA AAAGTCCGAGCAAACCGCTGGTGGTGAACGGCACCGCGCGGAAACCGCGAGCACCCCG
TtCutA_opt AAGCTGCCGTCGAAACCGCTGGTTGTCAATGGTACCGCACCGGAAACCGCAAGCACGCCG
** *****
protein K L P S K P L V V N G T A P K T A S T P

TtCutA GGCAGCGGCCGCCCCACCGTGGTGACCGAAAGCCTGGTGAGCGATGATAGCCTGCATGAT
TtCutA_opt GGTCTGGTTCGTCCGACCGTGGTTACGGAATCACTGGTGTCCGATGACAGCCTGCATGAT
** **
protein G S G R P T V V T E S L V S D D S L H D

TtCutA AACGCGCTGCATATTCATCAGGGCGTGAGCCCGAGCAGCAGCTGGCCGGAACAGCAGCCG
TtCutA_opt AATGCCCTGCATATTCACAGGGGTGTTTCCCCGTCGAGCTCTTGCCCGGAACAACAGCCG
** *
protein N A L H I H Q G V S P S S S W P E Q Q P

TtCutA ACCCGCCGAGCAGCGCGCGGAAAGCGTGGCGACCCGCGTGCCGGAACGCCCGGTGATT
TtCutA_opt ACGCCGCCGAGTTCGCGAGCAGAAAGCGTGGCAACCCGCTGTCCCGGAACGTCCGGTTATC
** *****
protein T P P S S A A E S V A T R V P E R P V I

TtCutA GTGAACGGCAGCAGCACCAACCGCAGCCCGAGCGCGACCGTGCCGAACCTGCTGGATAACC
TtCutA_opt GTCAACGGCTCATCGACCAATCGCTCCCCGTCAGCCACGGTTCGGAACCTGCTGGATAACC
** *****
protein V N G S S T N R S P S A T V P N L L D T

TtCutA AGCTTTCAGCAGCGCTGGCGATGGGCGCGGCGCTGCATCTGCCGTATAGCGCGGTGGTG
TtCutA_opt AGCTTCCAACAGCGTGTGCGAATGGGTGCTGGCCTGCACCTGCCGTATAGTGGCGTCTGTG

protein S F Q Q R V A M G A G L H L P Y S A V V

TtCutA GGCGATCCGAACGGCCTGGCGGGCTTTAGCCGCGCGCATACCCGCCAGCAGCTGGCGCCG
TtCutA_opt GGTGATCCGAATGGTCTGGCGGGCTTTCCCGCGCCCATACGCGTCAACAGCTGGCACCG
** *****
protein G D P N G L A G F S R A H T R Q Q L A P

TtCutA CTGGATCTGGCGACCGGCGATTATGCGGCGCATCAGGATCTGCCGATCTGAGCCCGGTG
TtCutA_opt CTGGATCTGGCTACCGGCGACTATGCTGCGCATCAGGATCTGCCGACCTGAGTCCGGTC

protein L D L A T G D Y A A H Q D L P H L S P V

TtCutA TATGAACATCGCACCCCGAGCCCGACCGTGCCGCGCAAATTTGAACCGAGCCCGATTAGC
TtCutA_opt TACGAACACCGTACCCCGTCCCCGACGGTGCCGCGTAAGTTTGAACCGTCTCCGATTAGT
** *****
protein Y E H R T P S P T V P R K F E P S P I S

TtCutA CAGCTGAGCGGCGCGGGCGGAAAGATTTTCGCGAGCGATTTTAGCCCGAAAGCGACCCAG
TtCutA_opt CAGCTGTCCGGTGGCGGCGGAAAGATTTTCGCTCAGACTTCTCGCCGAAGGCAACCCAG

protein Q L S G A G A K D F R S D F S P K A T Q

TtCutA AAAGCGCCGAGCAGCCGAGCCCGAACGGCGGCGCGCCGACCCGCCAGCAGGATAGCGCG
TtCutA_opt AAAGCTCCGAGCTCTCGTAGCCGAATGGTGGCGCACCGACGCTCAACAGGATTCTGCA

protein K A P S S R S P N G G A P T R Q Q D S A

TtCutA	ACCAGCCGCAGCCCGGCGCAGGATCAGCGCATTAACGGCCTGGCGCGCGAAAAACGGCCAT
TtCutA_opt	ACCTCGCGTAGCCCGGCACAAGACCAGCGTATCAACGGTCTGGCGCGCGAAAAATGGCCAT
	*** ** ***** ** ** ***** ** ***** ***** *****
protein	T S R S P A Q D Q R I N G L A R E N G H
TtCutA	GTGCGCGCGGCGAAAAAGCCAGACCGAAAGCTTTAGCGGCTGGCAGAAAACCAAACCCGC
TtCutA_opt	GTGCGTGCCGCAAAAAGTCAAACCGAATCTTTCAGTGGTTGGCAGAAGACCAAACGCGC
	***** ** ** ***** ** ***** ** ** ** ***** ***** **
protein	V R A A K S Q T E S F S G W Q K T K T R
TtCutA	AAAAAAGGCGGCGCGGCGGCGGCGGAACTGAAACAGAGCAGCGCGGCGAACGGCTTTCCG
TtCutA_opt	AAAAAGGGTGGCGCTGCAGCAGCAGAACTGAAACAAAGTTCCGCTGCGAACGGCTTTCCG
	***** ** ***** ** ** ** ***** ***** ** *** *****
protein	K K G G A A A A E L K Q S S A A N G F P
TtCutA	CAGAGCGAACCGCCGCCGAAAAACGAAGCGGATCGCAAAGGCGGC
TtCutA_opt	CAGAGCGAACCGCCGCCGAAGAATGAAGCGGATCGTAAAGGTGGC
	***** ***** ** ***** ***** ***** **
protein	Q S E P P P K N E A D R K G G

SUPPLEMENTAL FIGURE 1. Alignments of the native (AnCutA/TtCutA) and recoded (AnCutA_opt/TtCutA_opt) open reading frames corresponding to AnCutA and TtCutA proteins. Nucleotides unchanged during recoding are marked with asterisks. Amino acid sequences of the proteins are shown in bold.

SUPPLEMENTAL TABLES

SUPPLEMENTAL TABLE 1. Frequency of nucleotide incorporation into the tails synthesized by full-length wild-type AnCutA and TtCutA depending on the identity of 1 or 2 preceding nucleotides at the primer's end. Higher and lower frequencies, as compared to the average values (marked in grey), which are discussed in the main text, are indicated with green and red colors, respectively.

Primer's end	AnCutA								TtCutA							
	ss22-A ₄				ss22-U ₄				ss22-A ₄				ss22-U ₄			
	A	G	C	U	A	G	C	U	A	G	C	U	A	G	C	U
Average frequency of nucleotide incorporation [%]																
*	30.9	0.5	52.8	15.9	20.2	1.0	65.7	13.1	48.3	0.5	43.5	7.6	36.7	1.1	56.3	5.9
Frequency of nucleotide incorporation depending on <u>1</u> preceding nucleotide																
*A	42.2	0.3	47.0	10.5	28.0	0.6	63.0	8.4	58.5	0.4	36.1	4.9	45.9	0.7	50.1	3.3
*G	28.0	4.0	50.3	17.7	18.6	1.5	68.3	11.6	39.4	2.2	45.8	12.5	30.9	2.8	59.8	6.5
*C	25.1	0.6	52.5	21.8	22.6	1.4	55.7	20.3	39.5	0.7	48.6	11.3	33.7	1.3	56.9	8.0
*U	14.6	0.4	76.6	8.4	11.1	0.7	80.8	7.4	18.4	0.5	76.0	5.1	19.2	0.9	75.4	4.5
Frequency of nucleotide incorporation depending on <u>2</u> preceding nucleotides																
*AA	48.9	0.3	39.4	11.5	31.3	0.4	61.8	6.5	60.4	0.4	33.9	5.3	47.9	0.7	48.5	2.9
*AG	28.7	2.9	53.4	15.1	15.7	3.1	73.1	8.1	42.5	1.4	43.8	12.2	30.7	2.7	59.7	6.9
*AC	25.9	0.3	56.0	17.8	18.6	0.6	65.5	15.4	39.2	0.5	50.9	9.4	29.9	0.9	62.6	6.6
*AU	20.3	0.2	71.5	7.9	11.4	0.3	83.9	4.4	21.2	0.4	73.7	4.7	19.2	0.8	76.6	3.4
*GA	26.5	2.4	62.7	8.4	18.1	2.0	72.7	7.2	55.2	1.5	36.8	6.5	53.4	0.9	42.4	3.4
*GG	19.2	32.7	9.6	38.5	34.9	14.0	41.9	9.3	23.2	27.0	29.5	20.4	27.2	6.6	57.9	8.3
*GC	24.5	0.8	48.4	26.3	20.6	1.6	56.4	21.4	37.7	1.5	47.8	13.0	41.0	1.5	52.9	4.6
*GU	14.7	6.1	63.2	16.0	15.8	3.7	66.6	13.9	20.5	3.7	68.3	7.4	24.2	3.0	67.1	5.7
*CA	32.1	0.3	59.6	8.1	28.8	0.7	61.5	9.0	53.2	0.4	42.4	4.1	44.0	0.8	51.8	3.4
*CG	26.7	1.4	55.6	16.3	24.2	1.1	62.7	12.0	37.2	1.3	49.7	11.8	31.6	2.6	60.2	5.6
*CC	25.1	0.6	49.5	24.8	17.7	1.5	53.7	27.0	39.1	0.7	47.7	12.4	29.2	1.6	60.4	8.7
*CU	12.1	0.4	79.4	8.1	10.5	0.7	81.3	7.5	16.7	0.4	77.7	5.2	18.9	0.9	75.4	4.8
*UA	31.4	0.4	58.2	10.0	23.6	0.6	67.5	8.3	51.6	0.4	42.7	5.3	47.6	0.7	47.5	4.2
*UG	25.6	11.0	43.9	19.5	15.0	1.9	68.4	14.8	33.0	7.3	43.2	16.5	31.1	3.4	58.3	7.2
*UC	22.9	0.9	55.1	21.2	22.9	1.8	53.9	21.5	40.6	0.9	44.8	13.7	44.1	1.5	42.9	11.5
*UU	13.9	0.8	76.6	8.8	11.7	1.0	78.4	8.9	17.8	0.9	75.2	6.1	17.2	1.0	77.3	4.5

SUPPLEMENTAL TABLE 2. Plasmids used in this study.

Plasmid	Source or genotype
pET28M N-6xHis-SUMOTag	(Lebreton et al, 2008; Malakhov et al, 2004)
pET28 C-6xHisTag	pET-28b(+) from Novagen
pUC57-AnCutA	this study; recoded AnCutA ORF synthesized by Genscript
pUC57-TtCutA	this study; recoded TtCutA ORF synthesized by Genscript
pCC1	[pET28 C-6xHisTag] AnCutA ^{WT} , full-length
pCC2	[pET28 C-6xHisTag] AnCutA ^{D203A/D205A} (DADA), full-length
pCC3	[pET28M N-6xHis-SUMOTag] AnCutA _{tr} ^{WT}
pCC4	[pET28 C-6xHisTag] TtCutA ^{WT} , full-length
pCC5	[pET28M N-6xHis-SUMOTag] TtCutA _{tr1} ^{WT}
pCC6	[pET28 C-6xHisTag] TtCutA _{tr2} ^{WT}
pCC7	[pET28 C-6xHisTag] TtCutA _{tr3} ^{WT}
pCC8	[pET28 C-6xHisTag] TtCutA _{tr3} ^{R557A} (PAP1)
pCC9	[pET28 C-6xHisTag] TtCutA _{tr3} ^{R557Q} PAP2 (PAP2)
pCC10	[pET28 C-6xHisTag] TtCutA _{tr3} ^{E551D/N554E/R557A} (PAP3)
pCC11	[pET28 C-6xHisTag] TtCutA _{tr3} ^{E551T/R557A/L559S} (PAP4)
pCC12	[pET28 C-6xHisTag] TtCutA _{tr3} ^{R557N/N561R} (PAP5)
pCC13	[pET28 C-6xHisTag] TtCutA _{tr3} ^{E551D/R557N/N561R} (PAP6)
pCC14	[pET28 C-6xHisTag] TtCutA _{tr3} ^{R557H} (PUP1)
pCC15	[pET28 C-6xHisTag] TtCutA _{tr3} ^{E551D/R557H} (PUP2)
pCC16	[pET28 C-6xHisTag] TtCutA _{tr3} ^{E551D/N554E/R557H} (PUP3)
pCC17	[pET28M N-6xHis-SUMOTag] Cid1 ^{WT}
pCC18	[pET28M N-6xHis-SUMOTag] Cid1 ^{D330E/E333N/H336R} (PCP)

SUPPLEMENTAL TABLE 3. Oligonucleotides used in this study.

Oligonucleotide	Sequence (5'-3')	Purpose
<i>Oligonucleotides for cloning and site-directed mutagenesis</i>		
278F	ttttgtttaactttaagaaggagatata ccATGG TCACCGTCA CCGATGG	cloning of AnCutA WT ORF into pET28 C-6xHisTag
278R	atctcagtggtggtggtggtggtg ctcgag GCCACCTTTGCGC AGGCTTTC	cloning of AnCutA WT ORF into pET28 C-6xHisTag
280F	ttttgtttaactttaagaaggagatata ccATGG ACTCGCAAC CGGCAAC	cloning of TtCutA WT ORF into pET28 C-6xHisTag
280R	atctcagtggtggtggtggtggtg ctcgag GCCACCTTTACGA TCCGCTTC	cloning of TtCutA WT ORF into pET28 C-6xHisTag
CUCUmutF	CTCTGAC agcgct GTGGCCATTTCATCACCACGACCTGTAAA GAACTGG	site-directed mutagenesis (AnCutA D203A D205A)
CUCUmutR	TGCAAATGGCCAC agcgct GTTCAGAGCTACACAGCTTATTGCC CGATGAACC	site-directed mutagenesis (AnCutA D203A D205A)
278sumoF	gaagtctaccaggaacaaaccggt ggatcc AGCACCCATAGCG AATCTAC	cloning of AnCutA_tr into pET28M N-6xHis-SUMOTag
278sumo2	tcagtgggtggtggtggtggtg ctcgag ttaACGGGACGGCGGC GGTGCGG	cloning of AnCutA_tr into pET28M N-6xHis-SUMOTag
280sumoF	gaagtctaccaggaacaaaccggt ggatcc AGCGCCTATTCTG AAACCAAC	cloning of TtCutA_tr1 into pET28M N-6xHis-SUMOTag
280sumo2	tcagtgggtggtggtggtggtg ctcgag ttaGCGCAGCAGGATC GGACGCG	cloning of TtCutA_tr1 into pET28M N-6xHis-SUMOTag
12.280 S_F	ttttgtttaactttaagaaggagatata ccatg AGCGCCTATT CTGAAACC	cloning of TtCutA_tr2 into pET28 C-6xHisTag
12.280 sl_R	atctcagtggtggtggtggtggtg ctcgag GCGACCATCATTA AACGGTTC	cloning of TtCutA_tr2 into pET28 C-6xHisTag
280sIIIF	ttttgtttaactttaagaaggagatata ccatg AAGCCGGCCA AAAATTC	cloning of TtCutA_tr3 into pET28 C-6xHisTag
280sIIIR	atctcagtggtggtggtggtggtg ctcgag CAGCGTGGTACGA GAAGCCG	cloning of TtCutA_tr3 into pET28 C-6xHisTag
280PAP1F	GTTTAATGATGGT GCCAAcctagg CAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA WT→PAP1 (R557A)
280PAP1R	CGGCGGTATTG cctagg TTGGCACCATCATTAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PAP1 (R557A)
280PAP2F	GTTTAATGATGGT CAGAAcctagg CAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA WT→PAP2 (R557Q)
280PAP2R	CGGCGGTATTG cctagg TTCTGACCATCATTAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PAP2 (R557Q)
280PAP3F	GCGTTGAAGACCCGTTT GAGG ATGGTGCCAAcct agg CAATAC CGCCGAC	Site-directed mutagenesis TtCutA PAP1→PAP3 (E551D N554E R557A)
280PAP3R	ATTGcc gagg TTGGCACCATC CTCAA ACGG GTCTT CAACGCAC AGAGCATTG	site-directed mutagenesis TtCutA PAP1→PAP3 (E551D N554E R557A)
280PAP4F	TTGAAACACCGTTTAATGATGGTGCCAAc agcgg CAATACCGC CGACGAAAC	site-directed mutagenesis TtCutA PAP1→PAP4 (E551T R557A L559S)
280PAP4R	TTGcc gctg TTGGCACCATCATTAACGGT TGT TTCAACGCACA GAGCATTGTTC	site-directed mutagenesis TtCutA PAP1→PAP4 (E551T R557A L559S)
280PAP5F	TTTAATGATGGT AACAAcctagg CAGGACCGCCGACGAAACGA GCTTCC	site-directed mutagenesis TtCutA WT→PAP5 (R557N N561R)

280PAP5R	TTTCGTCGGCGGTCTGcctaggTTGTTACCATCATTAACGG TTCTTC	site-directed mutagenesis TtCutA WT→PAP5 (R557N N561R)
280PAP6F	TTTAATGATGGTAACAAcct tggCAGGACCGCCGACGAAACGA GCTTCC	site-directed mutagenesis TtCutA PUP2→PAP6 (E551D R557N N561R)
280PAP6R	TTTCGTCGGCGGTCTGccaggTTGTTACCATCATTAACGG ATCTTC	site-directed mutagenesis TtCutA PUP2→PAP6 (E551D R557N N561R)
280PUP1F	GTTTAATGATGGTCACAAcctaggCAATACCGCCGACGAAACG	site-directed mutagenesis TtCutA WT→PUP1 (R557H)
280PUP1R	GCGGTATTGcctaggTTGTGACCATCATTAACGGTTCTTC	site-directed mutagenesis TtCutA WT→PUP1 (R557H)
280PUP2F	GCGTTGAAGATCCGTTTAATGATGGTCACAAcctaggCAATAC CGCCGACGAAAC	site-directed mutagenesis TtCutA WT→PUP2 (E551D R557H)
280PUP2R	GCGGTATTGcctaggTTGTGACCATCATTAACGGATCTTCAA CGCACAGAGCATTG	site-directed mutagenesis TtCutA WT→PUP2 (E551D R557H)
280PUP3F	AGATCCGTTTGAAGATGGTCACAAcct cggCAATACCGCCGAC GAAACG	site-directed mutagenesis TtCutA PUP2→PUP3 (E551D N554E R557H)
280PUP3R	CGGTATTGccgaggTTGTGACCATCTTCAAACGGATCTTCAAC GCACA	site-directed mutagenesis TtCutA PUP2→PUP3 (E551D N554E R557H)
Cid1SumoF	tgattgaagtctaccaggaacaaaccggtggatccATGAACAT TTCTTCTGCACAA	cloning of Cid1 WT ORF into pET28M N-6xHis- SUMOtag
Cid1SumoR	cggatctcagtgggtgggtgggtgggtgctcgagtcacCTCAGAA TTGTCACCATCGG	cloning of Cid1 WT ORF into pET28M N-6xHis- SUMOtag
Cid1PCPF	GCGATTGAAgagcccTTCAATATTTACGTAATGTGGGTAGGA CAGTTAG	site-directed mutagenesis Cid1 WT→PCP (D330E E333N H336R)
Cid1PCPR	TACCCACATTACGTGAAATATTGAagggtcTTCAATCGCAAG AATATAC	site-directed mutagenesis Cid1 WT→PCP (D330E E333N H336R)
Sequencing primers		
T7promoter	TAATACGACTCACTATAGGG	Sequencing of inserts in pET28 C-6xHisTag
SumoF	TCATACTGTCAAAGACAGGG	Sequencing of inserts in pET28M N-6xHis-SUMOtag
T7terminator	TGCTAGTTATTGCTCAGCGG	Sequencing of inserts in pET28 C-6xHisTag and pET28M N-6xHis-SUMOtag
278 seq 1	GAACAATACCATGGCCCTGG	Sequencing of AnCutA
278 seq 2	AGTAACCAGTATTCCCGTGG	Sequencing of AnCutA
12_107_s1	ACATTTGCATCACCACGACC	Sequencing of AnCutA
12_107_s2	AACACCATTGCAACCTGGG	Sequencing of AnCutA
12_107_s3	CAATTCAACCTATCTGCCGG	Sequencing of AnCutA
12_107_s4	GCTGATTGTGACGGTTTCAG	Sequencing of AnCutA
280seq5	CAGCATTCTTCCAGTTTGCC	Sequencing of TtCutA
12_109_s1	GCAGCGCCTATTCTGAAACC	Sequencing of TtCutA
12_109_s2	TATCAAATACTGGACCCGCC	Sequencing of TtCutA
12_109_s3	CCGCGGTAACCGTCAGTTTC	Sequencing of TtCutA
12_109_s4	GAGCTTCATGCCGGATGAAC	Sequencing of TtCutA

<i>RNA oligonucleotides</i>		
5'-FAM-ss17-A ₃₅	FAM-r (CCCCACCACCAUCACUUAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAA)	RNA substrate used in biochemical assays
5'-FAM-ss17-A ₃₁ U ₄	FAM-r (CCCCACCACCAUCACUUAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAUUUU)	RNA substrate used in biochemical assays
5'-FAM-ss17-A ₃₁ C ₄	FAM-r (CCCCACCACCAUCACUUAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAACCCC)	RNA substrate used in biochemical assays
5'-FAM-ss22-A ₄	FAM-r (CGACUGGAGCACGAGGACACUGAAAA)	RNA substrate used in biochemical assays
5'-FAM-ss22-U ₄	FAM-r (CGACUGGAGCACGAGGACACUGUUUU)	RNA substrate used in biochemical assays
5'-FAM-ss22-C ₄	FAM-r (CGACUGGAGCACGAGGACACUGCCCC)	RNA substrate used in biochemical assays
<i>Oligonucleotides used in the 3'-RACE-seq procedure</i>		
RA3	TGGAATTCTCGGGTGCCAAGG	Adapter oligonucleotide
RTP	<u>GCCTTGGCACCCGAGAATTCCA</u>	Reverse transcription primer
RA5 _{ss22}	G TTCAGAGTTCTACAGTCCGACGATC CGACTGGAGCACGAGGA CACTG	First-round PCR primer (forward)
RP1	AATGATACGGCGACCACCGAGATCTACACGTT CAGAGTTCTAC AGTCCGA	Second-round PCR primer (forward)
RPI_X	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTT <u>CCTTGGCACCCGAGAATTCCA</u>	PCR primer with RNA-seq index (reverse)

SUPPLEMENTAL TABLE 4. Summary of protein purification strategies.

CutA variant	Epitope	Purification scale	SUMO cleavage + desalting + second nickel column	Buffering agent	[NaCl]	Additional purification	Figure
An FL WT An FL DADA Tt FL WT	C-6xHisTag	Small	-	Na-phosphate pH=8.0	300 mM	-	Figs. 2-4
An_tr WT Tt_tr1 WT	N-6xHis-SUMOTag	Large	Yes	Na-phosphate pH=4.5	300 mM	Gel filtration (Superdex S200) 20 mM Na-phosphate; 150 mM NaCl	Fig. 5
Tt_tr3 WT	C-6xHisTag	Large	-	Na-phosphate pH=4.5	300 mM	Gel filtration (Superdex S200) 20 mM Na-phosphate; 150 mM NaCl	Fig. 5
An_tr WT Tt_tr1 WT	N-6xHis-SUMOTag	Small	No	Na-phosphate pH=4.5	300 mM	-	Fig. 6
An FL WT Tt FL WT Tt_tr3 WT	C-6xHisTag	Small	-	Na-phosphate pH=4.5	300 mM	-	Fig. 6
Tt_tr3 (WT, PAP1-PAP6, PUP1-PUP3)	C-6xHisTag	Large	-	Na-phosphate pH=4.5	300 mM	Gel filtration (S200) 20 mM Na-phosphate; 150 mM NaCl	Figs. 8-13
Cid1 (WT, PCP)	N-6xHis-SUMOTag	Large	Yes	Tris-HCl pH=8.0	150 mM	Gel filtration (S200) 20 mM Tris-HCl 150 mM NaCl	Fig. 14

.....
Miejscowość i data

Kamil Kobyłecki

Pracownia Biologii RNA i Genomiki Funkcjonalnej

Instytut Biochemii i Biofizyki

Polskiej Akademii Nauk,

Pawińskiego 5a,

02-106 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

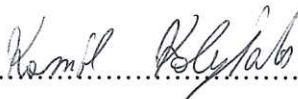
Kobyłecki K, Kuchta K, Dziembowski A, Ginalski K, Tomecki R.

"Biochemical and structural bioinformatics studies of fungal CutA nucleotidyltransferases explain their unusual specificity toward CTP and increased tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails",

opublikowanej w roku 2017 w RNA 23(12): 1902-1926

mój udział polegał na:

- przygotowaniu konstruktów plazmidowych;
- przeprowadzeniu ekspresji i czyszczenia białek;
- przeprowadzeniu większości eksperymentów biochemicznych opisanych w pracy;
- przygotowaniu bibliotek do eksperymentu 3'-RACE-seq,
- przeprowadzeniu wstępnej analizy danych z sekwencjonowania pod nadzorem R. Tomeckiego.

.....


Podpis

12 listopada 2019

.....
Miejscowość i data

Krzysztof Kuchta
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

Kobylecki K, Kuchta K, Dziembowski A, Ginalski K, Tomecki R.
Biochemical and structural bioinformatics studies of fungal CutA
nucleotidyltransferases explain their unusual specificity toward CTP and increased
tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails.
2017 RNA 23: 1902-1926

mój udział polegał na:

- stworzeniu modelu 3D TtCutA,
- wykonaniu analizy sekwencyjno-strukturalnej,
- wykonaniu analizy danych z sekwencjonowania 3'-RACE,
- współplanowaniu eksperymentów biochemicznych opisanych w pracy, polegających na zmianie specyficzności substratowej enzymów,
- przygotowaniu rysunku 1, rysunku 6, rysunku 7, rysunku 8, rysunku 9, rysunku 10, rysunku 14,
- opisanu w manuskrypcie przeprowadzonych przeze mnie badań.

.....

Podpis

12 listopad 2019

.....
Miejscowość i data

Andrzej Dziembowski
Pracownia Biologii RNA i Genomiki Funkcjonalnej
Instytut Biochemii i Biofizyki
Polskiej Akademii Nauk,
Pawińskiego 5a,
02-106 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

Kobylecki K, Kuchta K, Dziembowski A, Ginalski K, Tomecki R.

*Biochemical and structural bioinformatics studies of fungal CutA
nucleotidyltransferases explain their unusual specificity toward CTP and increased
tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails.*

2017 RNA 23: 1902-1926

mój udział polegał na:

- współkierowaniu projektem naukowym obejmującym badania opisane w tej pracy,
- współplanowaniu eksperymentów biochemicznych opisanych w pracy,
- pisaniu manuskryptu.



.....
Podpis

Warszawa, 12 listopad 2019

.....
Miejscowość i data

Krzysztof Ginalski
Pasteura 4/6 m. 56
02-093 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

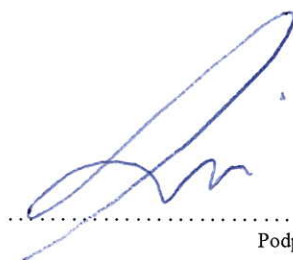
Kobylecki K, Kuchta K, Dziembowski A, Ginalski K, Tomecki R.

*Biochemical and structural bioinformatics studies of fungal CutA
nucleotidyltransferases explain their unusual specificity toward CTP and increased
tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails.*

2017 RNA 23: 1902-1926

mój udział polegał na:

- współkierowaniu projektem naukowym obejmującym badania opisane w tej pracy,
- współplanowaniu eksperymentów biochemicznych opisanych w pracy,
- zaplanowaniu badań bioinformatycznych,
- pisaniu manuskryptu.



.....
Podpis

Rafał Tomecki
Pracownia Biologii RNA i Genomiki Funkcjonalnej
Instytut Biochemii i Biofizyki
Polskiej Akademii Nauk,
Pawińskiego 5a,
02-106 Warszawa

OŚWIADCZENIE

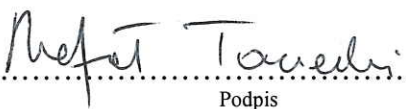
Oświadczam, że w pracy:

Kobyłecki K, Kuchta K, Dziembowski A, Ginalski K, Tomecki R.

*"Biochemical and structural bioinformatics studies of fungal CutA
nucleotidyltransferases explain their unusual specificity toward CTP and increased
tendency for cytidine incorporation at the 3'-terminal positions of synthesized tails",
opublikowanej w roku 2017 w RNA 23(12): 1902-1926*

mój udział polegał na:

- zaplanowaniu wszystkich eksperymentów biochemicznych;
- współkierowaniu projektem naukowym obejmującym badania opisane w tej pracy wraz z K. Ginalskim, w szczególności nadzorowaniu eksperymentów wykonywanych przez K. Kobyłeckiego;
- przeprowadzeniu eksperymentów biochemicznych na etapie poprawy manuskryptu po recenzjach;
- interpretacji wyników doświadczeń;
- wykonaniu rycin 2-5 i 11-13 w manuskrypcie oraz ryciny S1 i tabel S2-S4 w suplemencie;
- przygotowaniu ostatecznej wersji pozostałych rycin w manuskrypcie;
- napisaniu większości manuskryptu;
- przygotowaniu większości odpowiedzi na uwagi recenzentów.

.....

Podpis

FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases

Krzysztof Kuchta^{1,2,†}, Anna Muszewska^{1,3,†}, Lukasz Knizewski¹, Kamil Steczkiewicz¹,
Lucjan S. Wyrwicz⁴, Krzysztof Pawlowski⁵, Leszek Rychlewski⁶ and Krzysztof Ginalski^{1,*}

¹Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, Zwirki i Wigury 93, 02–089 Warsaw, Poland, ²College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Banacha 2C, 02–097 Warsaw, Poland, ³Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02–106 Warsaw, Poland, ⁴Laboratory of Bioinformatics and Biostatistics, M. Skłodowska-Curie Memorial Cancer Center and Institute of Oncology, WK Roentgena 5, 02–781 Warsaw, Poland, ⁵Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Nowoursynowska 166, 02–787 Warsaw, Poland and ⁶BioInfoBank Institute, Limanowskiego 24A, 60–744 Poznan, Poland

Received December 16, 2015; Accepted March 22, 2016

ABSTRACT

FAM46 proteins, encoded in all known animal genomes, belong to the nucleotidyltransferase (NTase) fold superfamily. All four human FAM46 paralogs (FAM46A, FAM46B, FAM46C, FAM46D) are thought to be involved in several diseases, with FAM46C reported as a causal driver of multiple myeloma; however, their exact functions remain unknown. By using a combination of various bioinformatics analyses (e.g. domain architecture, cellular localization) and exhaustive literature and database searches (e.g. expression profiles, protein interactors), we classified FAM46 proteins as active non-canonical poly(A) polymerases, which modify cytosolic and/or nuclear RNA 3' ends. These proteins may thus regulate gene expression and probably play a critical role during cell differentiation. A detailed analysis of sequence and structure diversity of known NTases possessing PAP/OAS1 SBD domain, combined with state-of-the-art comparative modelling, allowed us to identify potential active site residues responsible for catalysis and substrate binding. We also explored the role of single point mutations found in human cancers and propose that FAM46 genes may be involved in the development of other major malignancies including lung, colorectal, hepatocellular, head and neck, urothelial, endometrial and renal papillary carcinomas and melanoma. Identification of these novel enzymes taking part in

RNA metabolism in eukaryotes may guide their further functional studies.

INTRODUCTION

Proteins adopting the nucleotidyltransferase (NTase) fold play crucial roles in various biological processes, such as RNA stabilization and degradation (e.g. RNA polyadenylation), RNA editing, DNA repair, intracellular signal transduction, somatic recombination in B cells, regulation of protein activity, antibiotic resistance and chromatin remodelling (1). Almost all known members of this large and highly diverse superfamily transfer nucleoside monophosphate (NMP) from nucleoside triphosphate (NTP) to an acceptor hydroxyl group belonging to protein, nucleic acid or small molecule. They are characterized by the presence of a common α/β -fold structure composed of a three-stranded, mixed β -sheet flanked by four α -helices. This common core corresponding to the minimal NTase fold is usually decorated by various additional structural elements and additional domains, depending on the family. Sequence analysis of distinct members of this superfamily revealed the following common sequence patterns in NTase fold domain: hG[GS], [DE]h[DE]h and h[DE]h (where h indicates a hydrophobic amino acid) that include conserved active site residues. Three conserved aspartates/glutamates are involved in the coordination of divalent ions and activation of the acceptor hydroxyl group of the substrate. Two of them (from the [DE]h[DE]h motif) are located on the second core β -strand, while the third carboxylate (from the h[DE]h motif) is placed on the structurally adjacent third β -strand. The hG[GS] pattern is placed at the beginning of a short, sec-

*To whom correspondence should be addressed. Tel: +48 22 554 0800; Fax: +48 22 554 0801; Email: kgin@cent.uw.edu.pl

†These authors contributed equally to the paper as first authors.

ond core α -helix and has a crucial role in harbouring the substrate within the active site (1).

Human members of the NTase fold superfamily are encoded by 43 genes (1). Until now, only one group of potentially active human NTases, belonging to the so-called family-with-sequence-similarity-46 (FAM46), has not been characterized for its exact biological function. Little is known about FAM46 proteins apart from their involvement in various diseases. FAM46A, a gene preferentially expressed in the retina, was reported as a positional candidate for human retinal diseases since it maps within the RP25 locus to chromosome 6p12.1-q14.1 interval where several retinal dystrophy loci are located (2). It was also suggested that a variable number of tandem repeat polymorphisms in FAM46A may be associated with non-small cell lung cancer (3). Another FAM46 paralog, FAM46B, was identified to have lower expression in metastatic melanoma cells (United States Patent US 7615349 B2). FAM46B and FAM46C have been recently described as potential markers for refractory lupus nephritis (4) and multiple myeloma (5–10), respectively. Finally, it was reported that FAM46D is overexpressed in lung and glioblastoma tumors (11), as well as together with FAM46C, in the brain of autistic-like behaving transgenic mice (8).

Functional proteomics studies showed that FAM46A might have many potential protein interacting partners (12). One of them, ZFYVE9, detected in the yeast two-hybrid system, is involved in the recruitment of unphosphorylated SMAD2/SMAD3 to the transforming growth factor- β receptor (13). Another member of FAM46 family, FAM46C, is recognized as a type I interferon stimulated gene, which enhances the replication of certain viruses (14). In addition, FAM46C can be an anti-viral factor in acute infected long-tailed pygmy rice rats by Andes virus (15). It was also suggested that FAM46C is functionally related in some way to the regulation of translation (6).

To date, several studies have indicated that proteins belonging to FAM46 family might play an essential role in the cell; however, their exact functions remain unknown. In our previous work, we performed a comprehensive classification of the NTase fold proteins and assigned FAM46 proteins to this superfamily as potentially active members (1). This classification allowed us only to speculate that like other active NTases, FAM46 members may catalyze template-independent incorporation of NMP from NTP either to nucleic acid, protein or small molecule. In this study we present an in-depth bioinformatics analysis of the FAM46 family combined with an exhaustive literature and database searches and propose that the FAM46 proteins function as non-canonical poly(A) polymerases. Detailed insight into the sequence and structure diversity of NTases and their additional N- and C-terminal domains allowed us to generate a reliable 3D model for one of the family members (FAM46C) and to confidently identify the potential active site residues responsible not only for catalysis but also for substrate binding. In addition, the obtained structural model for human FAM46C sheds some new light on the molecular role of mutations found in cancer patients in the FAM46 genes. Finally, the broad sampling of sequenced genomes made it possible to track the evolutionary history of FAM46 proteins back to the origin and hypothesize that

the FAM46 family members are present not only in animals but also in all four sequenced Dictyosteliidae and two Entamoeba (Amoebozoa) genomes.

MATERIALS AND METHODS

Sequence searches

Four human FAM46 paralogs (FAM46A, FAM46B, FAM46C, FAM46D) were used as queries in PSI-BLAST (16) searches performed against the NCBI non-redundant (NR) protein sequence database with E-value threshold of 0.001 until profile convergence. Collected sequences were split into organism-specific sets and clustered with CD-HIT (17) in order to obtain unique sequences. All FAM46 family members were aligned using Mafft (18) with some manual adjustments. The alignment used for phylogeny reconstruction was additionally trimmed by TrimAl (19) to eliminate poorly aligned and thus uninformative regions.

Additionally, proteins containing both NTase and C-terminal four-helical up-and-down bundle fold domains were collected with PSI-BLAST searches performed against the NCBI NR database with E-value threshold of 0.001. Sequences (PDB SEQRES) of the following structures possessing this domain context: pdb12pbe, pdb13c18, pdb13jyy, pdb13jz0, pdb14ebs, pdb11v4a, pdb13k7d and pdb11kan, were used as queries.

Analysis of gene and protein features

The architecture of the human FAM46 genes was analysed using the UCSC genome browser (20). Protein localization was predicted with BaCelLo (21), CELLO (22), WoLF PSORT (23), Euk-mPLoc 2.0 (24) and MultiLoc (25). NetNES (26) was used to detect the nuclear export signal (NES). Protein phosphorylation motifs were detected with Eukaryotic Linear Motif (ELM) (27). Gene expression patterns were analysed using the BioGPS database (28). Genes with average z-scores higher than 5 (at least in one probe set) in 'Barcode on normal tissues' dataset were considered as expressed in specific tissue/cell. 'Barcode on normal tissues' dataset provides a survey across diverse normal human tissues from the U133plus2 Affymetrix microarray (28). The z-scores in this dataset are generated with the barcode function from the R package 'frma', which bases on barcode algorithm (29). The domain architecture was analysed using Meta-BASIC (30) and SMART (31).

Structural analysis of known NTases

Initially, known NTase fold families and structures were identified from literature (including our previous classification (1)) and various databases of catalogued protein families (PFAM (32), COG (33) and KOG (34)) and structures (PDB (35) and SCOP (36)). This initial set was then used for a comprehensive, transitive searches for all NTase fold superfamily members using our distant homology detection method Meta-BASIC (30) and Gene Relational DataBase (GRDB) system, as described in our previous work (1). Briefly, Meta-BASIC is a highly sensitive meta-profile alignment method capable of finding very distant similarity between proteins through a comparison of sequence profiles

enriched by predicted secondary structures (meta-profiles). The GRDB system includes precalculated Meta-BASIC connections between 16 230 PFAM, 4825 KOG and 4873 COG families and 38 498 representative proteins of known structure (PDB90). Each family and each structure is represented by its sequence (PDB90) or consensus sequence (PFAM, COG and KOG), sequence profile (generated with PSI-BLAST using the NCBI NR70 derivative) and secondary structure profile (predicted with PSIPRED (37)).

The structural diversity was analysed for all collected NTase superfamily structures clustered at 90% sequence identity. Structures were divided into groups based on their structural similarity using DALI (38). Structure-based alignments were generated for all considered domains (including both the conserved NTase fold and additional N- and C-terminal domains) after manually curated superposition of their structures. Secondary structures were assigned with DSSP (39).

3D model building

Potential templates were identified with Meta-BASIC and the consensus of fold recognition 3D-Jury approach (40) using human FAM46 proteins as queries. The sequence-to-structure alignment between FAM46 family members and all representative structures possessing both NTase and PAP/OAS1 SBD domains was built using a consensus alignment approach and 3D assessment (41) based on Meta-BASIC and 3D-Jury results, PSIPRED secondary structure predictions and conservation of critical active site residues and hydrophobic patterns. The 3D model of human FAM46C protein was built with MODELLER (42) using *Trypanosoma brucei* TUTase 4 (pdb12ikf) (43) as a template. Finally, the side chain rotamers were optimized using SCWRL3 (44). The overall quality of the modelled structure was checked with ProSA (45). Structure visualization was carried out with Pymol (<http://www.pymol.org>).

Analysis of protein interactors

Human members of the NTase fold superfamily were identified from the UniProt database (46) using our transitive Meta-BASIC search strategy as described above, starting with all collected NTase fold families and structures. Proteins interacting with human NTase superfamily members were identified using the BioGRID database (version 3.4.133) (47). GO annotations (molecular function and biological process) (48) for detected interactors were taken from the UniProt database. FAM46 interacting partners were also identified with manual literature searches.

Analysis of single point mutations in cancers

Missense mutations, found in cancer patients, in FAM46 genes were collected from publications and the following databases: cBioPortal (49), ICGC (50) and IntOGen (51). The sequence conservation in FAM46 family was measured based on Jensen–Shannon divergence (JSD) (52) using created FAM46 multiple sequence alignment. The JSD quantifies the similarity between probability distributions with scores ranging from 0 to 1 (53). A background amino acid

distribution, estimated from a large sequence set, is used to approximate the distribution of amino acid sites subject to no evolutionary pressure. Positions in an alignment that are found to have amino acid distributions very different from the background distribution are proposed to be functionally important or constrained by evolution. The JSD score was computed using the `score_conservation.py` program (52) with default parameters, e.g. using BLOSUM62 for the background distribution. Positions in FAM46 multiple sequence alignment with more than 30% gaps were omitted from JSD computations.

Phylogeny

In order to visualize the relationships between FAM46 family members, a phylogenetic analysis was performed with PhyML3.0 (54) using the LG and JTT models, with an estimated gamma parameter and proportion of invariable sites. An approximate branch support was calculated using the aLTR (55) option implemented in PhyML. Branches with supports lower than 0.5 were collapsed. The trees were drawn using iTol (56).

RESULTS AND DISCUSSION

FAM46 family

Firstly, we identified proteins belonging to FAM46 family with an exhaustive PSI-BLAST (16) searches performed against the NCBI NR protein sequence database using all four human FAM46 paralogs (FAM46A, FAM46B, FAM46C and FAM46D) as queries. These searches quickly converged at the third iteration; however, most family members can be easily detected even with a simple BLAST search. This is a feature of compact and very conserved protein families and graphical clustering of all these sequences corroborated this observation. As many of them turned out to be variants or mutants of the same protein (e.g. there are four FAM46 genes in the human genome and 14 proteins in the NR database) we selected 868 protein sequences unique for each organism using sequence clustering at different thresholds followed by manual assessment. It should be noted that some of the detected proteins contain long deletions within conserved regions, what might be due to erroneous gene/exon prediction.

Taxonomic distribution

FAM46 proteins are present in the proteomes of all animals. Supplementary Figure S1 summarizes the taxonomic distribution of all selected 868 FAM46 proteins unique for each organism. Four FAM46 paralogs can be identified in almost all sequenced Vertebrata (with high-quality genomes), but not in Tunicata (*Ciona intestinalis* and *Oikopleura dioica*), Hemichordata (*Saccoglossus kowaleskii*), Echinodermata (*Strongylocentrotus purpuratus*) or Cephalochordata (*Branchiostoma floridae*), which encode only a single FAM46 protein. Specifically, amphibian, bird, reptile and mammal genomes harbour four distinct FAM46 genes. On the other hand, fish proteomes contain six to seven FAM46 paralogs due to lineage specific duplications followed by fast differentiation of the retained paralogs (different in

each of the four analysed fishes). This evolutionary scenario has been already described in teleost fishes (57). An asymmetric acceleration of evolutionary rate in one of the paralogs after the duplication event, manifested by the high protein sequence divergence and usually leading to alignment problems in less conserved regions, was also observed in FAM46 paralogs. FAM46 family members are encoded in all sequenced animal phyla ranging from Arthropoda (*Daphnia*, *Drosophila*), Mollusca (*Crassostrea gigas*), Nematoda (*Caenorhabditis elegans*, *Brugia malayi*, *Loa loa*, *Trichinella spiralis*), Platyhelminthes (*Schistosoma mansoni*, *Clonorchis sinensis*), Cnidaria (*Nematostella vectensis*), Placozoa (*Trichoplax adherens*) and Porifera (*Amphimedon queenslandica*). FAM46 genes duplicated and diverged strongly in some Nematoda lineages leading to a variable number of paralogs in the analysed genomes. Moreover, FAM46 proteins are detectable in close metazoan relatives: Choanoflagellida (*Salpingoeca* sp., *Monosiga brevicollis*) and Ichthyosporea (*Sphaeroforma arctica*). Surprisingly, proteins belonging to this family can be also found in Amoebozoa (four *Entamoeba* species, *Polysphondylium pallidum*, *Acytostelium subglobosum* and three *Dictyostelium* species and *Acanthamoeba castellanii*) and one Diplomonadida (*Guillardia theta*). Choanoflagellida and Ichthyosporea, together with Metazoa, are a sister group of Fungi and Nucleariids. Noteworthy, FAM46 family members are absent in fungal and nuclearioid genomes sequenced within the Origins of Multicellularity Project by BROAD. Amoebozoa are sometimes grouped with Opisthokonta (Metazoa and Fungi) into a supertaxon Unikonta characterized by a single posterior flagellum in flagellated cells. Summarizing, the presence of FAM46 proteins in proteomes of Metazoa, Choanoflagellida and Amoebozoa suggests its origin in the ancestor of Unikonta with further divergence into four distinct conserved representatives in vertebrates.

Phylogeny inference

Phylogenetic relationships were analysed both for a set of 29 representative sequences and for a set of 868 Metazoa, Choanoflagellida, Diplomonadida and Amoebozoa FAM46 proteins. *Entamoeba*, *Giardia* and *Dictyostelids* form well-separated clades with uncertain branching order (Figure 1). They are, however, clearly separated from the Metazoa-Choanoflagellida clade. *Salpingoeca rosetta* and *M. brevicollis* form a sister clade to Metazoa. Some of invertebrate FAM46 proteins display higher variability at sequence level that can lead to long branches on the phylogenetic tree. The position of basal lineages within the Metazoa is uncertain and possible involvement of long branch attraction phenomenon should be taken into account.

The evolutionary history of FAM46 in the vertebrate genomes is a story of consecutive duplications leading to four highly similar paralogs. All vertebrate genomes analysed in this study retained all four FAM46 paralogs. Surprisingly, we detected the presence of FAM46 proteins in all sequenced Amoebozoa genomes.

The divergence time between Choanoflagellida and Metazoa is estimated ~600MYA (58,59). As FAM46 genes are present in Choanoflagellida, Metazoa and Amoebozoa genomes it is possible they were already in the ancestor of

all Unikonta, and therefore also in the ancestor of Opisthokonta. The most likely scenario involves an ancient deletion in the ancestor of Fungi. This evolutionary history claims FAM46 would be a very ancient gene, what is not reflected in its encoded amino acid sequence divergence. Provided FAM46 have an ancient origin early in the Unikonta, still the presence of two FAM46 genes in the *G. theta* genome requires clarification. Due to cohabitation, it is plausible that the FAM46 genes in *G. theta* genome appeared via horizontal gene transfer from Choanoflagellida to Diplomonadida. However, the low resolution of these deep branches renders the FAM46 phylogenetic tree (Supplementary Figure S2) uninformative for HGT inference. There is significant evidence for the transfer occurring in the opposite direction (60). The mechanism underlying algae to choanoflagellate transfer is supposed to be based on phagotrophy. We have insufficient data to hypothesize about the possibility of transfer happening from choanoflagellates to algae.

Gene structure

The organization, architecture and regulation of human FAM46 genes and their homologous counterparts in others organisms seem to contribute to their functional diversification. For instance, human FAM46 genes contain 2–3 exons of which 1–2 are coding (Supplementary Table S1) and they encode up to five different transcripts. In addition, antisense transcripts have been also detected, e.g. for human FAM46A (61). Interestingly, we found that the H3K27Ac pattern in the promoter region and along the coding region is completely different for each of human FAM46 genes, what can be related to distinct nucleosome density in these chromatin regions and different expression patterns (62). Additionally, the FAM46D gene is in a repeat dense area as denoted by RepeatMasker, which might be related to the overall chromosome X repeat density.

Domain architecture

Given the high variability of human FAM46 paralogs at the gene organization level, the encoded proteins are surprisingly similar at the sequence level, including the conservation of various motifs. FAM46 proteins seem to have a common two-domain architecture composed of an α/β region (according to secondary structure predictions) followed by an α -helical region. It should be noted that the FAM46 family is a distant outlier in the NTase fold superfamily and cannot be identified with standard sequence comparison methods such as PSI-BLAST. Using a highly sensitive tool for distant homology detection, Meta-BASIC (30), we mapped FAM46 proteins, with the above threshold scores, to several 3D structures including the terminal uridylyl transferase 4 (TUTase 4) from *T. brucei* (pdb12ikf) (43), 2'-5'-oligoadenylate synthase (OAS) from *S. scrofa* (pdb1px5) (63), CCA-adding enzyme from *A. fulgidus* (pdb14x4n) (64), cyclic AMP-GMP synthase from *V. cholerae* (pdb14u0n) (65), aminoglycoside 6-adenyltransferase from *B. subtilis* (pdb12pbe) and nuclear factors NF90 and NF45 from *M. musculus* (pdb14at7) (66). Importantly, FAM46 N-terminal α/β region has weak but

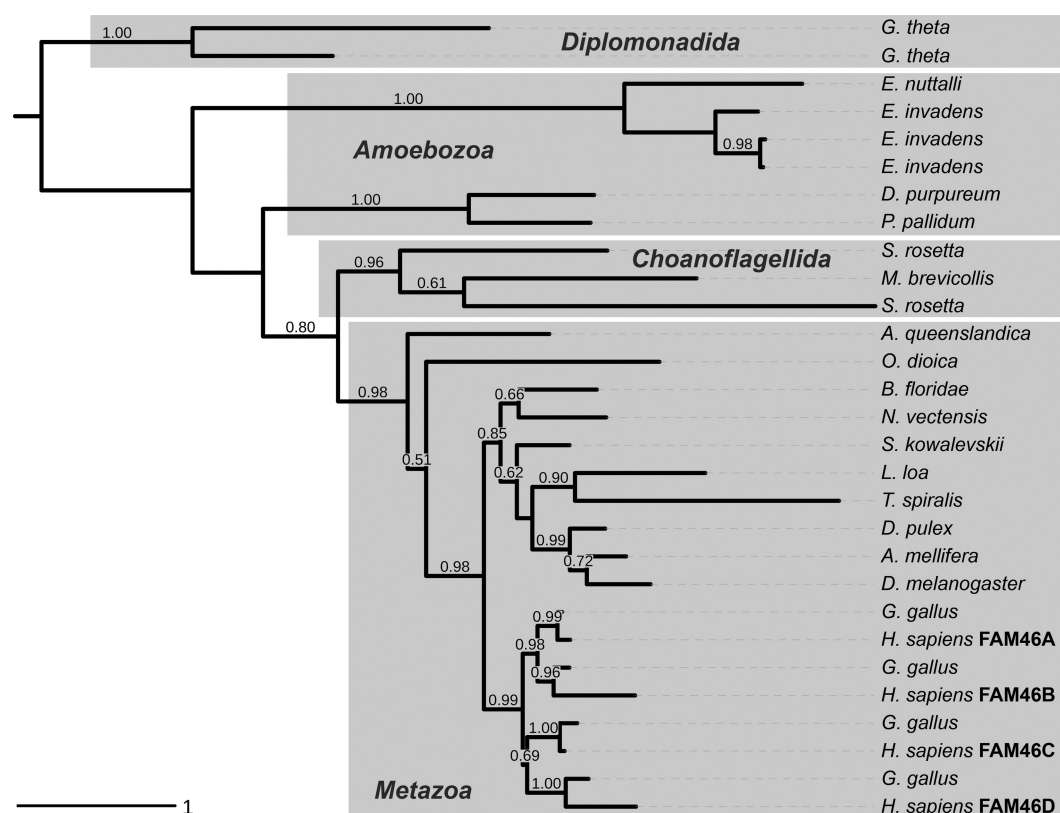


Figure 1. Phylogenetic tree of representative FAM46 protein sequences. Maximum likelihood (ML) analysis for selected 29 family members was carried out using the LG+G model. The approximate likelihood ratio test Shimodaira–Hasegawa-like (SH-like) branch supports above 0.5 are shown. Branches with support lower than 0.5 were collapsed.

evident sequence similarity to the NTase domain, which can be confirmed by several fold recognition servers. MetaBASIC suggested that the C-terminal α -helical part may be similar either to poly(A) polymerase/2'-5'-oligoadenylate synthetase 1 substrate binding domain (PAP/OAS1 SBD) or the domain of four-helical up-and-down bundle fold (4H), however, it assigned below threshold scores to these predictions. To figure out what protein fold is adopted by the FAM46 C-terminal region, we performed a comprehensive analysis of the structural diversity of all the available NTase superfamily structures, both for their conserved NTase and additional N- and C-terminal domains (Supplementary Figure S3).

While both the PAP/OAS1 SBD and 4H domains possess four core α -helices C-terminal to NTase domain, only PAP/OAS1 SBD retains the additional (the first core) α -helix located before NTase domain. According to secondary structure predictions, this helix is clearly seen in the FAM46 family members in the conserved region preceding the predicted NTase domain. In addition to a good mapping of predicted and observed core secondary structure elements, FAM46 proteins display also similar conservation of hydrophobic motifs and critical residues for NTP binding (see below) characteristic for the PAP/OAS1 SBD. In our previous studies we showed that such detailed analysis of below threshold MetaBASIC hits usually enables identification of highly diverged superfamily members which escape detection even with advanced sequence compari-

son methods. For instance, using this approach we identified restriction endonuclease-like (67) and RNase H-like (68) domains in many uncharacterized and poorly annotated protein families. Finally, we found that proteins embracing both the NTase and 4H domains are mainly encoded in bacteria and rarely found in archaeal genomes, with single representatives identified in eukaryotic species, including fungal *Myceliophthora thermophila* (gil367020986), *Tribulus terrestris* (gil367039397) and *Rhizophagus irregularis* (gil552919075), and soil-living amoeba *Dictyostelium discoideum* (gil66821023). This is consistent with the biological functions played by these proteins, as they participate in antibiotic resistance (e.g. *Staphylococcus aureus* kanamycin nucleotidyltransferase (69), *Enterococcus faecium* lincosamide antibiotic adenylyltransferase (70), *Bacillus subtilis* aminoglycoside 6-adenyltransferase) and nitrogen assimilation (e.g. *Escherichia coli* glutamine synthetase adenylyltransferase (71)). In contrast, NTases possessing the PAP/OAS1 SBD can be widely found in eukaryotes. Altogether, results of all these analyses strongly suggest that although displaying little sequence similarity, FAM46 proteins possess PAP/OAS1 SBD consisting of the five right-handed twisted α -helices (with an α 1-NTase- α 2 α 3 α 4 α 5 topology).

In addition, we found that a few FAM46 proteins possess additional domains inserted inside the NTase domain or located at N- or C-termini (Supplementary Figure S4). It should be noted, however, that the presence of some of these

additional domains may be a result of potentially incorrect gene/exon predictions.

PAP/OAS1 SBD in known NTase structures

To identify conserved PAP/OAS1 SBD residues, critical for binding NTP substrate in an NTase active site, we carried out an exhaustive sequence and structure analysis by generating the structural alignment of all the representative structures possessing both the NTase and PAP/OAS1 SBD domains (Figure 2). The PAP/OAS1 SBD specifically binds a nucleobase of the incoming NTP mainly by amino acids that provide, either directly or indirectly via water molecules, Watson–Crick hydrogen bonds. In addition, a conserved hydrophobic amino acid (e.g. V234 in poly(A) polymerase Pap1 (72) and Y212 in poly(U) polymerase Cid1 (73)), located at the beginning of the third core α -helix of PAP/OAS1 SBD, forms a flat hydrophobic surface for the incoming NTP nucleobase. Proteins containing the PAP/OAS1 SBD also possess another common residue, which is responsible for the recognition of a triphosphate moiety. Conserved lysine/arginine (e.g. K215 in Pap1) located in the second core α -helix of PAP/OAS1 SBD, together with a serine from the NTase domain hG[GS] motif, interact with NTP β - and γ -phosphate groups.

3D model of human FAM46C

Initially, we generated a sequence-to-structure alignment of FAM46 family members with all representative proteins of known structure possessing both the NTase and PAP/OAS1 SBD domains (using their structure-based alignment described above) (Figure 2). Although these structures display very little sequence similarity to the FAM46 proteins, in contrast to our previous work (1) where we focused in general on the most conserved regions of NTase fold common to all NTase superfamily members, here we were able to propose a reliable and complete sequence-to-structure alignment for all conserved regions of both domains. The alignment was guided by secondary structure predictions and conservation of (i) the NTase fold active site motifs, (ii) identified critical PAP/OAS1 SBD residues participating in substrate binding and (iii) hydrophobic patterns responsible for forming the hydrophobic core of the structure.

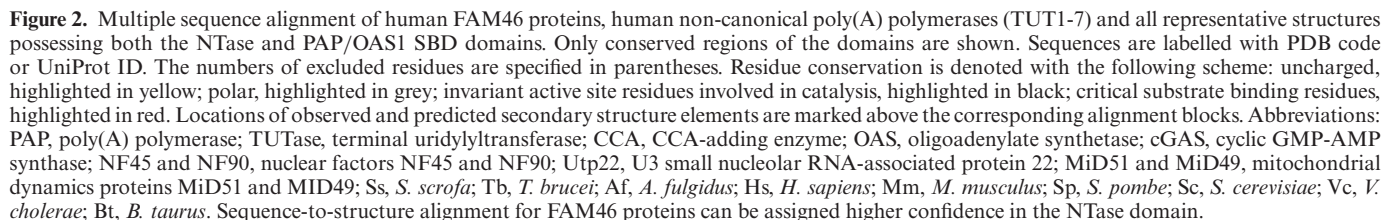
As a representative of FAM46 family for 3D modelling we selected human FAM46C, which is a potential biomarker for multiple myeloma. FAM46 proteins are very similar in sequence, for instance, four human paralogs share 56–75% amino acid identity within the common region encompassing both domains. In addition, the length of this region in these paralogs differs only by 1–2 residues.

The structure of *T. brucei* TUTase 4 (pdbl2ikf) (43), as assigned the highest Meta-BASIC score among the proteins possessing both NTase and PAP/OAS1 SBD domains, was used as a template to generate the 3D model of human FAM46C, based on the manually derived sequence-to-structure alignment. However, due to the lack of templates with similar insertion between the last core β -strand and the last core α -helix of NTase domain, we were unable to create a reliable model for 70 amino acids of FAM46C in

this region. Nevertheless, we can speculate that this insertion in FAM46C should fill the space usually occupied by residues responsible for binding incoming NTP nucleobase and RNA 5' end. Figure 3 presents a comparison of the FAM46C model and existing structures of non-canonical poly(A) polymerase Trf4p from *Saccharomyces cerevisiae*, which is a part of the Trf4p/Air2p/Mtr4p polyadenylation (TRAMP) complex (74), and the non-catalytic mitochondrial dynamic protein MiD51 from *M. musculus* (75). Importantly, in Trf4p, the region of 53 amino acids between the fourth and fifth core α -helices of the PAP/OAS1 SBD is crucial for binding the RNA 5' end and a nucleobase of the incoming NTP. The corresponding region in FAM46C is much shorter (only 11 amino acids) and probably is not able to form the interaction interface for a nucleobase. Therefore, it is likely that nucleobase binding residues are located within the 70 amino acids insertion between the last core β -strand and the last core α -helix of the FAM46C NTase domain. In addition, this conserved insertion, composed of the predicted two β -strands and two α -helices (with $\beta\alpha\beta\alpha$ order), may also participate in protein–protein interactions similar to the MiD51 receptor which binds the dynamin-related protein 1 (Drp1) via a well-conserved loop located in the NTase domain (75). However, it should be noted that FAM46C, in contrast to MiD51, seems to be an active NTase; therefore, even if the insertion is responsible for protein–protein interactions, it should also play a role in substrate recognition.

Active site

Figure 4 shows a comparison of active sites of human FAM46C (model), poly(A) polymerase Pap1 from *S. cerevisiae* (72), poly(U) polymerase Cid1 from *Schizosaccharomyces pombe* (73), CCA-adding enzyme from *A. fulgidus* (64), 2'-5'-oligoadenylate synthetase OAS1 from *S. scrofa* (76) and the cyclic GMP-AMP synthase (cGAS) from *M. musculus* (77). FAM46 proteins probably function as active NTases because they share all the key motifs in the NTase domain responsible for catalysis and substrate binding, including the [DE]h[DE]h and h[DE]h patterns with three conserved aspartate/glutamate residues (Asp90, Asp92 and Glu166) and hG[GS] motif with Gly73 and Ser74 in human FAM46C. Although we were not able to generate a complete 3D model of human FAM46C, we were able to identify additional residues responsible for NTP binding, which are located in the conserved secondary structure elements. Comparison of active sites of experimentally solved structures showed that proteins encompassing both NTase and PAP/OAS1 SBD usually bind a nucleobase or a ribose-moiety of incoming NTP by a serine or a threonine located just before or in the last core α -helix of NTase domain (Figures 2 and 4). Therefore, it is possible that FAM46C Ser248 may bind, directly or indirectly via a water molecule, 2'-OH or/and 3'-OH hydroxyl group of a ribose-base moiety as it is observed for Thr172 in poly(U) polymerase Cid1 (73) or it can participate in a nucleobase binding similarly to Thr190 in 2'-5'-oligoadenylate synthetase OAS1 (76). FAM46C shares also all the conserved residues in PAP/OAS1 SBD responsible for substrate binding. The FAM46C Leu282 probably interacts with a nucleobase of



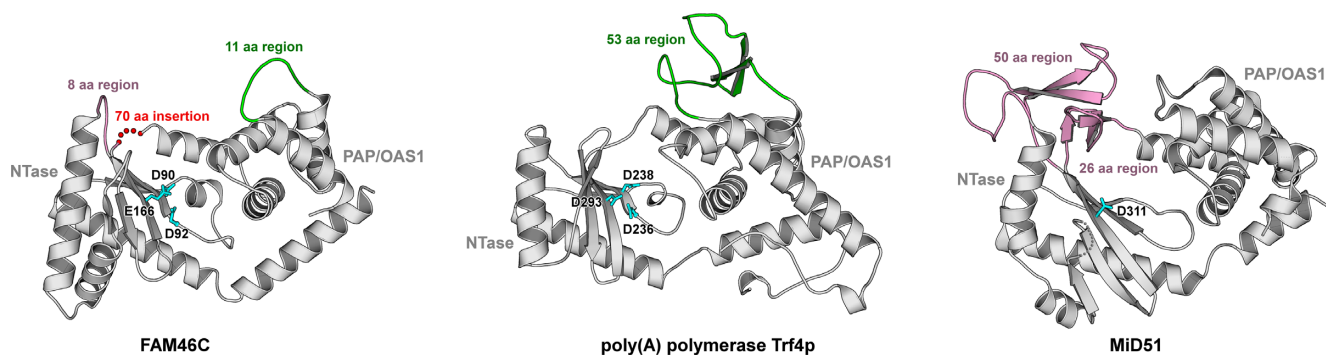


Figure 3. Comparison of 3D model of human FAM46C and available structures of non-canonical poly(A) polymerase Trf4p (pdb13nyb) and mitochondrial dynamics protein MiD51 (pdb4oaf). Regions in MiD51 responsible for protein–protein interactions and their potential counterpart in FAM46C are coloured pink. The region between the fourth and fifth core α -helices of PAP/OAS1 SBD in FAM46C and Trf4p (critical for nucleobase binding) is shown in green. The region not modelled in FAM46C (70 amino acids) is denoted by red dots. The conserved active site carboxylates are shown in blue.

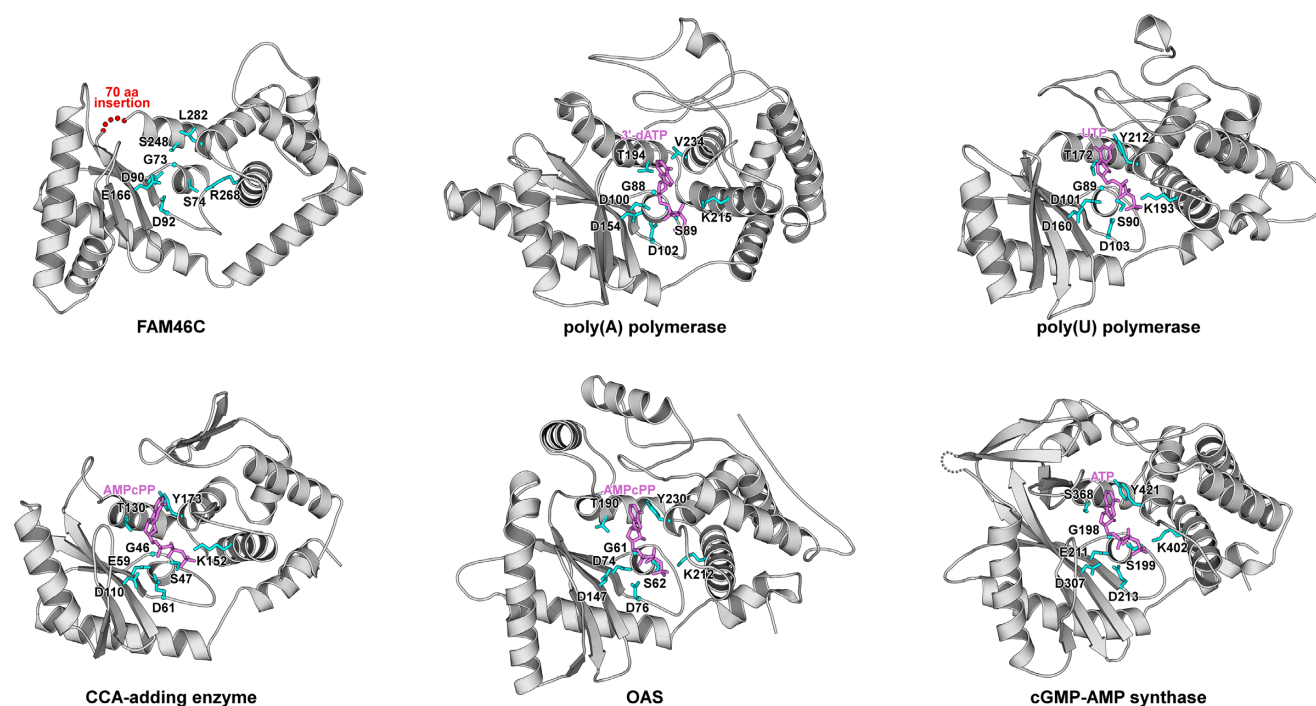


Figure 4. Comparison of the active sites of FAM46C, poly(A) polymerase (Pap1, pdb11fa0), poly(U) polymerase (Cid1, pdb4fhp), CCA-adding enzyme (pdb4x4r), OAS (OAS1, pdb4rwo) and cyclic GMP-AMP synthase (Mb21d1, pdb4k97). Only NTase and PAP/OAS1 SBD domains are shown. The region not modelled in FAM46C (70 amino acids) is denoted by red dots. Conserved amino acids critical for catalysis and substrate binding are shown in blue.

the incoming NTP like Tyr212 in poly(U) polymerase Cid1 or makes van der Waals contacts with the ribose-base moiety of NTP similar to Val234 in poly(A) polymerase Pap1 (72). Finally, NTP β - and γ -phosphates most likely interact with the conserved Arg268 in addition to Ser74 from the hG[GS] motif.

Cellular localization and tissue specificity

According to various servers predicting subcellular localization, the FAM46 proteins seem to be localized in both the cytoplasm and nucleus. In addition, three human paralogs (FAM46B, FAM46C and FAM46D) harbour potential leucine-rich NES, located at the end of the C-terminal

PAP/OAS1 SBD domain. As a consequence, it is likely that proteins belonging to the FAM46 family shuttle between the nucleus and cytoplasm.

We also analysed the gene expression data available in the BioGPS database (28) and found that each of the human FAM46 paralogs has a different tissue/cell expression pattern (Supplementary Table S1). Different expression patterns probably indicate various biological processes in which FAM46 proteins participate. According to the BioGPS database, FAM46A, FAM46B and FAM46C are potentially expressed in 81, 18 and 66 tissues/cells, respectively, while FAM46D can be found only in sperm (Supplementary Table S1).

Interacting partners

To get more hints at the biological function of FAM46, we compared all human NTase superfamily members according to GO molecular function and biological process of their protein interactors identified in the BioGRID database (47) (Supplementary Figure S5). We found that FAM46 binding partners share a common GO functions and processes mostly with interactors of human NTase fold proteins from the following four groups (described in Supplementary Table S2): interleukin enhancer-binding factors, non-canonical poly(A) polymerases (TUT), poly(A) polymerases (PAP) and zinc finger RNA-binding proteins (ZFR). Similarly to FAM46 family members, proteins belonging to these groups also possess additional PAP/OAS1 SBD domain. In addition, almost all of them retain the same biological function—DNA/RNA binding, including poly(A) RNA binding and participate in the same process—transcription.

We also analysed all 61 FAM46 protein interactors identified in the BioGRID database and the literature in more details (Supplementary Table S3). We found that each FAM46 paralog has a different set of interacting protein partners; therefore, it is likely that each paralog participates in a different biological process in the cell. Most of the interacting partners play important roles in development, including cellular proliferation and cell differentiation. We noticed that many FAM46 interactors share a common molecular functions (e.g. nucleic acids binding, including binding of the mRNA poly(A) tail) or biological processes (e.g. protein modification, transcription). Specifically, 26 FAM46 interacting partners may directly bind RNA and/or DNA. This group includes: transcription (co)factors (RHOXF2, TBX4, NR2F2, SOX5, NRF1, TRIP6), transcription activators (Znf322, Cxxc5), RNA stabilization factors (ELAVL1, BCCIP, HDLBP, Pabpc1, Pabpc4), proteins involved in transcription (POLR1A, POLR1E, POLR2J), proteins participating in mRNA translation (EIF4G3, Pabpc1, Pabpc4), proteins taking part in DNA repair (POLE2, Rad23b, WRAP53) and other proteins, which play various roles such as DNA helicase DDX11, putative RNA exonuclease (44M2.3), mitochondrial translation optimization factor 1 (MTO1), SUMO-conjugating enzyme UBC9 (UBE2I), 14-3-3 protein zeta/delta (Ywhaz) involved in signal transduction (78) and ATXN1, which may participate in RNA export (79). Similarly to human FAM46A (80), nine protein interactors (ELAVL1, BCCIP, EIF4G3, MTO1, Ywhaz, TRIP6, UBE2I, Pabpc1, Pabpc4) participate or may participate in RNA poly(A) tail binding. Another seven FAM46 interacting partners (EGLN2, DAZAP2, KEAP1, DCAF6, KLHDC2, ZNHIT6, MVP) were shown or suggested to cooperate with or regulate proteins, which are able to bind directly to nucleic acids. Specifically, the Egl nine homolog 2 (EGLN2) targets the transcriptional complex HIF- α subunit for proteasomal degradation (81). KEAP1 targets transcription factor Nrf2 for ubiquitination and degradation (82). DCAF6 enhances the transcriptional activity of nuclear receptors NR3C1 and AR (83). The physical interaction between KLHDC2 and a bZIP transcription factor (LZIP) leads to the repression of the LZIP-

dependent transcription (84). DAZAP2 regulates stress or germ granules–ribonucleoprotein complexes (85,86). MVP is a part of an evolutionary highly conserved ribonucleoprotein particles (vaults) (87,88), while ZNHIT6 may be involved in snoRNP biogenesis (89). Another functional group of FAM46 interactors contains proteins, which take part in protein modification (mostly in protein degradation). This group embraces proteases (serine protease HTRA1, caspase-like protease ESPL1), Kunitz-type protease inhibitor 1 (SPINT1), Kelch-like ECH-associated protein 1 (KEAP1) participating in Nrf2 ubiquitination and degradation (82), EGLN2 targeting HIF- α for proteasomal degradation (81), E3 ubiquitin-protein ligases (RNF14, PARK2), CUL4-associated factor 6 (DCAF6) functioning as substrate-recruiting module for CUL4-DDB1 E3 ubiquitin-protein ligase complex (90), SUMO-conjugating enzyme UBC9 (UBE2I), ubiquitin carboxyl-terminal hydrolase 4 (USP4), a regulatory subunit 6A of the 26S proteasome (Psmc3) and BAG6, which is crucial in ubiquitin-mediated protein degradation of defective or misallocated polypeptides (91). Among all FAM46 interactors, we were able to identify two kinases: Polo-like kinase 4 (PLK4) and non-receptor tyrosine-protein kinase SYK (SYK). PLK regulates cell cycle progression, mitosis and cytokinesis (92), while SYK mediates signal transduction and differentiation, particularly in B-cell development (93,94). Another interacting partner, the FYVE domain-containing protein 9 (ZFYVE9) identified as a SMAD2/3-binding protein, may also regulate the proliferation of hepatic cells during zebrafish embryogenesis (95). The last identified functional group contains proteins responsible for an intra and extracellular cell transport, including DYNC1H1—a heavy chain of cytoplasmic dynein 1 (96,97), RIN3—a small GT-Pase, which participates in intracellular membrane trafficking (98) and AP2B1, which plays a pivotal role in many vesicle trafficking pathways within the cell (99).

Mutations in cancers

Recent studies have identified numerous somatic mutations in various cancer patients leading to single point mutations in the human FAM46 proteins (Supplementary Table S4). For instance, the human FAM46C gene was reported as a causal driver of multiple myeloma (6). In addition, a single FAM46C mutation (Y247N) was identified in *hem6* mice with hypochromic anemia, which affects terminal spermiogenesis and terminal stages of erythroid differentiation (100). This study showed that male *hem6* mice produce sperm with defects detectable by phase contrast microscopy and fluorescence microscopy. To analyze the role of these mutations, we mapped them onto a 3D model of FAM46C (Figure 5) and found that they can be divided into two groups. The first group includes mutations that are located in a highly conserved area of the active site and its close vicinity, and probably may decrease/increase FAM46 catalytic activity and/or affect substrate binding (e.g. change the preference for the type of incorporated NTP). This group embraces the mouse *hem6* mutation and the majority of mutations found in multiple myeloma patients as well as several mutations from other cancers. The average JSD (52) score for all mutated amino acid positions in multi-

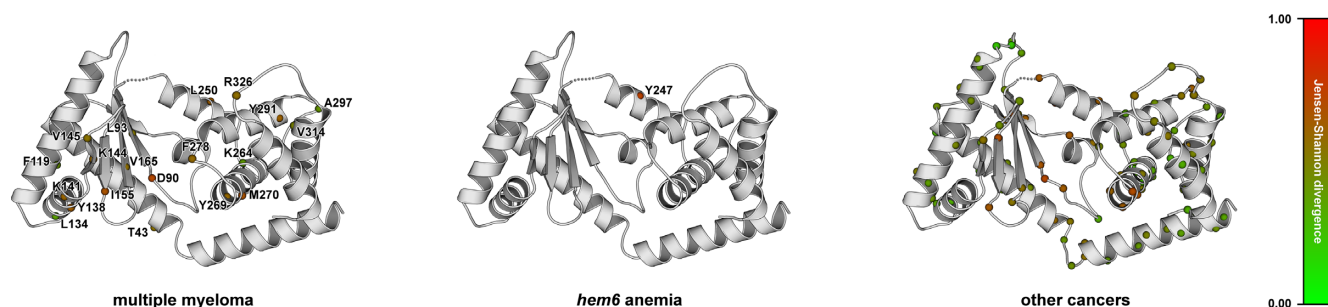


Figure 5. Missense mutations in FAM46 family members found in cancer patients and *hem6* mouse. The positions of the corresponding single point mutations, mapped onto a 3D model of human FAM46C, are shown as spheres. The spheres are coloured according to JSD score, which refers to the amino acid conservation in FAM46 family.

ple myeloma (reported in Supplementary Table S4) and for the mutated residue 247 in *hem6* FAM46C is 0.56 and 0.74, respectively (higher JSD scores correspond to higher sequence conservation). In benchmarks JSD approach, which considers also estimated conservation of sequentially neighbouring sites, performed better than traditional measures (e.g. Shannon entropy or Sum-of-pairs measure) in identifying functionally important residues (52). In comparison, the average JSD score for five FAM46 active site residues: glutamic/aspartic acids, glycine and serine is 0.65. Mutations belonging to the second group are located mostly on the protein surface (usually with JSD scores below 0.4), in evolutionary low conserved regions. Those mutations may affect protein–protein interactions or alternatively might not play any crucial role in the reported cancers.

In addition, we selected all mutations of highly conserved residues with a JSD score higher than 0.65 (the average JSD for five FAM46 active site residues from conserved NTase motifs). It allowed us to identify mutations found in a number of malignancies (highlighted in orange in Supplementary Table S4), which probably have the largest impact on protein activity and may be connected with diagnosed cancers. Consequently, we suggest that, in addition to multiple myeloma, FAM46 genes may be also involved in pathogenesis of various other cancer subtypes including liver hepatocellular carcinoma, bladder urothelial carcinoma, head and neck squamous cell carcinoma, uterine corpus endometrial carcinoma, kidney renal papillary cell carcinoma, lung adenocarcinoma, ductal adenocarcinoma, colorectal adenocarcinoma, primary plasma cell leukemia and skin cutaneous melanoma.

Potential function

The results of our sequence and structure analyses suggest that the FAM46 proteins are active NTases, which have both the NTase fold and PAP/OAS1 SBD domains. Active NTases possessing the PAP/OAS1 SBD are known to participate in tRNA maturation (CCA-adding enzymes), RNA degradation (TUTases, poly(A) polymerase in TRAMP complex), mRNA maturation (poly(A) polymerases) and in a defense response to viruses and bacteria (2'-5'-oligoadenylate synthases and cyclic GMP-AMP synthases). Although it was shown that both FAM46A and FAM46C are induced by interferon I and II (14) and that FAM46C is one of the interferon-stimulated genes (ISGs)

which modify viral (YFV and VEEV) replication during infection, it is unlikely that FAM46 proteins are antiviral enzymes like OASes. Unlike replication inhibiting ISGs (such as OASL, Mab-21 and C6orf150), FAM46C slightly enhances the replication of certain viruses (14). FAM46 family members do not possess also an H(X₅)CC(X₆)C motif (conserved among vertebrate cGAS members) located between the NTase and PAP/OAS1 SBD domains. This motif, which resembles most closely HCCC-type zinc-ribbons found in TAZ domains, is required for efficient cytosolic DNA recognition (101). Finally, we investigated the possibility that FAM46 proteins may be novel non-canonical poly(A) polymerases participating in RNA 3' end modification like TUTases or poly(A) polymerases GLD-2 and GLD-3. This hypothesis is consistent with M. Tian studies (100), where it was shown that mutated FAM46C may modulate the poly(A) tails of specific transcripts during erythroid differentiation. The author identified a single FAM46C mutation (Y247N) in *hem6* mice and showed that it might cause an accelerated, progressive shrinkage of the poly(A) tail in four transcripts (alpha-globin, Alas2, Hbb-b1 and Ftl1) and probably does not have any effect on poly(A) tails in two transcripts (Fth1 and beta-actin). Additionally, a Y247N mutation led to an increase of expression levels of 152 transcripts, resulted in a decrease of expression levels of 29 transcripts, and did not have any effect on 29 erythroid transcripts (100). It should be noted, however, that M. Tian used an indirect approach to analyze the poly(A) tail lengths in the six aforementioned transcripts. His strategy of indirect poly(A) tail length assay assumed that the poly(A) tails do not possess any nucleotides other than adenosines; therefore, he was not able to identify the real length of the modified poly(A) tails if they also contain other nucleotides. Thus, it is likely that he observed shortening of mRNA poly(A) tails for specific transcripts if FAM46C is responsible for the addition of adenosines to the RNA 3' end. The FAM46C mutation (Y247N) might have weakened the processivity of FAM46C resulting in poly(A) tails shrinkage. On the other hand, FAM46C may be a non-canonical poly(A) polymerase which adds cytidines or uridines to the RNA 3' end. In this scenario, FAM46C may participate in transcript degradation by modifying the poly(A) tails. This hypothesis is consistent with the fact that up to 152 transcripts increased expression levels in a *hem6* mutant. In this case, the observed shortening of the poly(A) tails in four

transcripts may be a side effect of cell deregulation. In both scenarios, FAM46 proteins may play a very important role in mRNA stability as active non-canonical poly(A) polymerases rather than some other factors, which prevent early mRNA degradation by disrupting interactions between ribonuclease docking complex and RNA as suggested by M. Tian (100). Our functional assignment is also in line with the facts that mouse FAM46C may bind directly or through a complex to RNA CU-rich motifs (100) and FAM46A may bind to poly(A) tails (80). According to Chapman *et al.*, the expression of FAM46C is highly correlated with the expression of ribosomal proteins and initiation and elongation factors involved in protein translation (6). They proposed that FAM46C is functionally related in some way to the regulation of translation (e.g. as a mRNA stability factor), however, they did not assign any exact function to this protein. Recent studies revealed that the poly(A) tail length impacts gene expression in some processes such as inflammation, learning and memory (102), and there is a clear correlation between the poly(A) tail length and translational efficiency in early development stages in zebrafish and African clawed frogs (103). Therefore, it is possible that the correlation observed by Chapman *et al.* is the effect of length change of the poly(A) tails.

Both the M. Tian and Chapman *et al.* studies are consistent with the results of our analysis of FAM46 interactors and interacting partners of all remaining human NTase fold proteins. We found that FAM46 binding partners share a common GO functions and processes mostly with interactors of those active NTase fold superfamily members which belong to non-canonical poly(A) polymerases and poly(A) polymerases. We showed that over half of the 61 identified FAM46 interactors participate in DNA and/or RNA binding, including nine proteins which can bind mRNA poly(A) tails. Many of FAM46 interacting partners are involved in transcription or translation, like transcription (co)factors (RHOXF2, TBX4, NR2F2, SOX5, NRF1, TRIP6), transcription activators (Znf322, Cxxc5), RNA stabilization factors (ELAVL1, BCCIP, HDLBP, Pabpc1, Pabpc4), proteins involved in transcription (POLR1A, POLR1E, POLR2J), proteins participating in mRNA translation (EIF4G3, Pabpc1, Pabpc4), and proteins taking part in transcription regulation (EGLN2, KEAP1, DCAF6, KLHDC2) and mitochondrial translation optimization (MTO1). Finally, some FAM46 protein interactors regulate or are a part of ribonucleoprotein complexes.

Our domain architecture analysis revealed that proteins belonging to FAM46 family possess only two domains: NTase and PAP/OAS1 SBD, with single exceptions of some additional domains present in a few proteins. Importantly, we were not able to detect any additional conserved domains such as ferredoxin-like, which plays a critical role in processivity of canonical poly(A) polymerases or TUTases (Supplementary Figure S3). The ferredoxin-like domain provides additional interactions with RNA and may enhance its binding, allowing the NTase enzyme to add up to several hundred nucleotides. Therefore, FAM46 proteins acting as non-canonical poly(A) polymerases probably can add only a few nucleotides to the RNA 3' end.

FAM46 family members seem to be localized both in the cytoplasm and nucleus, like two other human non-canonical poly(A) polymerases, PAPD4 and PAPD5 (104). Considering the physiological functions of FAM46 interactors, we can speculate about the biological processes, in which FAM46 proteins may participate. FAM46A probably cooperates with a subunit RPB11-a of DNA-directed RNA polymerase II, eukaryotic translation initiation factor 4 gamma (eIF4G), high-density lipoprotein-binding protein (HDLBP, Vigilin), while FAM46C may bind to polyadenylate-binding proteins (Pabpc1, Pabpc4) and (together with FAM46A) to ELAV-like protein 1. As a consequence, proteins belonging to FAM46 family can be involved in mRNA (de)stabilization either in the nucleus or cytoplasm. DNA-directed RNA polymerase II transcribes all protein-coding genes and synthesizes many functional non-coding RNAs. The eIF4G3 subunit is a scaffold protein in eIF4F complex, which participates in the recruitment of eukaryotic mRNAs to the ribosome (105). Pabpc1 and Pabpc4 belong to cytoplasmic poly(A) binding proteins (PABPC), which bind specifically to the poly(A) tail of mRNA and are required for poly(A) shortening, ribosome recruitment and translation initiation (106). Another protein interactor, *Xenopus* Vigilin, can selectively protect *in vitro* vitellogenin mRNA from cleavage by endonuclease PMR-1 (107), while ELAVL1 is described in the literature usually as a stabilization factor, which prevents the degradation of mRNAs possessing short tails (108–110). FAM46 proteins can be also involved in a ribosome biogenesis (like POLR1A, POLR1E interactors (111,112)) or they can (de)stabilize a nuclear pool of extra-ribosomal RPL23 and the pre-60S trans-acting factor eIF6 (like BC-CIP interactor (113)). By interacting with telomerase Cajal body protein 1 (WRAP53), FAM46 family members may change 3' ends of small Cajal body RNAs, which are involved in modifying splicing RNAs (114). Together with the Box C/D snoRNA protein 1 (ZNHIT6), they may also participate in snoRNP biogenesis, which is essential for the processing and modification of rRNA (89). Finally, FAM46 proteins (together with DAZAP2 and major vault proteins (MVP)) may modify RNAs which build ribonucleoprotein complexes like stress granules (85) and vaults composed of MVP, vault poly(adenosine diphosphate-ribose) polymerases (VPARP), telomerase-associated proteins (TEP1) and small untranslated RNAs (vRNAs) (87,88).

The FAM46 family members seem to be highly regulated proteins. The process, in which these new non-canonical poly(A) polymerases participate, is probably determined by their tissue-specific expression and gene organization. As reported in the BioGPS database, tissue expression levels are different for each human FAM46 paralog. Moreover, the human FAM46 proteins are likely to be regulated by phosphorylation. Each human paralog has many phosphorylation patterns detectable with ELM predictor (27) with high probability scores (data not shown). For instance, FAM46A, FAM46B and FAM46C have two potential phosphoserine sites (a LIG.PLK pattern) recognized by the Polo-like kinase, which is a known human FAM46C interacting partner.

CONCLUSION

A comprehensive analysis of various biological information available in literature and databases combined with numerous sequence and structure analyses (including a state-of-the-art distant homology detection, fold recognition and 3D modelling) allowed us to propose that FAM46 members function as cytoplasmic and/or nuclear non-canonical poly(A) polymerases. Four human FAM46 paralogs thus complement the group of already known non-canonical poly(A) polymerases in humans embracing seven proteins: RBM21 (U6 TUTase, Star-PAP, TUT6), hGLD2 (PAPD4, TUT2), hmtPAP (PAPD1, TUT1), POLS (TUT5), PAPD5 (TUT3), ZCCHC6 (TUT7) and ZCCHC11 (TUT4). ZCCHC6 and ZCCHC11 mono-uridylate the 3' end of specific miRNAs involved in cell differentiation and Homeobox (Hox) gene control (115). The hmtPAP produces poly(A) tails in mitochondria (116). The RBM21 catalyzes the uridylation of U6 snRNA involved in pre-mRNA splicing (117). The hGLD2 generates poly(A) tails of selected cytoplasmic mRNAs (118). The PAPD5 participates in the polyadenylation-mediated degradation of aberrant pre-rRNA and in replication-dependent histone mRNA degradation (119). Unfortunately, we are not able to predict the exact type of RNA that can be modified by FAM46 proteins. However, taking into account all the identified FAM46 interacting partners, we can speculate that FAM46 proteins could modify the 3' end of mRNAs, small Cajal body RNAs and vRNAs. In addition, they may also participate in snoRNP and ribosome biogenesis, and (de)stabilize a nuclear pool of extra-ribosomal RPL23 and the pre-60S trans-acting factor eIF6.

The FAM46 family members as well as all the known non-canonical poly(A) polymerases share the two following domains: a PAP/OAS1 SBD with an inserted NTase domain right after the first core α -helix. In this work, we showed that proteins with such domain architecture, in addition to highly conserved NTase domain patterns ([DE]h[DE]h, h[DE]h and hG[GS]), possess also three additional, conserved amino acids critical for NTP binding. These residues embrace serine or threonine in the last α -helix of the NTase domain, and lysine/arginine and a hydrophobic amino acid located in the second and third PAP/OAS1 SBD core α -helix, respectively. Although the FAM46 proteins retain serine or cysteine in the last α -helix of the NTase domain, it is possible that the conserved insertion between the last core β -strand and α -helix in FAM46 NTase domain may substitute the role of the conserved Ser/Thr at least for some family members, enabling them to catalyze the modification of selected RNA 3' ends.

We also performed a systematic search for missense mutations in human FAM46 genes, found in cancer patients. Collected mutation data from various databases and literature, combined with sequence/structure analyses suggest that, in addition to multiple myeloma, FAM46 genes may be also involved in the development of other major malignancies including lung, colorectal, hepatocellular, head and neck, urothelial, endometrial and renal papillary carcinomas and melanoma. We identified several single point mutations of highly conserved FAM46 amino acids that may affect the enzyme catalytic activity, processivity and substrate

binding (e.g. by changing the preference for the type of incorporated NTP). Consequently, these mutations can lead to deregulation of specific RNAs as an oncogenic mechanism in multiple myeloma and other cancers. This is consistent with previous studies which showed a correlation between RNA deregulation (e.g. mRNA (120), microRNA (121,122), long non-coding RNA (123), small non-coding RNA (124)) and various diseases including cancers.

Summarizing, this work provides functional and structural annotation for novel and highly important enzymes involved in RNA metabolism in eukaryotes and thus may guide functional studies of these previously uncharacterized proteins. Further experimental investigations should address the predicted activity and clarify potential substrates to provide more insight into the detailed biological roles of these newly detected non-canonical poly(A) polymerases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Wayne Dawson for proofreading the manuscript.

FUNDING

Foundation for Polish Science [TEAM/2010-6 to K.G.]; Polish National Science Centre [2011/02/A/NZ2/00014 and 2014/15/B/NZ1/03357 to K.G.]; National Centre for Research and Development [INNOCENT-K2/HI2/19/184217/NCBR/13 to L.R.]; European Commission [FP7-KBBE-2011-289646 to L.R.]. A.M. and K.S. were the recipients of the fellowship from the Ministry of Science and Higher Education. Funding for open access charge: Polish National Science Centre [2014/15/B/NZ1/03357].

Conflict of interest statement. None declared.

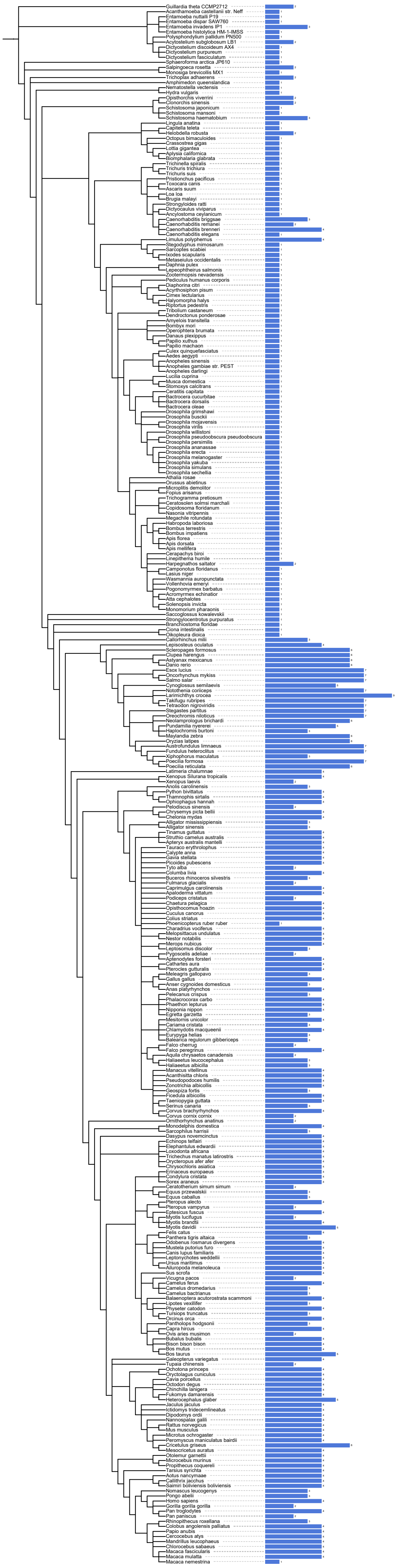
REFERENCES

- Kuchta, K., Knizewski, L., Wyrwicz, L.S., Rychlewski, L. and Ginalski, K. (2009) Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res.*, **37**, 7701–7714.
- Barragan, I., Borrego, S., Abd El-Aziz, M.M., El-Ashry, M.F., Abu-Safieh, L., Bhattacharya, S.S. and Antinolo, G. (2008) Genetic analysis of FAM46A in Spanish families with autosomal recessive retinitis pigmentosa: characterisation of novel VNTRs. *Ann. Hum. Genet.*, **72**, 26–34.
- Etokebe, G.E., Bulat-Kardum, L., Munthe, L.A., Balen, S. and Dembic, Z. (2014) Association of variable number of tandem repeats in the coding region of the FAM46A gene, FAM46A rs11040 SNP and BAG6 rs3117582 SNP with susceptibility to tuberculosis. *PLoS One*, **9**, e91385.
- Benjachat, T., Tongyoo, P., Tantivitayakul, P., Somporn, P., Hirankarn, N., Prom-On, S., Pisitkun, P., Leelahavanichkul, A., Avihingsanon, Y. and Townamchai, N. (2015) Biomarkers for refractory Lupus nephritis: a microarray study of kidney tissue. *Int. J. Mol. Sci.*, **16**, 14276–14290.
- Boyd, K.D., Ross, F.M., Walker, B.A., Wardell, C.P., Tapper, W.J., Chiecchio, L., Dagrada, G., Konn, Z.J., Gregory, W.M., Jackson, G.H. et al. (2011) Mapping of chromosome 1p deletions in myeloma

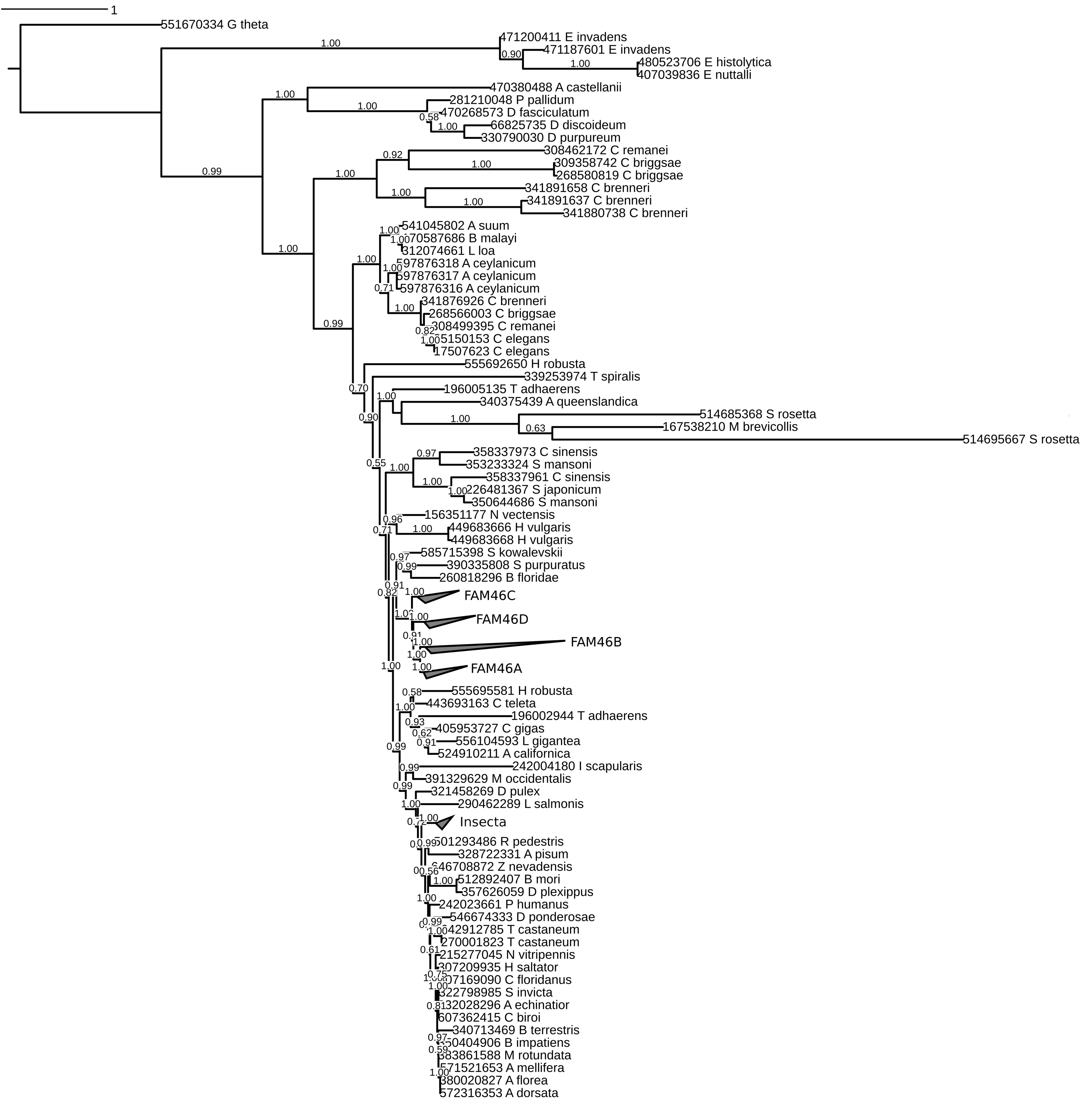
- identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as being genes in regions associated with adverse survival. *Clin. Cancer Res.*, **17**, 7776–7784.
6. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467–472.
 7. Kortum, K.M., Langer, C., Monge, J., Bruins, L., Zhu, Y.X., Shi, C.X., Jedlowski, P., Egan, J.B., Ojha, J., Bullinger, L. *et al.* (2015) Longitudinal analysis of 25 sequential sample-pairs using a custom multiple myeloma mutation sequencing panel (M(3)P). *Ann. Hematol.*, **94**, 1205–1211.
 8. Hamilton, S.M., Spencer, C.M., Harrison, W.R., Yuva-Paylor, L.A., Graham, D.F., Daza, R.A., Hevner, R.F., Overbeek, P.A. and Paylor, R. (2011) Multiple autism-like behaviors in a novel transgenic mouse model. *Behav. Brain Res.*, **218**, 29–41.
 9. Kuehl, W.M. and Bergsagel, P.L. (2012) Molecular pathogenesis of multiple myeloma and its premalignant precursor. *J. Clin. Invest.*, **122**, 3456–3463.
 10. Barbieri, M., Manzoni, M., Fabris, S., Ciceri, G., Todoerti, K., Simeon, V., Musto, P., Cortelezzi, A., Baldini, L., Neri, A. *et al.* (2015) Compendium of FAM46C gene mutations in plasma cell dyscrasias. *Br. J. Haematol.*, doi:10.1111/bjh.13793.
 11. Bettoni, F., Filho, F.C., Grosso, D.M., Galante, P.A., Parmigiani, R.B., Geraldo, M.V., Henrique-Silva, F., Oba-Shinjo, S.M., Marie, S.K., Soares, F.A. *et al.* (2009) Identification of FAM46D as a novel cancer/testis antigen using EST data and serological analysis. *Genomics*, **94**, 153–160.
 12. Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P. and Gauthier, J.M. (2004) Functional proteomics mapping of a human signaling pathway. *Genome Res.*, **14**, 1324–1332.
 13. Etokebe, G.E., Kuchler, A.M., Haraldsen, G., Landin, M., Osmundsen, H. and Dembic, Z. (2009) Family-with-sequence-similarity-46, member A (Fam46a) gene is expressed in developing tooth buds. *Arch. Oral Biol.*, **54**, 1002–1007.
 14. Schoggins, J.W., Wilson, S.J., Panis, M., Murphy, M.Y., Jones, C.T., Bieniasz, P. and Rice, C.M. (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, **472**, 481–485.
 15. Campbell, C.L., Torres-Perez, F., Acuna-Retamar, M. and Schountz, T. (2015) Transcriptome markers of viral persistence in naturally-infected andes virus (bunyaviridae) seropositive long-tailed pygmy rice rats. *PLoS One*, **10**, e0122935.
 16. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
 17. Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.
 18. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
 19. Capella-Gutierrez, S., Silla-Martinez, J.M. and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
 20. Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B. *et al.* (2013) The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.*, **41**, D64–D69.
 21. Pierleoni, A., Martelli, P.L., Fariselli, P. and Casadio, R. (2006) BaCelLo: a balanced subcellular localization predictor. *Bioinformatics*, **22**, e408–e416.
 22. Yu, C.S., Chen, Y.C., Lu, C.H. and Hwang, J.K. (2006) Prediction of protein subcellular localization. *Proteins*, **64**, 643–651.
 23. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, **35**, W585–W587.
 24. Chou, K.C. and Shen, H.B. (2010) A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One*, **5**, e9931.
 25. Hoglund, A., Donnes, P., Blum, T., Adolph, H.W. and Kohlbacher, O. (2006) MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics*, **22**, 1158–1165.
 26. la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K. and Brunak, S. (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng. Des. Sel.*, **17**, 527–536.
 27. Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Kruger, D., Grebnev, G., Kuban, M. *et al.* (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.*, **42**, D259–D266.
 28. Wu, C., Macleod, I. and Su, A.I. (2013) BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res.*, **41**, D561–D565.
 29. McCall, M.N., Uppal, K., Jaffee, H.A., Zilliox, M.J. and Irizarry, R.A. (2011) The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res.*, **39**, D1011–D1015.
 30. Ginalski, K., von Grotthuss, M., Grishin, N.V. and Rychlewski, L. (2004) Detecting distant homology with Meta-BASIC. *Nucleic Acids Res.*, **32**, W576–W581.
 31. Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. U. S. A.*, **95**, 5857–5864.
 32. Finn, R.D., Tate, J., Misty, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
 33. Tatusov, R.L., Galperin, M.Y., Natale, D.A. and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*, **28**, 33–36.
 34. Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S. *et al.* (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.*, **5**, R7.
 35. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
 36. Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
 37. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
 38. Holm, L. and Sander, C. (1996) Mapping the protein universe. *Science*, **273**, 595–603.
 39. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
 40. Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, **19**, 1015–1018.
 41. Ginalski, K. and Rychlewski, L. (2003) Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment. *Proteins*, **53**(Suppl. 6), 410–417.
 42. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
 43. Stagno, J., Aphasizheva, I., Rosengarth, A., Luecke, H. and Aphasizhev, R. (2007) UTP-bound and Apo structures of a minimal RNA uridylyltransferase. *J. Mol. Biol.*, **366**, 882–899.
 44. Wang, Q., Canutescu, A.A. and Dunbrack, R.L. Jr (2008) SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat. Protoc.*, **3**, 1832–1847.
 45. Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, **35**, W407–W410.
 46. UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
 47. Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, **34**, D535–D539.
 48. Gene Ontology Consortium. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.

49. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, pii.
50. Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
51. Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A. and Lopez-Bigas, N. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods*, **10**, 1081–1082.
52. Capra, J.A. and Singh, M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics*, **23**, 1875–1882.
53. Lin, J. (1991) Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory*, **37**, 145–151.
54. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
55. Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.*, **55**, 539–552.
56. Letunic, I. and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.*, **39**, W475–W478.
57. Brunet, F.G., Roest Crolius, H., Paris, M., Aury, J.M., Gibert, P., Jaillon, O., Laudet, V. and Robinson-Rechavi, M. (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol. Biol. Evol.*, **23**, 1808–1816.
58. Peterson, K.J. and Butterfield, N.J. (2005) Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. *Proc. Natl Acad. Sci. U.S.A.*, **102**, 9547–9552.
59. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I. *et al.* (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*, **451**, 783–788.
60. Sun, G., Yang, Z., Ishwar, A. and Huang, J. (2010) Algal genes in the closest relatives of animals. *Mol. Biol. Evol.*, **27**, 2879–2889.
61. Matsuzaka, Y., Tounai, K., Denda, A., Tomizawa, M., Makino, S., Okamoto, K., Keicho, N., Oka, A., Kalski, J.K., Tamiya, G. *et al.* (2002) Identification of novel candidate genes in the diffuse panbronchiolitis critical region of the class I human MHC. *Immunogenetics*, **54**, 301–309.
62. Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Hofer, T. and Rippe, K. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.*, **19**, 1185–1192.
63. Hartmann, R., Justesen, J., Sarkar, S.N., Sen, G.C. and Yee, V.C. (2003) Crystal structure of the 2'-specific and double-stranded RNA-activated interferon-induced antiviral protein 2'-5'-oligoadenylate synthetase. *Mol. Cell*, **12**, 1173–1185.
64. Kuhn, C.D., Wilusz, J.E., Zheng, P.A., Beal, P.A. and Joshua-Tor, L. (2015) On-enzyme refolding permits small RNA and tRNA surveillance by the CCA-adding enzyme. *Cell*, **160**, 644–658.
65. Zhu, D., Wang, L., Shang, G., Liu, X., Zhu, J., Lu, D., Wang, L., Kan, B., Zhang, J.R. and Xiang, Y. (2014) Structural biochemistry of a *Vibrio cholerae* dinucleotide cyclase reveals cyclase activity regulation by folates. *Mol. Cell*, **55**, 931–937.
66. Wolkowicz, U.M. and Cook, A.G. (2012) NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res.*, **40**, 9356–9368.
67. Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L. and Ginalski, K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.*, **40**, 7016–7045.
68. Majorek, K.A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K. and Bujnicki, J.M. (2014) The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.*, **42**, 4160–4179.
69. Sakon, J., Liao, H.H., Kanikula, A.M., Benning, M.M., Rayment, I. and Holden, H.M. (1993) Molecular structure of kanamycin nucleotidyltransferase determined to 3.0-Å resolution. *Biochemistry*, **32**, 11977–11984.
70. Morar, M., Bhullar, K., Hughes, D.W., Junop, M. and Wright, G.D. (2009) Structure and mechanism of the lincosamide antibiotic adenylyltransferase LinB. *Structure*, **17**, 1649–1659.
71. Xu, Y., Zhang, R., Joachimiak, A., Carr, P.D., Huber, T., Vasudevan, S.G. and Ollis, D.L. (2004) Structure of the N-terminal domain of *Escherichia coli* glutamine synthetase adenylyltransferase. *Structure*, **12**, 861–869.
72. Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L. and Bohm, A. (2000) Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science*, **289**, 1346–1349.
73. Lunde, B.M., Magler, I. and Meinhart, A. (2012) Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic Acids Res.*, **40**, 9815–9824.
74. Hamill, S., Wolin, S.L. and Reinisch, K.M. (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc. Natl Acad. Sci. U. S. A.*, **107**, 15045–15050.
75. Losón, O.C., Liu, R., Rome, M.E., Meng, S., Kaiser, J.T., Shan, S.O. and Chan, D.C. (2014) The mitochondrial fission receptor Mif51 requires ADP as a cofactor. *Structure*, **22**, 367–377.
76. Lohofener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D.J. and Fedorov, R. (2015) The activation mechanism of 2'-5'-oligoadenylate synthetase gives new insights into OAS/cGAS triggers of innate immunity. *Structure*, **23**, 851–862.
77. Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G. *et al.* (2013) Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell*, **153**, 1094–1107.
78. Fu, H., Subramanian, R.R. and Masters, S.C. (2000) 14-3-3 proteins: structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.*, **40**, 617–647.
79. Irwin, S., Vandelft, M., Pinchev, D., Howell, J.L., Graczyk, J., Orr, H.T. and Truant, R. (2005) RNA association and nucleocytoplasmic shuttling by ataxin-1. *J. Cell Sci.*, **118**, 233–242.
80. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
81. Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J. *et al.* (2001) Targeting of HIF-1α to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science*, **292**, 468–472.
82. Zhang, D.D. and Hannink, M. (2003) Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Mol. Cell Biol.*, **23**, 8137–8151.
83. Tsai, T.C., Lee, Y.L., Hsiao, W.C., Tsao, Y.P. and Chen, S.L. (2005) NR1P, a novel nuclear receptor interaction protein, enhances the transcriptional activity of nuclear receptors. *J. Biol. Chem.*, **280**, 20000–20009.
84. Zhou, H.J., Wong, C.M., Chen, J.H., Qiang, B.Q., Yuan, J.G. and Jin, D.Y. (2001) Inhibition of LZIP-mediated transcription through direct interaction with a novel host cell factor-like protein. *J. Biol. Chem.*, **276**, 28933–28938.
85. Kim, J.E., Ryu, I., Kim, W.J., Song, O.K., Ryu, J., Kwon, M.Y., Kim, J.H. and Jang, S.K. (2008) Proline-rich transcript in brain protein induces stress granule formation. *Mol. Cell Biol.*, **28**, 803–813.
86. Forbes, M.M., Rothamel, S., Jenny, A. and Marlow, F.L. (2015) Maternal dazap2 regulates germ granules by counteracting dynein in zebrafish primordial germ cells. *Cell Rep.*, **12**, 49–57.
87. Kedersha, N.L. and Rome, L.H. (1986) Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J. Cell Biol.*, **103**, 699–709.
88. Kedersha, N.L., Heuser, J.E., Chugani, D.C. and Rome, L.H. (1991) Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. *J. Cell Biol.*, **112**, 225–235.
89. McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E. and Watkins, N.J. (2007) A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol. Cell Biol.*, **27**, 6782–6793.

90. Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T. and Zheng, N. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature*, **443**, 590–593.
91. Rodrigo-Brenni, M.C., Gutierrez, E. and Hegde, R.S. (2014) Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell*, **55**, 227–237.
92. Arquint, C., Gabryjczyk, A.M., Imseng, S., Bohm, R., Sauer, E., Hiller, S., Nigg, E.A. and Maier, T. (2015) STIL binding to Polo-box 3 of PLK4 regulates centriole duplication. *Elife*, **4**, e07888.
93. Cheng, A.M., Rowley, B., Pao, W., Hayday, A., Bolen, J.B. and Pawson, T. (1995) Syk tyrosine kinase required for mouse viability and B-cell development. *Nature*, **378**, 303–306.
94. Carnevale, J., Ross, L., Puissant, A., Banerji, V., Stone, R.M., DeAngelo, D.J., Ross, K.N. and Stegmaier, K. (2013) SYK regulates mTOR signaling in AML. *Leukemia*, **27**, 2118–2128.
95. Liu, N., Li, Z., Pei, D. and Shu, X. (2013) Zfyve9a regulates the proliferation of hepatic cells during zebrafish embryogenesis. *Int. J. Dev. Biol.*, **57**, 773–778.
96. Fiorillo, C., Moro, F., Yi, J., Weil, S., Brisca, G., Astrea, G., Severino, M., Romano, A., Battini, R., Rossi, A. *et al.* (2014) Novel dynein DYNC1H1 neck and motor domain mutations link distal spinal muscular atrophy and abnormal cortical development. *Hum. Mutat.*, **35**, 298–302.
97. Vallee, R.B., McKenney, R.J. and Ori-McKenney, K.M. (2012) Multiple modes of cytoplasmic dynein regulation. *Nat. Cell Biol.*, **14**, 224–230.
98. Hutagalung, A.H. and Novick, P.J. (2011) Role of Rab GTPases in membrane traffic and cell physiology. *Physiol. Rev.*, **91**, 119–149.
99. Collins, B.M., McCoy, A.J., Kent, H.M., Evans, P.R. and Owen, D.J. (2002) Molecular architecture and functional model of the endocytic AP2 complex. *Cell*, **109**, 523–535.
100. Tian, M. (2010) *The molecular cloning and characterization of Fam46c RNA stability factor*. PhD dissertation, Harvard University.
101. Kranzusch, P.J., Lee, A.S., Berger, J.M. and Doudna, J.A. (2013) Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.*, **3**, 1362–1368.
102. Weill, L., Belloc, E., Bava, F.A. and Mendez, R. (2012) Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat. Struct. Mol. Biol.*, **19**, 577–585.
103. Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H. and Bartel, D.P. (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, **508**, 66–71.
104. Schmidt, M.J. and Norbury, C.J. (2010) Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. *Wiley Interdiscip. Rev. RNA*, **1**, 142–151.
105. Villa, N., Do, A., Hershey, J.W. and Fraser, C.S. (2013) Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J. Biol. Chem.*, **288**, 32932–32940.
106. Kuhn, U. and Wahle, E. (2004) Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta*, **1678**, 67–84.
107. Cunningham, K.S., Dodson, R.E., Nagel, M.A., Shapiro, D.J. and Schoenberg, D.R. (2000) Vigilin binding selectively inhibits cleavage of the vitellogenin mRNA 3'-untranslated region by the mRNA endonuclease polysomal ribonuclease 1. *Proc. Natl Acad. Sci. U. S. A.*, **97**, 12498–12502.
108. Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M. Jr, Tuschl, T., Ohler, U. and Keene, J.D. (2011) Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell*, **43**, 327–339.
109. Fan, X.C. and Steitz, J.A. (1998) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J.*, **17**, 3448–3460.
110. Peng, S.S., Chen, C.Y., Xu, N. and Shyu, A.B. (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J.*, **17**, 3461–3470.
111. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. and Cramer, P. (2013) RNA polymerase I structure and transcription regulation. *Nature*, **502**, 650–655.
112. Laferte, A., Favry, E., Sentenac, A., Riva, M., Carles, C. and Chedin, S. (2006) The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev.*, **20**, 2030–2040.
113. Wyler, E., Wandrey, F., Badertscher, L., Montellese, C., Alper, D. and Kutay, U. (2014) The beta-isoform of the BRCA2 and CDKN1A(p21)-interacting protein (BCCIP) stabilizes nuclear RPL23/uL14. *FEBS Lett.*, **588**, 3685–3691.
114. Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P. and Artandi, S.E. (2009) A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science*, **323**, 644–648.
115. Thornton, J.E., Du, P., Jing, L., Sjekloca, L., Lin, S., Grossi, E., Sliz, P., Zon, L.I. and Gregory, R.I. (2014) Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). *Nucleic Acids Res.*, **42**, 11777–11791.
116. Chang, J.H. and Tong, L. (2012) Mitochondrial poly(A) polymerase and polyadenylation. *Biochim. Biophys. Acta*, **1819**, 992–997.
117. Trippe, R., Guschina, E., Hossbach, M., Urlaub, H., Luhrmann, R. and Benicke, B.J. (2006) Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA*, **12**, 1494–1504.
118. Kwak, J.E., Wang, L., Ballantyne, S., Kimble, J. and Wickens, M. (2004) Mammalian GLD-2 homologs are poly(A) polymerases. *Proc. Natl Acad. Sci. U.S.A.*, **101**, 4407–4412.
119. Mullen, T.E. and Marzluff, W.F. (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev.*, **22**, 50–65.
120. Lokody, I. (2014) RNA dynamics: destabilizing mRNAs promotes metastasis. *Nat. Rev. Cancer*, **14**, 578.
121. Palmero, E.I., de Campos, S.G., Campos, M., de Souza, N.C., Guerreiro, I.D., Carvalho, A.L. and Marques, M.M. (2011) Mechanisms and role of microRNA deregulation in cancer onset and progression. *Genet. Mol. Biol.*, **34**, 363–370.
122. Jansson, M.D. and Lund, A.H. (2012) MicroRNA and cancer. *Mol. Oncol.*, **6**, 590–610.
123. Gao, W., Chan, J.Y. and Wong, T.S. (2014) Long non-coding RNA deregulation in tongue squamous cell carcinoma. *Biomed. Res. Int.*, **2014**, 405860.
124. Ravo, M., Cordella, A., Rinaldi, A., Bruno, G., Alexandrova, E., Saggese, P., Nassa, G., Giurato, G., Tarallo, R., Marchese, G. *et al.* (2015) Small non-coding RNA deregulation in endometrial carcinogenesis. *Oncotarget*, **6**, 4677–4691.



Supplementary Figure S1. Schematic representation of the protein family distribution and abundance among taxa. The bar length is proportional to the number of FAM46 homologs in a given genome. The difference in the number of FAM46 paralogs among vertebrate genomes may be a consequence of either the incompleteness or poor quality of the available proteomes.

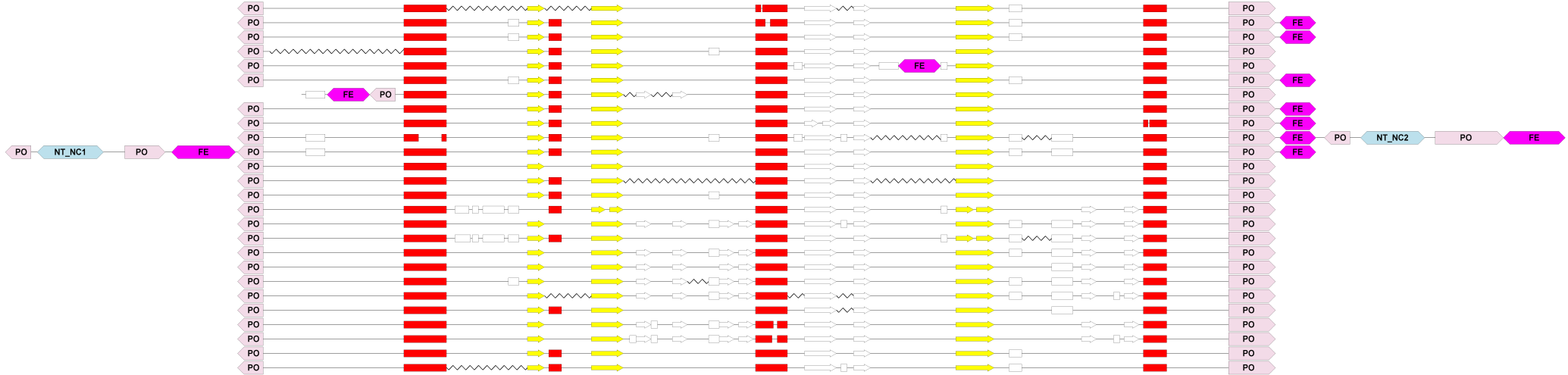


Supplementary Figure S2. Phylogenetic tree of FAM46 protein family. ML analysis was carried out using the JTT+G model. The approximate likelihood ratio test SH-like branch supports above 0.5 are shown. Branches with support lower than 0.4 were collapsed.

NTase NT

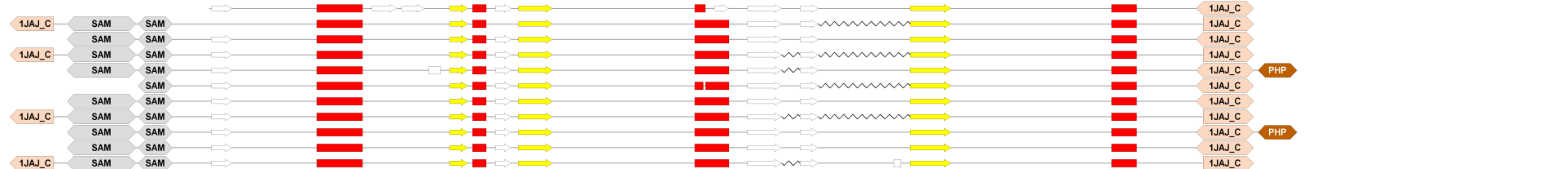
Group I

1px5_A
1q78_A
2o1p_A
2ikf_A
2b4v_A
4it6_A
3pq1_A
3ouy_A
4x4n_A
4m5d_A_NC1
4m5d_A_NC2
3nyb_A
4e7x_A
3hj1_A
4u0n_A
4km5_A
4u03_A
4k8v_C
4ley_A
4o68_A
4jlx_A
4ig8_A
4woy_A
4nxt_A
4at7_A
4at7_B



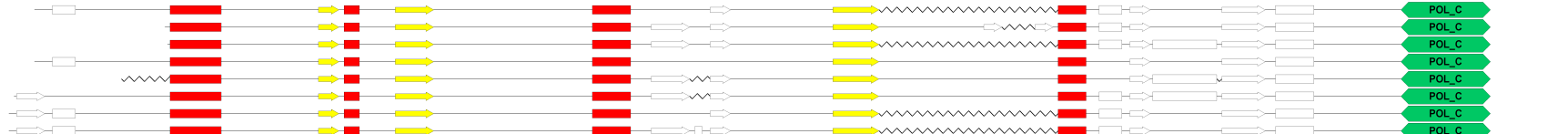
Group II

1jai_A
1jms_A
1xs1_A
2ihm_A
2w9m_A
4p4m_A
4k4g_A
4lzd_A
3au2_A
3v7l_A
4i27_A



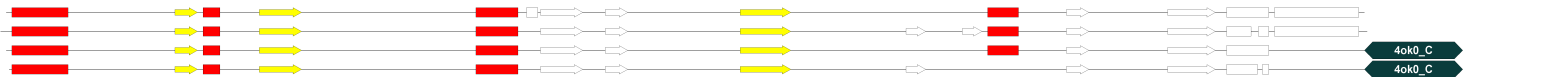
Group III

1ou5_A
1miv_A
1vfg_A
2la3_A
3wfq_E
3wfo_A
3aql_A
3h37_A



Group IV

2ewr_A
4wh5_A
4ok0_A
4wqk_A



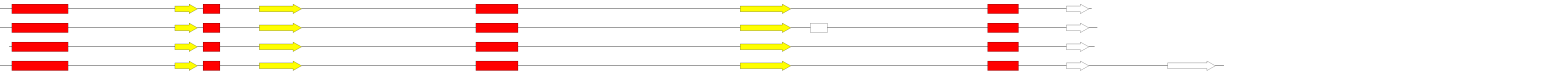
Group V

2pbe_A
3c18_A
3jyy_A
4ebj_A



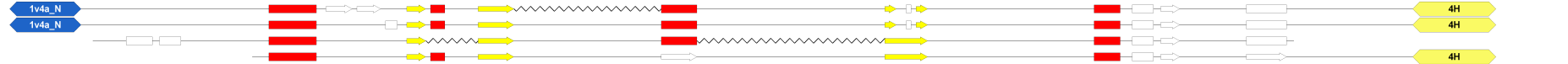
Group VI

1wot_A
1no5_A
1ylq_A
2rff_A



Group VII

1v4a_A
3k7d_A
3nku_A
1kan_A



Group VIII

1vj7_A
2be3_A
3l9d_A



Group IX

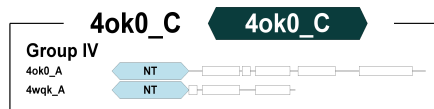
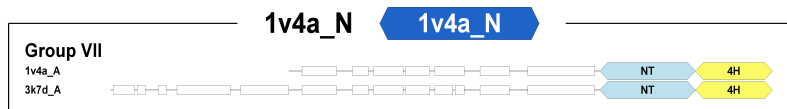
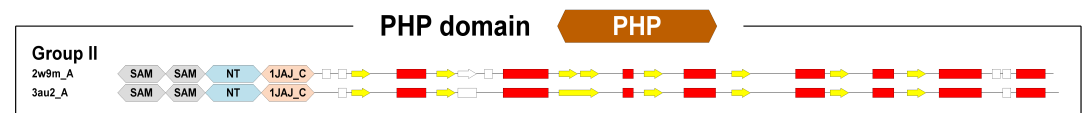
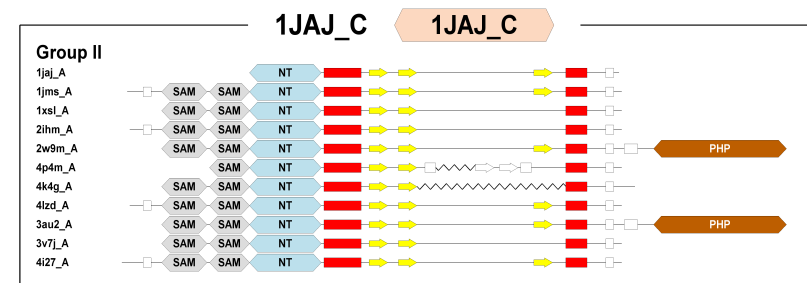
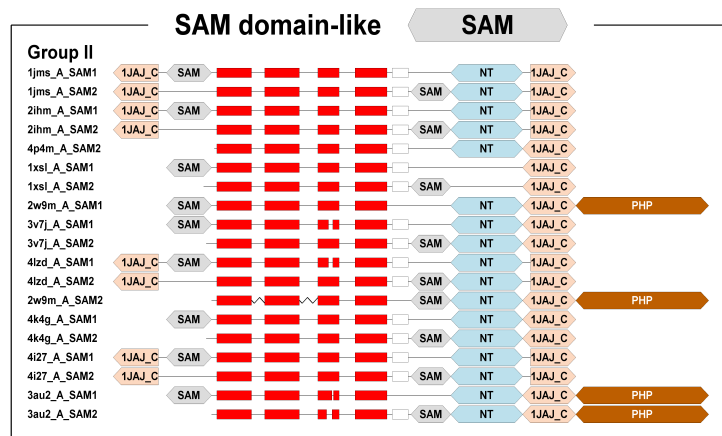
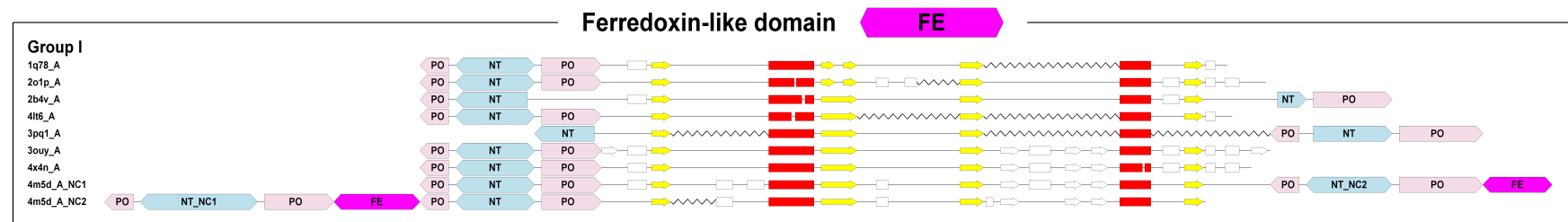
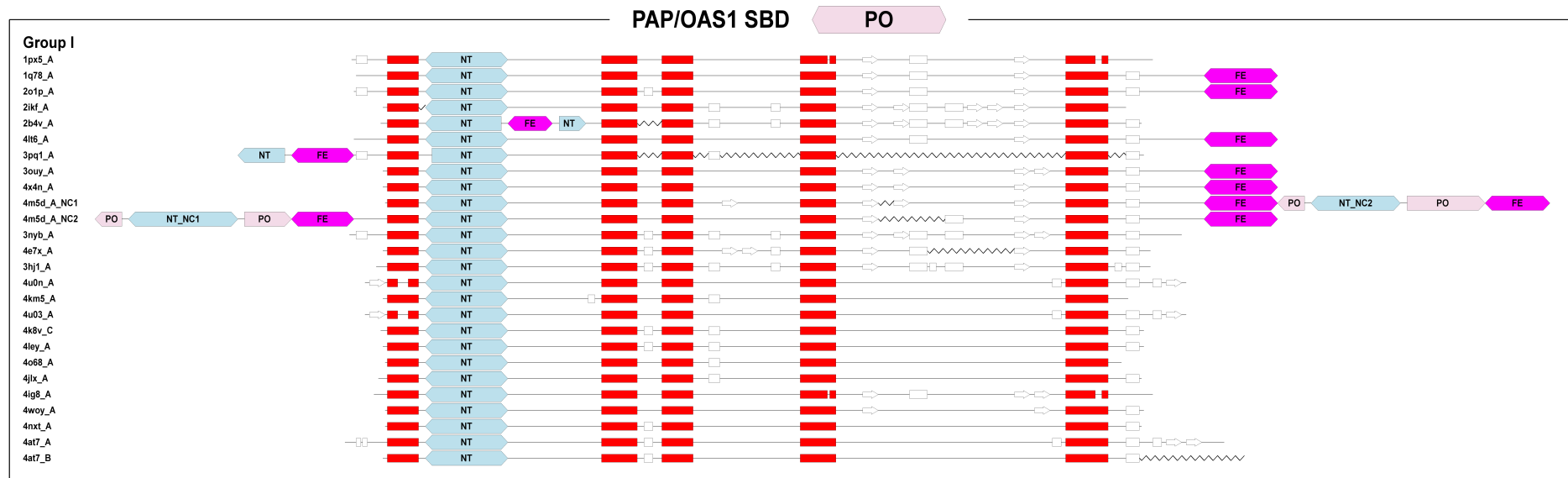
2nrk_A

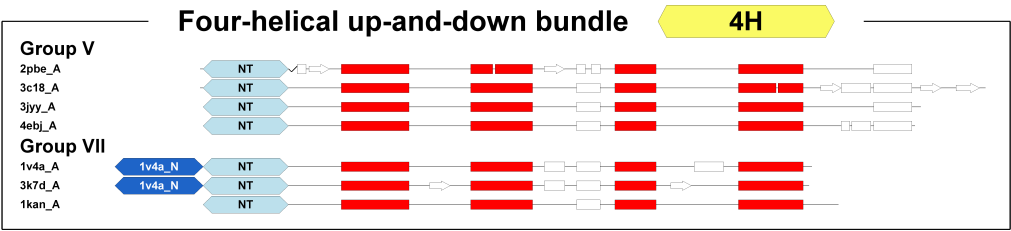
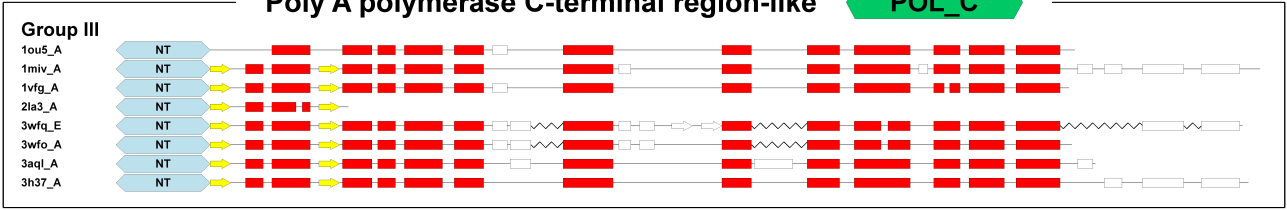


Group X

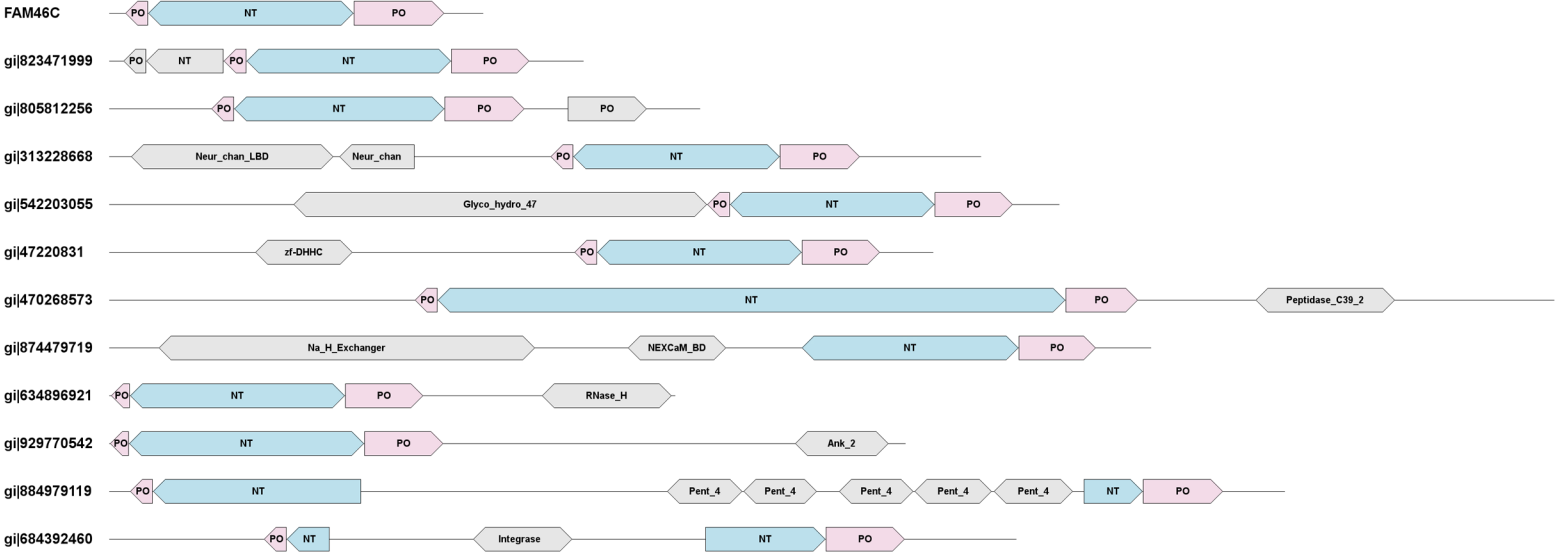
2ga9_D_NC
2ga9_D_CT







Supplementary Figure S3. Summary of NTase fold superfamily structures. Structure-based alignment of secondary structure elements is shown for all domains present in all representative structures. Structures were divided into 10 groups based on their structural similarity. The β -strands and α -helices are shown as arrows and rectangles, respectively. Core α -helices are denoted in red, while core β -strands are colored yellow. Zigzag lines represent unsolved regions in 3D structures. Domains are labeled with following abbreviations: NT, NTase; PO, PAP/OAS1 SBD; FE, Ferredoxin-like domain; SAM, SAM domain-like; 1jaj_C, C-terminal domain of pdb|1jaj; PHP, PHP domain; 1v4a_N, N-terminal domain of pdb|1v4a; 4ok0_C, C-terminal domain of pdb|4ok0; HD, HD-domain/PDEase-like; POL_C, Poly A polymerase C-terminal region-like; 4H, Four-helical up-and-down bundle; Pox_polyA, Poxvirus poly(A) polymerase N-terminal domain. Suffixes '_CT' and '_NC' refer to catalytic and non-catalytic NTase domains.



Supplementary Figure S4. Domain architecture of FAM46 proteins. All possible domain contexts identified in FAM46 family are shown for selected proteins. Domains are labeled with following abbreviations: NT, NTase; PO, PAP/OAS1 SBD; Neur_chan_LBD, Neurotransmitter-gated ion-channel ligand binding domain (PF02931); Neur_chan, Neurotransmitter-gated ion-channel transmembrane region (PF02932); Glyco_hydro_47, Glycosyl hydrolase family 47 (PF01532); zf-DHHC, DHHC palmitoyltransferase (PF01529); Peptidase_C39_2, Peptidase_C39 like family (PF13529); Na_H_Exchanger, Sodium/hydrogen exchanger family (PF00999); NEXaM_BD, Regulatory region of Na⁺/H⁺ exchanger NHE binds to calmodulin (PF16644); RNase_H, RNase H (PF00075); Ank_2, Ankyrin repeats (3 copies) (PF12796); Pent_4, Pentapeptide repeats domains (PF13599); Integrase, Integrase core domain (PF00665).

GO Functions

FA46A	13	28	2	0	14	31	27	16	2	5	17	62	70	7	3	6	16	16	8	10	13	29	13	10	15	5	20	31	11	39	9	14	20	29	4	16	3	19	20	19	39	8		
FA46B	13	16	5	2	0	7	2	7	9	7	1	2	6	13	14	4	0	2	8	7	3	3	2	8	3	3	4	2	3	5	1	9	3	6	8	7	3	8	1	6	4	4	6	2
FA46C	28	5	47	2	0	11	21	15	12	1	3	11	32	39	5	3	4	7	8	6	4	8	15	7	6	9	3	13	17	11	24	6	8	13	22	2	9	2	10	15	9	25	5	
FA46D	2	2	2	2	0	1	1	2	2	1	1	2	2	2	1	0	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	
CB054	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
CGAS	14	2	11	1	0	22	7	6	2	1	3	10	11	11	2	1	2	3	4	3	1	3	9	6	2	7	3	10	11	11	12	5	6	5	8	1	3	2	6	10	5	13	3	
DPOLB	31	7	21	1	0	7	159	33	15	1	5	11	71	76	5	3	5	14	16	5	6	11	29	12	7	14	5	16	31	6	46	6	13	19	27	3	29	3	19	15	19	34	25	
DPOLL	27	9	15	2	0	6	33	83	16	1	4	9	51	51	4	1	6	16	23	13	6	12	21	12	14	13	4	12	28	4	33	5	10	18	27	4	24	4	18	17	20	25	8	
DPOLM	18	7	12	2	0	2	15	16	39	1	2	7	20	26	5	2	4	9	10	3	6	10	12	4	4	7	3	7	13	2	16	3	9	11	13	3	11	1	8	8	8	14	4	
FKRP	2	1	1	1	0	1	1	1	1	11	3	1	2	3	1	0	6	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	
FKTN	5	2	3	1	0	3	5	4	2	3	8	3	6	7	1	0	2	3	3	2	1	2	5	3	2	3	1	2	4	3	6	2	3	3	6	1	3	2	4	3	4	6	2	
GLD2	17	6	11	2	0	10	11	9	7	1	3	24	20	20	3	1	1	6	8	3	2	4	13	6	3	9	2	13	16	10	20	7	12	15	15	2	7	2	8	15	4	15	6	
ILF2	62	13	32	2	0	11	71	51	20	2	6	20	552	405	7	5	7	21	27	12	8	16	58	30	12	35	9	27	55	9	91	11	21	37	65	8	31	4	34	30	33	69	17	
ILF3	70	14	39	2	0	11	76	51	26	3	7	20	515	431	8	7	7	22	31	12	9	18	55	25	12	33	7	28	53	9	107	11	25	38	63	8	31	5	33	31	31	73	18	
IPIL1	7	4	5	1	0	2	5	4	5	1	3	7	8	16	1	2	3	5	2	5	5	4	1	3	3	1	3	3	2	5	1	4	3	4	3	3	1	4	3	4	7	1		
IPIL2	3	0	3	0	0	1	3	1	2	0	0	1	5	7	1	7	0	0	2	1	0	1	2	0	1	2	0	2	3	1	2	1	2	0	2	0	1	0	2	1	4	3	1	
IPRI	6	2	4	1	0	2	5	6	4	6	2	1	7	7	2	0	18	4	2	3	3	4	4	3	2	2	1	1	5	2	4	1	2	6	6	2	3	1	2	2	3	6	1	
MB211	16	8	7	2	0	3	14	16	9	1	3	6	21	22	3	0	4	23	15	6	3	4	15	7	3	5	3	4	13	2	18	2	6	12	14	2	12	1	9	9	6	14	4	
MB212	16	7	8	2	0	4	16	23	10	1	3	8	27	31	5	2	2	15	37	4	2	7	13	4	15	6	3	8	19	3	18	3	8	10	14	2	11	2	10	9	18	16	5	
MB213	8	3	6	1	0	3	5	13	3	1	2	3	12	12	2	1	3	6	4	16	2	3	6	7	3	4	2	4	13	3	8	2	5	5	7	2	3	1	5	5	6	6	3	
MID49	10	3	4	1	0	1	6	6	6	1	1	2	8	9	5	0	3	3	2	2	14	10	5	2	2	5	1	3	4	1	9	3	3	4	6	2	5	1	4	2	4	6	1	
MID51	13	2	8	1	0	3	11	12	10	2	2	4	16	18	5	1	4	4	7	3	10	28	8	4	4	9	2	6	9	3	13	3	6	7	9	2	8	7	7	6	8	13	3	
NOL6	29	8	15	1	0	9	29	21	12	1	5	13	58	55	4	2	4	15	13	6	5	8	73	14	6	15	10	22	33	8	38	8	21	18	31	2	18	2	14	17	12	34	10	
OAS1	13	3	7	1	0	6	12	12	4	1	3	6	30	25	1	0	3	7	4	7	2	4	14	31	3	9	3	10	14	5	16	4	7	11	14	1	7	2	8	6	10	17	5	
OAS2	10	3	6	1	0	2	7	14	4	1	2	3	12	12	3	1	2	3	15	3	2	4	6	3	17	4	2	6	6	2	7	3	5	5	7	2	6	2	4	3	16	7	4	
OAS3	15	4	9	1	0	7	14	13	7	1	3	9	35	33	3	2	2	5	6	4	5	9	15	9	4	40	6	10	15	6	26	7	12	9	16	1	7	3	11	10	12	14	7	
OASL	5	2	3	1	0	3	5	4	3	1	1	2	9	7	1	0	1	3	3	2	1	2	10	3	2	6	14	4	7	2	8	2	3	4	7	1	3	1	3	3	10	7	2	
PAPD1	20	3	13	1	0	10	16	12	7	1	2	13	27	28	3	2	1	4	8	4	3	6	22	10	6	10	4	34	26	10	22	8	19	13	14	1	7	3	8	13	10	19	7	
PAPD5	31	5	17	2	0	11	31	28	13	1	4	16	55	53	3	3	5	13	19	13	4	9	33	14	6	15	7	26	86	11	36	7	21	16	24	1	12	2	18	19	14	29	9	
PAPD7	11	1	1	1	0	1	6	4	2	1	3	10	9	9	2	1	2	2	3	3	1	3	8	5	2	6	2	10	11	11	11	5	6	5	7	1	2	2	3	10	4	11	3	
PAPOA	39	9	24	1	0	12	46	33	16	1	6	20	91	107	5	2	4	18	18	8	9	13	38	16	7	26	8	22	36	11	14	18	26	37	3	21	3	29	23	19	48	12		
PAPOB	9	3	6	1	0	5	6	5	3	1	2	7	11	11	1	1	1	2	3	2	3	3	8	4	3	7	2	8	7	5	14	14	6	6	8	1	5	2	4	5	3	7	3	
PAPOG	14	6	8	1	0	6	13	10	9	1	3	12	21	25	4	2	2	6	8	5	3	6	21	7	5	12	3	19	21	6	18	6	31	13	13	3	9	2	9	12	7	16	7	
STPAP	20	8	13	2	0	5	19	18	11	2	3	15	37	38	3	0	6	12	10	5	4	7	18	11	5	9	4	13	16	5	26	6	13	55	20	4	13	3	12	12	8	23	5	
STRBP	29	7	22	2	0	8	27	27	13	1	6	15	65	63	4	2	6	14	14	7	6	9	31	14	7	16	7	14	24	7	37	8	13	20	86	2	17	2	17	22	2	44	9	
TAM41	4	3	2	1	0	1	3	4	3	1	1	2	8	8	3	0	2	2	2	2	2	2	2	1	2	1	1	1	1	1	3	1	3	4	2	10	3	1	2	1	2	2	1	
TDT	16	8	9	1	0	3	29	24	11	1	3	7	31	31	3	1	3	12	11	3	5	8	18	7	6	7	3	7	12	2	21	5	9	13	17	3	50	3	12	8	8	20	6	
TM102	3	1	2	1	0	2	3	4	1	1	2	2	4	5	1	0	1	2	1	1	1	7	2	2	2	3	1	3	2	2	3	2	2	3	2	1	3	9	3	2	2	3	2	
TRNT1	19	6	10	1	0	6	19	18	8	1	4	8	34	33	4	2	2	9	10	5	4	7	14	8	4	11	3	8	18	3	29	4	9	12	17	2	12	3	77	10	11	25	9	
TUT4	20	4	15	1	0	10	15	17	8	1	3	15	30	31	3	1	2	9	9	5	2	6	17	6	3	10	3	13	19	10	23	5	12	12	22	1	8	2	10	34	6	25	8	
TUT7	19	4	9	1	0	5	19	20	8	1	4	4	33	31	4	4	3	6	18	6	4	8	12	10	16	12	6	10	14	4	19	3	7	8	22	2	8	2	11	6	54	18	6	
ZFR	39	6	25	1	0	13	34	25	14	1	6	15	69	73	7	3	6	14	16	6	6	13	34	17	7	14	7	19	29	11	48	7	16	23	44	2	20	3	25	25	18	116	11	
ZFR2	8	2	5	1	0	3	25	8	4	1	2	6	17	18	1	1	1	4	5	3	1	3	10	5	4	7	2	7	9	3	12	3	7	5	9	1	6	2	9	8	6	11	26	
FA46A	13	28	2	0	14	31	27	16	2	5	17	62	70	7	3	6</																												

B

GO Functions

FAM46	135	89	55	45	44	37	34	30	30	28	27	22	21	17	16	15	14	6	4	3	0
ILF	89	508	118	117	84	77	70	60	54	46	68	36	26	33	20	19	11	8	8	5	0
TUT	55	118	179	68	54	50	49	44	49	45	51	28	19	20	14	14	12	6	4	3	0
PAP	45	117	68	167	52	46	38	44	33	27	40	29	16	21	15	8	12	6	3	3	0
ZFR	44	84	54	52	125	49	44	36	26	22	29	26	14	21	13	13	13	6	2	3	0
POL_BETA	37	77	50	46	49	129	27	29	33	22	21	19	15	29	12	10	7	5	3	3	0
STRBP	34	70	49	38	44	27	86	31	27	21	27	17	13	17	9	10	8	6	2	2	0
NOL6	30	60	44	44	36	29	31	73	21	20	26	14	12	18	9	7	9	5	2	2	0
POL_LAMBDA	30	54	49	33	26	33	27	21	83	36	29	18	16	24	13	9	6	4	4	4	0
MAB	28	46	45	27	22	22	21	20	36	55	24	13	14	15	10	9	6	4	2	2	0
OAS	27	68	51	40	29	21	27	26	29	24	82	12	10	11	11	8	8	3	2	3	0
CCA	22	36	28	29	26	19	17	14	18	13	12	77	8	12	7	7	6	4	2	3	0
POL_MU	21	26	19	16	14	15	13	12	16	14	10	8	39	11	10	8	2	2	3	1	0
DNTT	17	33	20	21	21	29	17	18	24	15	11	12	11	50	9	5	3	3	3	3	0
MID	16	20	14	15	13	12	9	9	13	10	11	7	10	9	32	8	3	3	2	7	0
ITPRIP	15	19	14	8	13	10	10	7	9	9	8	7	8	5	8	38	3	7	3	1	0
cGAS	14	11	12	12	13	7	8	9	6	6	8	6	2	3	3	3	22	3	1	2	0
FUKUTIN	6	8	6	6	6	5	6	5	4	4	3	4	2	3	3	7	3	16	1	2	0
MMP37	4	8	4	3	2	3	2	2	4	2	2	2	3	3	2	3	1	1	10	1	0
TMEM102	3	5	3	3	3	3	2	2	4	2	3	3	1	3	7	1	2	2	1	9	0
C2orf54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAM46	ILF	TUT	PAP	ZFR	POL_BETA	STRBP	NOL6	POL_LAMBDA	MAB	OAS	CCA	POL_MU	DNTT	MID	ITPRIP	cGAS	FUKUTIN	MMP37	TMEM102	C2orf54	

GO Processes

FAM46	407	237	131	119	101	87	65	63	63	56	56	55	53	49	35	33	19	13	11	9	0
ILF	237	1803	373	364	308	264	192	161	218	115	168	114	125	115	66	91	69	15	22	21	0
TUT	131	373	602	190	179	173	109	136	161	112	133	81	75	77	39	44	35	14	15	13	0
PAP	119	364	190	308	138	115	102	67	115	68	99	61	67	50	23	42	26	13	12	11	0
ZFR	101	308	179	138	378	190	101	75	79	49	95	61	86	50	30	36	17	14	12	10	0
STRBP	87	264	173	115	190	342	88	79	69	46	98	63	80	61	22	29	17	9	7	10	0
POL_BETA	65	192	109	102	101	88	270	78	49	45	50	39	68	32	19	28	12	10	9	8	0
POL_LAMBDA	63	161	136	67	75	79	78	273	86	85	42	60	61	34	19	28	16	7	8	7	0
OAS	63	218	161	115	79	69	49	86	296	71	82	47	36	27	18	26	31	5	7	10	0
MAB	56	115	112	68	49	46	45	85	71	195	39	27	29	29	12	25	11	5	8	6	0
NOL6	56	168	133	99	95	98	50	42	82	39	217	30	38	27	13	22	17	8	7	4	0
POL_MU	55	114	81	61	61	63	39	60	47	27	30	191	36	41	20	16	16	3	9	5	0
DNTT	53	125	75	67	86	80	68	61	36	29	38	36	185	31	16	17	12	6	8	5	0
MID	49	115	77	50	50	61	32	34	27	29	27	41	31	179	30	18	12	5	5	18	0
ITPRIP	35	66	39	23	30	22	19	19	18	12	13	20	16	30	112	10	9	2	12	2	0
CCA	33	91	44	42	36	29	28	28	26	25	22	16	17	18	10	178	11	5	4	3	0
MMP37	19	69	35	26	17	17	12	16	31	11	17	16	12	12	9	11	137	2	6	0	0
cGAS	13	15	14	13	14	9	10	7	5	5	8	3	6	5	2	5	2	22	2	2	0
FUKUTIN	11	22	15	12	12	7	9	8	7	8	7	9	8	5	12	4	6	2	37	0	0
TMEM102	9	21	13	11	10	10	8	7	10	6	4	5	5	18	2	3	0	2	0	23	0
C2orf54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FAM46	ILF	TUT	PAP	ZFR	STRBP	POL_BETA	POL_LAMBDA	OAS	MAB	NOL6	POL_MU	DNTT	MID	ITPRIP	CCA	MMP37	cGAS	FUKUTIN	TMEM102	C2orf54	

Supplementary Figure S5. Comparison of human NTase superfamily members according to GO molecular function and biological process of their identified protein interactors. The heat maps provide the numbers of common GO functions and processes. (A) Comparison for all human NTase fold proteins labeled with UniProt names. (B) Comparison for human NTase fold groups described in Supplementary Table S2.

Supplementary Table S1. Selected gene and protein features for all four human FAM46 paralogs. Subcellular localization was predicted using various theoretical methods (consensus prediction).

	FAM46A	FAM46B	FAM46C	FAM46D
Location	Chr6: 81491439-81752774	Chr1: 27005020-27012836	Chr1: 117605934-117628372	ChrX: 80335504-80445311
Exons	3	2	2	3
Coding exons	2	2	1	1
Strand	reverse	reverse	forward	forward
Transcripts	5	1	1	2
Tissue/cell with FAM46 expression	cervix, platelets, pituitary gland, adipose tissue omental, lymph nodes, chorion villus cells, fetal cartilage, endometrium, astrocytes, colonic mucosa, vagina, salivary gland, Cd4+cd25+high regulatory T cells, breast ductal cells, omental adipose tissue, nipple cross section, lymphocytes, breast lobular cells, monocytes, sigmoid colon, cultured endothelial progenitor cells, coronary artery, tongue squamous cells, epididymis corpus, mammary gland, stomach, breast epithelial cells, Cd31+ cells, bronchus, spleen, adipose tissue, alveolar macrophages, gingival epithelium, naive B cells, neutrophils, blood, prostate gland, penis, lung, rectum, breast, Cd3+ T cells, nasopharyngeal epithelial cells, sinus mucosa, liver, sigmoid colon mucosa, promyelocytes, bone marrow, skin, myometrium, synovial membrane, bronchial epithelial cells, parotid gland, saphenous vein, naive Cd4+ T cells, Cd14+ monocytes, ovary, oral mucosa, amniotic fluid, heart atrium, breast stroma, vulva, airway epithelial cells, dorsal root ganglia, adrenal cortex, trachea, thyroid gland, mesenchymal stem cells, Cd34+ cells, cecum, placenta, fallopian tube epithelium, colon,	pharyngeal mucosa, vulva, tonsil epithelium, gingival epithelium, parotid gland, skin, breast ductal cells, vagina, salivary gland, head and neck epithelial cells, breast epithelial cells, nipple cross section, penis, saphenous vein, esophagus, oral mucosa, amniotic fluid, tongue main corpus	cervix, platelets, lymph nodes, tonsil, colonic mucosa, vagina, salivary gland, Cd4+cd25+high regulatory T cells, kidney medulla, deltoid muscle, lymphocytes, breast lobular cells, cultured endothelial progenitor cells, Cd34+ cells, neutrophils, tongue squamous cells, memory B cells, centrocytes, stomach, testes, Cd31+ cells, bronchus, spleen, skeletal muscle, breast ductal cells, Cd4+ T cells, naive B cells, blood, prostate gland, lung, amniotic fluid, breast, Cd3+ T cells, nasopharyngeal epithelial cells, PBMC, sigmoid colon mucosa, promyelocytes, bone marrow, T cells, epididymis corpus, bronchial epithelial cells, plasma cells, naive Cd4+ T cells, Cd14+ monocytes, tonsil epithelium, ovary, heart atrium, breast stroma, Cd49a+ cells, airway epithelial cells, parotid gland, trachea, thyroid gland, vastus lateralis muscle, sperm, cecum, placenta, fallopian tube epithelium, colon, kidney, kidney cortex, endometrial endothelial cell, rectum mucosa, tongue main corpus, sinus mucosa, urethra	sperm

	memory B cells, endometrial endothelial cell, PBMC, rectum mucosa, esophagus, Cd49a+ cells, trigeminal ganglia, urethra			
Cellular localization	cytoplasmic, nuclear	cytoplasmic, nuclear	cytoplasmic, nuclear	cytoplasmic, nuclear

Supplementary Table S2. Human NTase fold proteins. Proteins were clustered into 21 functional groups.

Group	Description	UniProt ID (Name)
FAM46	Family with sequence identity 46	Q96IP4 (FA46A), Q96A09 (FA46B), Q5VWP2 (FA46C), Q8NEK8 (FA46D)
ILF	Interleukin enhancer-binding factor	Q12906 (ILF3), Q12905 (ILF2)
TUT	Non-canonical poly(A) RNA polymerase	Q9NVV4 (PAPD1), Q6PIY7 (GLD2), Q8NDF8 (PAPD5), Q5TAX3 (TUT4), Q5XG87 (PAPD7), Q9H6E5 (STPAP), Q5VYS8 (TUT7)
PAP	Poly(A) polymerase	P51003 (PAPOA), Q9NRJ5 (PAPOB), Q9BWT3 (PAPOG)
MAB	Protein mab-21-like	Q13394 (MB211), Q8N8X9 (MB213), Q9Y586 (MB212)
OAS	2'-5'-oligoadenylate synthase	Q9Y6K5 (OAS3), P29728 (OAS2), P00973 (OAS1), Q15646 (OASL)
cGAS	Cyclic GMP-AMP synthase	Q8N884 (CGAS)
ZFR	Zinc finger RNA-binding protein	Q96KR1 (ZFR), Q9UPR6 (ZFR2)
POL_BETA	DNA polymerase beta	P06746 (DPOLB)
STRBP	Spermatid perinuclear RNA-binding protein	Q96SI9 (STRBP)
POL_LAMBDA	DNA polymerase lambda	Q9UGP5 (DPOLL)
NOL6	Nucleolar protein 6	Q9H6R4 (NOL6)
CCA	CCA tRNA nucleotidyltransferase	Q96Q11 (TRNT1)
POL_MU	DNA-directed DNA/RNA polymerase mu	Q9NP87 (DPOLM)
DNTT	DNA nucleotidylexotransferase	P04053 (TDT)
MID	Mitochondrial dynamics protein	Q96C03 (MID49), Q9NQG6 (MID51)
ITRIP	Inositol 1,4,5-trisphosphate receptor-interacting protein-like	Q3MIP1 (IPIL2), Q8IWB1 (IPRI), Q6GPH6 (IPIL1)
FUKUTIN	Fukutin and Fukutin-related proteins	O75072 (FKTN), Q9H9S5 (FKRP)
MMP37	Phosphatidate cytidyltransferase	Q96BW9 (TAM41)
TMEM102	Transmembrane protein 102	Q8N9M5 (TM102)
C2orf54	Uncharacterized protein C2orf54	Q08AI8 (CB054)

Supplementary Table S3. Protein interacting partners of FAM46 proteins. References to publications, where FAM46A, FAM46B, FAM46C and FAM46D protein interactors were detected are provided in corresponding first four columns.

FAM46A	FAM46B	FAM46C	FAM46D	Protein interactor	Organism	Description of protein interactor
	+	+	+	RHOXF2	human	Rhox homeobox family member 2 (RHOXF2) is encoded by a homeobox gene on the X chromosome. Homeobox genes encode transcription factors that have critical roles in specifying spatial positioning of cells, controlling cellular proliferation and regulating the differentiation of cells during embryonic development (3). The common feature of all homeobox genes is that they encode a homeobox domain, a motif that binds not only to specific DNA but also to specific proteins and sometimes even to RNA (3-5). Human RHOXF2 encodes a protein PEPP-2 and is specifically expressed in testis. It regulates spermatogenesis and neuronal development (6,7). It was shown that PEPP-2 interacts with both itself and various other proteins such as nuclear transcription factors and cytoplasmic RNA decay-promoting proteins (8-10).
	+	+		DAZAP2	human	DAZAP2 was identified as a protein interacting with DAZ and DAZL through DAZ repeats and plays important roles in the physiological responses to various environmental conditions (12). A germ-cell-specific DAZAP2 participates in stress granules (SG) formation in cytosol. During environmental stress, 40S ribosomal subunit complex stalls on mRNA and do not form 80S translational complex. Established RNA-protein complex is stabilized by RNA-binding proteins and/or by the association with DAZAP2 that interacts with eIF4G (13). Recently, it has also been shown that DAZAP2 regulates another ribonucleoprotein complex within zebrafish oocytes - germ granules. In primordial germ cells DAZAP2 limits germ granule Dynein-dependent fragmentation by counteracting Dynein (14).
+	+			ATXN1	human	Abnormal ATXN1 polyglutamine tract occurs in spinocerebellar ataxin type 1 patients. Human Ataxin-1 protein possesses two domains: Capicua transcriptional repressor modulator (PF12547) and AXH domain (PF08517) (16). AXH domain has significant sequence similarity to the transcription factor HBH1 and shares the OB-fold with many oligonucleotide-binding proteins (17). It was shown that ATXN1 has RNA-binding activity <i>in vitro</i> (18), supporting the possible RNA-binding role in the cell. Additionally, ATXN1 interacts with various transcription factors, including Capicua (19), CBF1 (20), cofactors such as LANP that are implicated in transcriptional repression (19), transcriptional corepressor SMRT (21) and selectively with histone deacetylase 3 and 4 (21). In mammals cells, ATXN1 might participate in several Notch signaling pathway controlled developmental and pathological processes (20). Although ATXN1 is a nuclear protein, it is also capable of nuclear export. It was suggested that Ataxin-1 may participate in RNA export as ATXN1 localization to inclusions and inclusion dynamics within the nucleus are RNA and transcription dependent and Ataxin-1 nuclear inclusions recruit the mRNA export factor, TAP/NXF1 (22).
+	+			SH2D2A	human	Human H2 domain-containing protein 2A (SH2D2A, TSA _d) is composed of two domains: SH2 domain (PF00017) and inactive protein tyrosine kinase domain (PF07714). TSA _d participates in actin reorganization in the T cells what is important during differentiation, migration through tissues and proliferation (24-26). It modulates the activity of tyrosine kinase Lck through interaction of its C-terminus with Lck (27,28) as well as together with vascular endothelial (VE) growth factor (VEGF) regulates c-Src tyrosine kinase (29). TSA _d is required for VEGF-induced, c-Src-mediated regulation of endothelial cell junctions and for vascular permeability (29).

+		+		ELAVL1	human	ELAV-like protein 1 (ELAVL1) or Human antigen R (HuR) possesses three highly conserved RNA recognition motifs (RRMs) and may bind mRNA 3'UTR (e.g. cytochrome c 3'UTR (31)), 5'UTR (e.g. p27 5'UTR (32)) through U rich elements and a poly(A) tail (33). Two first N-terminal domains (RRM1 and RRM2) recognize all types of mRNA adenylate- and uridyate-rich elements (AREs) (34) and the third RRM3 domain possibly binds mRNA poly(A) tails (35) what is consistent with RRM3 enzymatic function - terminal adenosyl transferase (36). HuR is described in the literature usually as a stabilization factor, which prevents the degradation of mRNAs possessing short tail (34,37,38). This FAM46 interactor is mainly located within the nucleus but it can shuttle between nucleus and the cytoplasm upon cell stimulation (39). It was reported that HuR may regulate mRNA translation (31,32,40), ARE-mediated mRNA turnover (38), and mRNA splicing (41) and mRNA polyadenylation in neurons by blocking both cleavage and poly(A) addition sites containing U-rich sequences (42).
+		+		BCCIP	human mouse	CDKN1A-interacting protein (BCCIP) takes part in cell cycle control by regulating the expression of the CDK inhibitor p21 (46). Additionally, it is involved in homologous recombination, suppression of spontaneous DNA damage and cell cycle control (47). It was shown that BCCIP is essential in mouse embryonic development (48) and neural development (49). One of two BCCIP mammalian isoforms generated by alternative pre-mRNA splicing - BCCIP β stabilizes a nuclear pool of extra-ribosomal RPL23 and the pre-60S trans-acting factor eIF6 (50). Human BBIP may bind a poly(A) tail (33).
+				TBX4	human	Protein encoded by T-box transcription factor TBX4 (TBX4) possesses conserved DNA-binding T-box domain that play a crucial role in chicken and mice lower limb development (51-53), and in human lower limbs and pelvis developmental pathways (54).
+				CD93	human	Cluster of differentiation 93 (CD93) is a transmembrane glycoprotein whose expression patterns and the structure suggests a role in adhesion and/or leukocyte migration (56,57). It was suggested that CD93 participates in cell-cell interactions during hematopoietic and vascular development (58), and regulation of inflammation and immunity (57).
+				RIN3	human	Ras and Rab interactor 3 (RIN3) is a stimulator and a stabilizer of an active small GTPase Rab5. It plays a role in the transport pathway from plasma membrane to early endosomes (59). Rab proteins belong to small GTPases, which participate in intracellular membrane trafficking (60).
+				SYK	human	Non-receptor tyrosine-protein kinase SYK (SYK) mediates signal transduction and differentiation, particularly in B-cell development (61,62).
+				TNS2	human	Tensin-2 (TNS2) functions as a platform for dis/assembly of signaling complexes at focal adhesions. It is likely that Tensin-2 has actin-, phosphotyrosine- and integrin-binding activities (63). It was shown that mouse tensin is required for the normal renal function (64), is involved in skeletal muscle regeneration (65) and is critical for tension-regulated cell migration (66). Human Tension-2 negatively regulates Akt/PKB signaling cascade and may inhibit cell survival, proliferation and migration (67).
+				NR2F2	human	COUP transcription factor 2 (NR2F2) belongs to the steroid/thyroid hormone receptor superfamily which can act as either transcriptional activators or repressors in a cell type-dependent manner (68). The functions of COUP-TFs in neural development, cell-fate specification, organogenesis, angiogenesis and metabolism have been extensively investigated (68,69).

+				HDLBP	human	High-density lipoprotein-binding protein (HDLBP, vigilin) contains 15 KH domains (PF00013) (70) and is present in both the nucleus and the cytoplasm in the human HEP-2 cells (71). KH domain was first identified in human heterogeneous nuclear ribonucleoprotein (hnRNP) K and is able to bind RNA or single stranded DNA (72,73). The domain was identified in proteins associated with splicing, transcriptional regulation and translational control (73). The yeast vigilin homolog Scp160p was found in polyribosome-bound mRNP complexes including poly(A) binding protein – Pab1p (74). <i>Xenopus</i> vigilin can selectively protect <i>in vitro</i> vitellogenin mRNA from cleavage by endonuclease PMR-1. The high affinity of vigilin for selected mRNA 3'UTRs bearing the PMR-1 cleavages sites directly influences the susceptibility to this RNAs cleavage <i>in vitro</i> (75). However, human HDLBP may interact with the proto-oncogene <i>c-fms</i> mRNA 3'UTR sequence decreasing <i>c-fms</i> mRNA stability and inhibiting <i>c-fms</i> translation (76). Human HDLBP was suggested to be a part of a complex encompassing tRNA and various proteins including translation elongation factors (71) as well as it was shown that vigilin associates with human 80S ribosomes (77). It was also proposed that HDLBP accelerates the export of tRNA from the nucleus to cytosol (71).
+				HTRA1	human	Serine protease HTRA1 belongs to HtrA family and is involved in protein quality control (78). HTRA1 contains two domains: the N-terminal Trypsin-like peptidase domain (PF13365) and the C-terminal PDZ domain (PF13180), which binds the C-termini of mislocalized or misfolded polypeptides (79). HTRA1 is a secreted protease that has the capacity to degrade numerous extracellular matrix proteins and growth factors (80). It was demonstrated that HTRA1 is implicated in many pivotal biological processes in mammals, including proliferation (81), growth (82), apoptosis (83,84) and embryogenesis (85).
+				COL4A2	human	Collagen alpha-2(IV) chain (COL4A2) is a structural constituent of extracellular matrix and is expressed almost exclusively in the basal membrane (86).
+				SPINT1	human	Kunitz-type protease inhibitor 1 (SPINT1, HAI-1) participates in extracellular matrix organization. It acts as an inhibitor of serine proteases such as matrilysin (87,88), hepsin (89,90) and hepatocyte growth factor (HGF) activator (91), which have important roles in epithelial development, cell growth and homeostasis (92).
+				RNF14	human	E3 ubiquitin-protein ligase RNF14 consists of four domains: N-terminal RWD domain (PF05773), zinc finger domain of C3HC4 type (RING finger) (PF13923) and two C-terminal IBR domains (PF01485). RNF14 expressed in insect Sf9 cells catalyzes E2-dependent autoubiquitination <i>in vitro</i> . It was also suggested that it may act as an Ub-ligase (E3) in the ubiquitination of certain nuclear protein(s) (93). It can regulate the expression of cyclin D1 (one of the major enhancers of cell cycle progression) (94) and enhance androgen receptor dependent transcriptional activation (95).
+				DYNC1H1	human	Cytoplasmic dynein 1 heavy chain 1 (DYNC1H1) is a heavy chain of cytoplasmic dynein 1. Cytoplasmic dynein transports diverse intracellular cargoes, controls subcellular targeting and motor activity, and generates forces involved in cell division (96,97).
+				KEAP1	human	Kelch-like ECH-associated protein 1 (KEAP1) targets transcription factor Nrf2, a key regulator for cancer prevention by anticarcinogens, for ubiquitination and degradation by the proteasome under basal conditions (98). KEAP1 functions as a substrate adaptor protein for a Cul3-dependent E3 ubiquitin ligase complex (99).
+				PSG3	human	Pregnancy-specific beta-1-glycoprotein 3 (PSG3) is one of the major placental polypeptides, which is up-regulated upon trophoblast differentiation (100).

+				EGLN2	human	Egl nine homolog 2 (EGLN2, PHD1) is a prolyl hydroxylase, which plays a role in regulation of gene expression by oxygen (101). Its <i>C. elegans</i> homolog EGL-9 targets the hypoxia-inducible transcriptional factors HIF-1 α and HIF-2 α for proteasomal degradation (102).
+				ZFYVE9	human	Zinc finger FYVE domain-containing protein 9 (ZFYVE9) was identified as a SMAD2/3-binding protein, which is required in SMAD2-dependent TGF- β signaling (103). It may regulate also the proliferation of hepatic cells during zebrafish embryogenesis (104).
+				BAG6	human	Mammalian large proline-rich protein (BAG6) takes part in two crucial and distinct pathways: ubiquitin-mediated protein degradation of defective or misallocated polypeptides (105) and tail-anchored (TA) transmembrane protein biogenesis (106). It is essential for survival and maintenance of male germ cells (107).
+				DCAF6	human	DDB1- and CUL4-associated factor 6 (DCAF6) is a nuclear receptor interaction protein, which enhances the transcriptional activity of nuclear receptors NR3C1 and AR (108). It may function as substrate-recruiting module for CUL4-DDB1 E3 ubiquitin-protein ligase complex (109).
+				DDX11	human	DDX11 is a DNA helicase involved in cellular proliferation (110). It possesses both ATPase and DNA helicase activities, which are strictly dependent on DNA, divalent cations and ATP. DDX11 can unwind both DNA/DNA and RNA/DNA substrates (111). The DNA helicase DDX11 is required for sister chromatid cohesion in mammalian cells (thus participates in normal mitotic progression) (112) and viral genome maintenance (113).
+				EIF4G3	human	Eukaryotic translation initiation factor 4 gamma (EIF4G) is a scaffold protein in eIF4F complex. eIF4F complex participates in a recruitment of eukaryotic mRNAs to the ribosome by association of eIF4F complex with a cap structure (114). There are three EIF4G isoforms in mammals (EIF4G1, EIF4G2 and EIF4G3) (115). EIF4G may bind a poly(A) tail (33).
+				ESPL1	human	Separin (ESPL1) is a caspase-like protease which takes part in cohesin cleavage in anaphase. This process is crucial for sister chromatid separation and for the completion of cytokinesis (116).
+				FNDC3A	human	Fibronectin type-III domain-containing protein 3A (FNDC3A) and Fibronectin type III domain-containing protein 3B (FNDC3B) possess few Fibronectin type III domains (PF16893 and PF16893). It was proposed that FNDC3A mediates spermatid-Sertoli adhesion during mouse spermatogenesis (117), while FNDC3B reciprocally regulates differentiation of adipocytes and osteoblast (118).
+				FNDC3B	human	
+				KIAA0922	human	Transmembrane protein 131-like (KIAA0922) is a regulator of thymocyte proliferation and antagonizes canonical Wnt signaling (119).
+				KLHDC2	human	Kelch domain-containing protein 2 (KLHDC2) takes part in transcriptional regulation. The physical interaction between KLHDC2 and a bZIP transcription factor (LZIP) leads to the repression of the LZIP-dependent transcription (120).
+				44M2.3	human	Putative RNA exonuclease NEF-sp consists of three domains: an active 3'-5' exonuclease (PF01612) and two RRM domains (PF00076).

+				MTO1	human	Mitochondrial Translation Optimization Factor 1 (MTO1) is involved in the mitochondrial tRNA modification (121) and controls mitochondrial translation rate in a highly tissue-specific manner (122). In mammals, MTO1 is predominantly expressed in high-energy demand tissue (123). It may bind a poly(A) tail (33).
+				POLE2	human	POLE2 is a DNA polymerase epsilon subunit 2 (124). DNA polymerase epsilon participates in a lagging strand replication, a cell cycle regulation, a double-stranded break repair (DSBR), a base excision repair (BER) and a mismatch repair (NER). It has a DN polymerase and 3'-5' exonuclease activity (125).
+				POLR1A	human	POLR1A is a DNA-directed RNA polymerase I subunit RPA1. The RNA polymerase (Pol) I participates in a ribosome biogenesis and regulates eukaryotic cell growth (126).
+				POLR1E	human	POLR1E is a DNA-directed RNA polymerase I subunit RPA49. The RNA polymerase (Pol) I participates in a ribosome biogenesis and regulates eukaryotic cell growth (126,127).
+				POLR2J	human	POLR2J is a DNA-directed RNA polymerase II subunit RPB11-a. RNA polymerase II (Pol II) transcribes all protein-coding genes and synthesize many functional non-coding RNAs (128).
+				SHKBP1	human	SH3KBP1-binding protein 1 may prevent epidermal growth factor receptor (EGFR) degradation. EGFR plays a critical role in the regulation of many biological responses such as cell proliferation, differentiation and cellular adhesion (129).
+				TMEM39A	human	Transmembrane protein 39A possesses the Tmp39 protein domain (PF10271) which function is still unknown.
+				USP4	human	Ubiquitin carboxyl-terminal hydrolase 4 (USP4) plays a role in mitogen-activated protein (MAP) kinase signaling (130), pre-mRNA splicing (131), controls p53 stability (132), has a DNA damage response role (133) and promotes homologous recombination (134).
+				WDR60	human	It was shown that WD repeat-containing protein 60 (WDR60) mutations could cause skeletal ciliopathies. Therefore, it was suggested that WDR60 plays a role in ciliogenesis (135).
+				WDR62	human	WD repeat-containing protein 62 (WDR62) is probably required for cerebral cortical development and plays a role in neuronal precursors proliferation and migration (136,137).
+				WRAP53	human	Telomerase Cajal body protein 1 (WRAP53) associates with telomerase enzyme, establishes telomerase components, and small Cajal body RNAs that are involved in modifying splicing RNAs (138).
+				ZNHIT6	human	Box C/D snoRNA protein 1 (ZNHIT6) may be involved in biogenesis of snoRNP, which is essential for the processing and modification of rRNA (139).
	+			SOX5	human	Transcription factor SOX-5 (SOX5) is a member of a family of transcription factors involved in regulation of embryonic development (140). It plays a critical role in chondroblast and neuronal progenitor development (141,142) and during skeletal development and bone healing (143-145).

	+			Ywhaz	mouse	14-3-3 protein zeta/delta (Ywhaz) contains 14-3-3 protein domain (PF00244). 14-3-3 proteins have ability to bind diverse signaling proteins including kinases, phosphatases and transmembrane receptors. They play a role in both signal transduction and progression. 14-3-3 proteins participates in the regulation of diverse biological processes, including neuronal development, cell growth control, and viral and bacterial pathogenesis (147). Human Ywhaz may bind a poly(A) tail (33).
		+		NRF1	human	Nuclear respiratory factor 1 (NRF1) regulates pathways in lipid metabolism, amino acid metabolism, proteasomal degradation, the citric acid cycle and the mitochondrial respiratory chain (149).
		+		TRIP6	human	Thyroid receptor-interacting protein 6 (TRIP6) is involved in cytoskeletal organization, coordinated cell migration and tissue invasion (150). In nucleus it acts as a transcriptional cofactor and may contribute to genome stability (150). TRIP6 is engaged in key cellular processes such as cell proliferation, differentiation and survival (150). It may bind a poly(A) tail (33).
		+		PLK4	human	Polo-like kinases (PLK) regulate cell cycle progression, mitosis and cytokinesis (151). Human PLK4 is one of the main players during centriole duplication (151,152). Centrioles are important for cell shape, activity and segregation of duplicated chromosomes during cell division (151,153).
		+		MVP	human	In mammals, vaults (evolutionary highly conserved ribonucleoproteins particles) are composed of major vault proteins (MVP), vault poly(adenosine diphosphate-ribose) polymerases (VPARP), telomerase-associated proteins (TEP1) and small untranslated RNAs (vRNA) (155,156). Vaults are involved in a diversity of cellular processes, including multidrug resistance, transport mechanisms, signal transmission and immune response (157,158). It was shown that some human vRNAs produce regulatory small RNAs (svRNAs) in a Dicer-dependent mechanism (159).
		+		PARK2	human	E3 ubiquitin-protein ligase parkin (PARK2) functions within E3 ubiquitin ligase complex and is involved in protein degradation (160). It participates in the removal and/or detoxification of abnormally folded or damaged proteins (161), regulates autophagy (162), protects from apoptosis by the transcriptional repression of p53/TP53 (163) and could enhance cell viability and protect cells from oxidative stress (164).
		+		UBE2I	human	SUMO-conjugating enzyme UBC9 encoded by UBE2I takes part in reminiscent of ubiquitination - sumoylation process. This post-transcriptional modification adds a small ubiquitin-like modifier (SUMO) polypeptide to the ε-amino group of a lysine residue (167). Sumoylation governs cellular processes such as DNA replication, cell division (168), the DNA damage response (169), transcription and nuclear trafficking (170,171). In mice, Ubc9 is essential for embryonic stem cell (ESC) survival, playing a critical role in reprogramming fibroblasts to iPS cells and maintaining ESC pluripotency (167). Ubc9-deficient mice embryos die at the early postimplantation stage (171,172). Loss of Ubc9 leads to major chromosome condensation and segregation defects (171,172). Deregulation of sumoylation has been found in different human disorders and diseases (88), such as neurodegenerative diseases (173,174), heart failure (175) and cancer (176). UBE2I may bind a poly(A) tail (33).
		+		AP2B1	human	AP2B1 encodes AP2 complex subunit beta, which functions in clathrin-mediated endocytosis. Thus, AP2 plays a pivotal role in many vesicle trafficking pathways within the cell (177).
		+		Znf322	mouse	Zinc finger protein 322 (Znf322) regulates transcriptional activation in MAPK signaling pathways (179).

		+		Rad23b	rat	UV excision repair protein RAD23 homolog B (Rad23b) is involved in DNA repair and protein degradation. The protein degradation function of RAD23 contributes to cell cycle progression, stress response, endoplasmic reticulum proteolysis and DNA repair (181).
		+		Psmc3	mouse	Psmc3 is a regulatory subunit 6A of the 26S proteasome, which catalyzes degradation of ubiquitinated proteins (182). PSMC3 may control cell proliferation (175) and as a 19S protein component can participate in transcription regulation (183), nucleotide excision repair (184) and a regulation of mitosis (185).
		+		Cxxc5	mouse	CXXC-type zinc finger protein 5 (Cxxc5) is a nuclear factor (186), a negative regulator for Wnt/ β -catenin signaling involved in osteoblast differentiation (187) and transcription activator for myelin genes on oligodendrocyte differentiation (188).
		+		Immunoglobulin heavy chain V gene segment	mouse	We were not able to identify protein id via protein description, thus we could not recognize the specific function of detected interactor. However, we can speculate that immunoglobulin heavy chain V is probably a subunit of an antibody, which has a crucial role in the immune system (189).
		+		Pabpc1	mouse	Polyadenylate-binding protein 1 (Pabpc1) and Polyadenylate-binding protein 4 (Pabpc4) belong to cytoplasmic poly(A) binding proteins (PABPC), which bind specifically to the poly(A) tail of mRNA and are required for poly(A) shortening, ribosome recruitment and translation initiation (190). PABC binds eukaryotic translation initiation factor 4 gamma (eIF4G), which recruits the 40S ribosomal subunit (191). PABPC1 antagonizes uridylation of polyadenylated mRNAs (192) and may be responsible for the microRNA mediated gene silencing (193). Pabpc1 are differentially expressed during male germ cell development (194) while Pabpc4 serves a critical role in erythroid differentiation (195).
		+		Pabpc4	mouse	
			+	C1orf94	human	Uncharacterized

References

1. Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173-1178.
2. Rolland, T., Tasan, M., Charlotiaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R. *et al.* (2014) A proteome-scale map of the human interactome network. *Cell*, **159**, 1212-1226.
3. MacLean, J.A., 2nd and Wilkinson, M.F. (2010) The RhoX genes. *Reproduction*, **140**, 195-213.
4. Shyu, A.B. and Wilkinson, M.F. (2000) The double lives of shuttling mRNA binding proteins. *Cell*, **102**, 135-138.
5. Svingen, T. and Tonissen, K.F. (2006) Hox transcription factors and their elusive mammalian gene targets. *Heredity (Edinb)*, **97**, 88-96.
6. Wayne, C.M., MacLean, J.A., Cornwall, G. and Wilkinson, M.F. (2002) Two novel human X-linked homeobox genes, hPEPP1 and hPEPP2, selectively expressed in the testis. *Gene*, **301**, 1-11.
7. Niu, A.L., Wang, Y.Q., Zhang, H., Liao, C.H., Wang, J.K., Zhang, R., Che, J. and Su, B. (2011) Rapid evolution and copy number variation of primate RHOXF2, an X-linked homeobox gene involved in male reproduction and possibly brain function. *BMC Evol Biol*, **11**, 298.

8. Lehner, B. and Sanderson, C.M. (2004) A protein interaction framework for human mRNA degradation. *Genome Res*, **14**, 1315-1323.
9. Prasad, T.S., Kandasamy, K. and Pandey, A. (2009) Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol Biol*, **577**, 67-79.
10. Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabo, G., Rual, J.F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E. *et al.* (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*, **125**, 801-814.
11. Venkatesan, K., Rual, J.F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I. *et al.* (2009) An empirical framework for binary interactome mapping. *Nat Methods*, **6**, 83-90.
12. Tsui, S., Dai, T., Roettger, S., Schempp, W., Salido, E.C. and Yen, P.H. (2000) Identification of two novel proteins that interact with germ-cell-specific RNA-binding proteins DAZ and DAZL1. *Genomics*, **65**, 266-273.
13. Kim, J.E., Ryu, I., Kim, W.J., Song, O.K., Ryu, J., Kwon, M.Y., Kim, J.H. and Jang, S.K. (2008) Proline-rich transcript in brain protein induces stress granule formation. *Mol Cell Biol*, **28**, 803-813.
14. Forbes, M.M., Rothhamel, S., Jenny, A. and Marlow, F.L. (2015) Maternal dazap2 Regulates Germ Granules by Counteracting Dynein in Zebrafish Primordial Germ Cells. *Cell Rep*, **12**, 49-57.
15. Suter, B., Fontaine, J.F., Yildirimman, R., Rasko, T., Schaefer, M.H., Rasche, A., Porras, P., Vazquez-Alvarez, B.M., Russ, J., Rau, K. *et al.* (2013) Development and application of a DNA microarray-based yeast two-hybrid system. *Nucleic Acids Res*, **41**, 1496-1507.
16. Banfi, S., Servadio, A., Chung, M.Y., Kwiatkowski, T.J., Jr., McCall, A.E., Duvick, L.A., Shen, Y., Roth, E.J., Orr, H.T. and Zoghbi, H.Y. (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat Genet*, **7**, 513-520.
17. Chen, Y.W., Allen, M.D., Veprintsev, D.B., Lowe, J. and Bycroft, M. (2004) The structure of the AXH domain of spinocerebellar ataxin-1. *J Biol Chem*, **279**, 3758-3765.
18. Yue, S., Serra, H.G., Zoghbi, H.Y. and Orr, H.T. (2001) The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Hum Mol Genet*, **10**, 25-30.
19. Lam, Y.C., Bowman, A.B., Jafar-Nejad, P., Lim, J., Richman, R., Fryer, J.D., Hyun, E.D., Duvick, L.A., Orr, H.T., Botas, J. *et al.* (2006) ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell*, **127**, 1335-1347.
20. Tong, X., Gui, H., Jin, F., Heck, B.W., Lin, P., Ma, J., Fondell, J.D. and Tsai, C.C. (2011) Ataxin-1 and Brother of ataxin-1 are components of the Notch signalling pathway. *EMBO Rep*, **12**, 428-435.
21. Tsai, C.C., Kao, H.Y., Mitzutani, A., Banayo, E., Rajan, H., McKeown, M. and Evans, R.M. (2004) Ataxin 1, a SCA1 neurodegenerative disorder protein, is functionally linked to the silencing mediator of retinoid and thyroid hormone receptors. *Proc Natl Acad Sci U S A*, **101**, 4047-4052.
22. Irwin, S., Vandelft, M., Pinchev, D., Howell, J.L., Graczyk, J., Orr, H.T. and Truant, R. (2005) RNA association and nucleocytoplasmic shuttling by ataxin-1. *J Cell Sci*, **118**, 233-242.
23. Grossmann, A., Benlasfer, N., Birth, P., Hegele, A., Wachsmuth, F., Apelt, L. and Stelzl, U. (2015) Phospho-tyrosine dependent protein-protein interaction network. *Mol Syst Biol*, **11**, 794.

24. Hem, C.D., Sundvold-Gjerstad, V., Granum, S., Koll, L., Abrahamsen, G., Buday, L. and Spurkland, A. (2015) T cell specific adaptor protein (TSAd) promotes interaction of Nck with Lck and SLP-76 in T cells. *Cell Commun Signal*, **13**, 31.
25. Burkhardt, J.K., Carrizosa, E. and Shaffer, M.H. (2008) The actin cytoskeleton in T cell activation. *Annu Rev Immunol*, **26**, 233-259.
26. Matsumoto, T., Bohman, S., Dixelius, J., Berge, T., Dimberg, A., Magnusson, P., Wang, L., Wikner, C., Qi, J.H., Wernstedt, C. *et al.* (2005) VEGF receptor-2 Y951 signaling and a role for the adapter molecule TSAd in tumor angiogenesis. *EMBO J*, **24**, 2342-2353.
27. Sundvold-Gjerstad, V., Granum, S., Mustelin, T., Andersen, T.C., Berge, T., Shapiro, M.J., Shapiro, V.S., Spurkland, A. and Lea, T. (2005) The C terminus of T cell-specific adapter protein (TSAd) is necessary for TSAd-mediated inhibition of Lck activity. *Eur J Immunol*, **35**, 1612-1620.
28. Granum, S., Andersen, T.C., Sorlie, M., Jorgensen, M., Koll, L., Berge, T., Lea, T., Fleckenstein, B., Spurkland, A. and Sundvold-Gjerstad, V. (2008) Modulation of Lck function through multisite docking to T cell-specific adapter protein. *J Biol Chem*, **283**, 21909-21919.
29. Sun, Z., Li, X., Massena, S., Kutschera, S., Padhan, N., Gualandi, L., Sundvold-Gjerstad, V., Gustafsson, K., Choy, W.W., Zang, G. *et al.* (2012) VEGFR2 induces c-Src signaling and vascular permeability in vivo via the adaptor protein TSAd. *J Exp Med*, **209**, 1363-1377.
30. Abdelmohsen, K., Srikantan, S., Yang, X., Lal, A., Kim, H.H., Kuwano, Y., Galban, S., Becker, K.G., Kamara, D., de Cabo, R. *et al.* (2009) Ubiquitin-mediated proteolysis of HuR by heat shock. *EMBO J*, **28**, 1271-1282.
31. Kawai, T., Lal, A., Yang, X., Galban, S., Mazan-Mamczarz, K. and Gorospe, M. (2006) Translational control of cytochrome c by RNA-binding proteins TIA-1 and HuR. *Mol Cell Biol*, **26**, 3295-3307.
32. Kullmann, M., Gopfert, U., Siewe, B. and Hengst, L. (2002) ELAV/Hu proteins inhibit p27 translation via an IRES element in the p27 5'UTR. *Genes Dev*, **16**, 3087-3099.
33. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393-1406.
34. Fan, X.C. and Steitz, J.A. (1998) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J*, **17**, 3448-3460.
35. Ma, W.J., Chung, S. and Furneaux, H. (1997) The Elav-like proteins bind to AU-rich elements and to the poly(A) tail of mRNA. *Nucleic Acids Res*, **25**, 3564-3569.
36. Meisner, N.C., Hintersteiner, M., Seifert, J.M., Bauer, R., Benoit, R.M., Widmer, A., Schindler, T., Uhl, V., Lang, M., Gstach, H. *et al.* (2009) Terminal adenosyl transferase activity of posttranscriptional regulator HuR revealed by confocal on-bead screening. *J Mol Biol*, **386**, 435-450.
37. Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Jr., Tuschl, T., Ohler, U. and Keene, J.D. (2011) Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell*, **43**, 327-339.
38. Peng, S.S., Chen, C.Y., Xu, N. and Shyu, A.B. (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J*, **17**, 3461-3470.
39. Doller, A., Pfeilschifter, J. and Eberhardt, W. (2008) Signalling pathways regulating nucleo-cytoplasmic shuttling of the mRNA-binding protein HuR. *Cell Signal*, **20**, 2165-2173.

40. Galban, S., Kuwano, Y., Pullmann, R., Jr., Martindale, J.L., Kim, H.H., Lal, A., Abdelmohsen, K., Yang, X., Dang, Y., Liu, J.O. *et al.* (2008) RNA-binding proteins HuR and PTB promote the translation of hypoxia-inducible factor 1alpha. *Mol Cell Biol*, **28**, 93-107.
41. Izquierdo, J.M. (2008) Hu antigen R (HuR) functions as an alternative pre-mRNA splicing regulator of Fas apoptosis-promoting receptor on exon definition. *J Biol Chem*, **283**, 19077-19084.
42. Zhu, H., Zhou, H.L., Hasman, R.A. and Lou, H. (2007) Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J Biol Chem*, **282**, 2203-2210.
43. Colland, F., Jacq, X., Trouplin, V., Mouglin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P. and Gauthier, J.M. (2004) Functional proteomics mapping of a human signaling pathway. *Genome Res*, **14**, 1324-1332.
44. Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K. *et al.* (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*, **162**, 425-440.
45. Tian, M. (2010) The molecular cloning and characterization of Fam46c RNA stability factor. *PhD dissertation*, UMI 3415385, Harvard University.
46. Meng, X., Lu, H. and Shen, Z. (2004) BCCIP functions through p53 to regulate the expression of p21Waf1/Cip1. *Cell Cycle*, **3**, 1457-1462.
47. Lu, H., Yue, J., Meng, X., Nickoloff, J.A. and Shen, Z. (2007) BCCIP regulates homologous recombination by distinct domains and suppresses spontaneous DNA damage. *Nucleic Acids Res*, **35**, 7160-7170.
48. Lu, H., Huang, Y.Y., Mehrotra, S., Droz-Rosario, R., Liu, J., Bhaumik, M., White, E. and Shen, Z. (2011) Essential roles of BCCIP in mouse embryonic development and structural stability of chromosomes. *PLoS Genet*, **7**, e1002291.
49. Huang, Y.Y., Lu, H., Liu, S., Droz-Rosario, R. and Shen, Z. (2012) Requirement of mouse BCCIP for neural development and progenitor proliferation. *PLoS One*, **7**, e30638.
50. Wyler, E., Wandrey, F., Badertscher, L., Montellese, C., Alper, D. and Kutay, U. (2014) The beta-isoform of the BRCA2 and CDKN1A(p21)-interacting protein (BCCIP) stabilizes nuclear RPL23/uL14. *FEBS Lett*, **588**, 3685-3691.
51. Rodriguez-Esteban, C., Tsukui, T., Yonei, S., Magallon, J., Tamura, K. and Izpisua Belmonte, J.C. (1999) The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity. *Nature*, **398**, 814-818.
52. Naiche, L.A. and Papaioannou, V.E. (2003) Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. *Development*, **130**, 2681-2693.
53. Takeuchi, J.K., Koshiba-Takeuchi, K., Suzuki, T., Kamimura, M., Ogura, K. and Ogura, T. (2003) Tbx5 and Tbx4 trigger limb initiation through activation of the Wnt/Fgf signaling cascade. *Development*, **130**, 2729-2739.
54. Bongers, E.M., Duijf, P.H., van Beersum, S.E., Schoots, J., Van Kampen, A., Burckhardt, A., Hamel, B.C., Losan, F., Hoefsloot, L.H., Yntema, H.G. *et al.* (2004) Mutations in the human TBX4 gene cause small patella syndrome. *Am J Hum Genet*, **74**, 1239-1248.
55. Wang, J., Huo, K., Ma, L., Tang, L., Li, D., Huang, X., Yuan, Y., Li, C., Wang, W., Guan, W. *et al.* (2011) Toward an understanding of the protein interaction network of the human liver. *Mol Syst Biol*, **7**, 536.

56. Greenlee, M.C., Sullivan, S.A. and Bohlson, S.S. (2008) CD93 and related family members: their role in innate immunity. *Curr Drug Targets*, **9**, 130-138.
57. Greenlee-Wacker, M.C., Galvan, M.D. and Bohlson, S.S. (2012) CD93: recent advances and implications in disease. *Curr Drug Targets*, **13**, 411-420.
58. Petrenko, O., Beavis, A., Klaine, M., Kittappa, R., Godin, I. and Lemischka, I.R. (1999) The molecular characterization of the fetal stem cell marker AA4. *Immunity*, **10**, 691-700.
59. Kajiho, H., Saito, K., Tsujita, K., Kontani, K., Araki, Y., Kurosu, H. and Katada, T. (2003) RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J Cell Sci*, **116**, 4159-4168.
60. Hutagalung, A.H. and Novick, P.J. (2011) Role of Rab GTPases in membrane traffic and cell physiology. *Physiol Rev*, **91**, 119-149.
61. Cheng, A.M., Rowley, B., Pao, W., Hayday, A., Bolen, J.B. and Pawson, T. (1995) Syk tyrosine kinase required for mouse viability and B-cell development. *Nature*, **378**, 303-306.
62. Carnevale, J., Ross, L., Puissant, A., Banerji, V., Stone, R.M., DeAngelo, D.J., Ross, K.N. and Stegmaier, K. (2013) SYK regulates mTOR signaling in AML. *Leukemia*, **27**, 2118-2128.
63. Lo, S.H. (2004) Tensin. *Int J Biochem Cell Biol*, **36**, 31-34.
64. Lo, S.H., Janmey, P.A., Hartwig, J.H. and Chen, L.B. (1994) Interactions of tensin with actin and identification of its three distinct actin-binding domains. *J Cell Biol*, **125**, 1067-1075.
65. Ishii, A. and Lo, S.H. (2001) A role of tensin in skeletal-muscle regeneration. *Biochem J*, **356**, 737-745.
66. Chen, H. and Lo, S.H. (2003) Regulation of tensin-promoted cell migration by its focal adhesion binding and Src homology domain 2. *Biochem J*, **370**, 1039-1045.
67. Hafizi, S., Ibraimi, F. and Dahlback, B. (2005) C1-TEN is a negative regulator of the Akt/PKB signal transduction pathway and inhibits cell survival, proliferation, and migration. *FASEB J*, **19**, 971-973.
68. Lin, F.J., Qin, J., Tang, K., Tsai, S.Y. and Tsai, M.J. (2011) Coup d'Etat: an orphan takes control. *Endocr Rev*, **32**, 404-421.
69. Alfano, C., Magrinelli, E., Harb, K. and Studer, M. (2014) The nuclear receptors COUP-TF: a long-lasting experience in forebrain assembly. *Cell Mol Life Sci*, **71**, 43-62.
70. Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M.A., Nilges, M., Gibson, T.J. and Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, **85**, 237-245.
71. Kruse, C., Willkomm, D.K., Grunweller, A., Vollbrandt, T., Sommer, S., Busch, S., Pfeiffer, T., Brinkmann, J., Hartmann, R.K. and Muller, P.K. (2000) Export and transport of tRNA are coupled to a multi-protein complex. *Biochem J*, **346**, 107-115.
72. Garcia-Mayoral, M.F., Hollingworth, D., Masino, L., Diaz-Moreno, I., Kelly, G., Gherzi, R., Chou, C.F., Chen, C.Y. and Ramos, A. (2007) The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation. *Structure*, **15**, 485-498.
73. Valverde, R., Edwards, L. and Regan, L. (2008) Structure and function of KH domains. *FEBS J*, **275**, 2712-2726.

74. Lang, B.D. and Fridovich-Keil, J.L. (2000) Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. *Nucleic Acids Res*, **28**, 1576-1584.
75. Cunningham, K.S., Dodson, R.E., Nagel, M.A., Shapiro, D.J. and Schoenberg, D.R. (2000) Vigilin binding selectively inhibits cleavage of the vitellogenin mRNA 3'-untranslated region by the mRNA endonuclease polysomal ribonuclease 1. *Proc Natl Acad Sci U S A*, **97**, 12498-12502.
76. Woo, H.H., Yi, X., Lamb, T., Menzl, I., Baker, T., Shapiro, D.J. and Chambers, S.K. (2011) Posttranscriptional suppression of proto-oncogene c-fms expression by vigilin in breast cancer. *Mol Cell Biol*, **31**, 215-225.
77. Vollbrandt, T., Willkomm, D., Stossberg, H. and Kruse, C. (2004) Vigilin is co-localized with 80S ribosomes and binds to the ribosomal complex through its C-terminal domain. *Int J Biochem Cell Biol*, **36**, 1306-1318.
78. Clausen, T., Southan, C. and Ehrmann, M. (2002) The HtrA family of proteases: implications for protein composition and cell fate. *Mol Cell*, **10**, 443-455.
79. Truebestein, L., Tennstaedt, A., Monig, T., Krojer, T., Canellas, F., Kaiser, M., Clausen, T. and Ehrmann, M. (2011) Substrate-induced remodeling of the active site regulates human HTRA1 activity. *Nat Struct Mol Biol*, **18**, 386-388.
80. Tiaden, A.N. and Richards, P.J. (2013) The emerging roles of HTRA1 in musculoskeletal disease. *Am J Pathol*, **182**, 1482-1488.
81. Xia, J., Wang, F., Wang, L. and Fan, Q. (2013) Elevated serine protease HtrA1 inhibits cell proliferation, reduces invasion, and induces apoptosis in esophageal squamous cell carcinoma by blocking the nuclear factor-kappaB signaling pathway. *Tumour Biol*, **34**, 317-328.
82. Baldi, A., De Luca, A., Morini, M., Battista, T., Felsani, A., Baldi, F., Catricala, C., Amantea, A., Noonan, D.M., Albini, A. *et al.* (2002) The HtrA1 serine protease is down-regulated during human melanoma progression and represses growth of metastatic melanoma cells. *Oncogene*, **21**, 6684-6688.
83. Chien, J., Staub, J., Hu, S.I., Erickson-Johnson, M.R., Couch, F.J., Smith, D.I., Crowl, R.M., Kaufmann, S.H. and Shridhar, V. (2004) A candidate tumor suppressor HtrA1 is downregulated in ovarian cancer. *Oncogene*, **23**, 1636-1644.
84. Chien, J., Aletti, G., Baldi, A., Catalano, V., Muretto, P., Keeney, G.L., Kalli, K.R., Staub, J., Ehrmann, M., Cliby, W.A. *et al.* (2006) Serine protease HtrA1 modulates chemotherapy-induced cytotoxicity. *J Clin Invest*, **116**, 1994-2004.
85. De Luca, A., De Falco, M., De Luca, L., Penta, R., Shridhar, V., Baldi, F., Campioni, M., Paggi, M.G. and Baldi, A. (2004) Pattern of expression of HtrA1 during mouse development. *J Histochem Cytochem*, **52**, 1609-1617.
86. Glentis, A., Gurchenkov, V. and Matic Vignjevic, D. (2014) Assembly, heterogeneity, and breaching of the basement membranes. *Cell Adh Migr*, **8**, 236-245.
87. Lin, C.Y., Anders, J., Johnson, M. and Dickson, R.B. (1999) Purification and characterization of a complex containing matriptase and a Kunitz-type serine protease inhibitor from human milk. *J Biol Chem*, **274**, 18237-18242.
88. Chen, Y.W., Wang, J.K., Chou, F.P., Chen, C.Y., Rorke, E.A., Chen, L.M., Chai, K.X., Eckert, R.L., Johnson, M.D. and Lin, C.Y. (2010) Regulation of the matriptase-prostasin cell surface proteolytic cascade by hepatocyte growth factor activator inhibitor-1 during epidermal differentiation. *J Biol Chem*, **285**, 31755-31762.
89. Kirchhofer, D., Peek, M., Lipari, M.T., Billeci, K., Fan, B. and Moran, P. (2005) Hepsin activates pro-hepatocyte growth factor and is inhibited by hepatocyte growth factor activator inhibitor-1B (HAI-1B) and HAI-2. *FEBS Lett*, **579**, 1945-1950.

90. Herter, S., Piper, D.E., Aaron, W., Gabriele, T., Cutler, G., Cao, P., Bhatt, A.S., Choe, Y., Craik, C.S., Walker, N. *et al.* (2005) Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. *Biochem J*, **390**, 125-136.
91. Kataoka, H., Shimomura, T., Kawaguchi, T., Hamasuna, R., Itoh, H., Kitamura, N., Miyazawa, K. and Koono, M. (2000) Hepatocyte growth factor activator inhibitor type 1 is a specific cell surface binding protein of hepatocyte growth factor activator (HGFA) and regulates HGFA activity in the pericellular microenvironment. *J Biol Chem*, **275**, 40453-40462.
92. Hong, Z., Nowakowski, M., Spronk, C., Petersen, S.V., Andreasen, P.A., Kozminski, W., Mulder, F.A. and Jensen, J.K. (2015) The solution structure of the MANEC-type domain from hepatocyte growth factor activator inhibitor-1 reveals an unexpected PAN/apple domain-type fold. *Biochem J*, **466**, 299-309.
93. Ito, K., Adachi, S., Iwakami, R., Yasuda, H., Muto, Y., Seki, N. and Okano, Y. (2001) N-Terminally extended human ubiquitin-conjugating enzymes (E2s) mediate the ubiquitination of RING-finger proteins, ARA54 and RNF8. *Eur J Biochem*, **268**, 2725-2732.
94. Kikuchi, H., Uchida, C., Hattori, T., Isobe, T., Hiramatsu, Y., Kitagawa, K., Oda, T., Konno, H. and Kitagawa, M. (2007) ARA54 is involved in transcriptional regulation of the cyclin D1 gene in human cancer cells. *Carcinogenesis*, **28**, 1752-1758.
95. Kang, H.Y., Yeh, S., Fujimoto, N. and Chang, C. (1999) Cloning and characterization of human prostate coactivator ARA54, a novel protein that associates with the androgen receptor. *J Biol Chem*, **274**, 8570-8576.
96. Fiorillo, C., Moro, F., Yi, J., Weil, S., Brisca, G., Astrea, G., Severino, M., Romano, A., Battini, R., Rossi, A. *et al.* (2014) Novel dynein DYNC1H1 neck and motor domain mutations link distal spinal muscular atrophy and abnormal cortical development. *Hum Mutat*, **35**, 298-302.
97. Vallee, R.B., McKenney, R.J. and Ori-McKenney, K.M. (2012) Multiple modes of cytoplasmic dynein regulation. *Nat Cell Biol*, **14**, 224-230.
98. Zhang, D.D. and Hannink, M. (2003) Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Mol Cell Biol*, **23**, 8137-8151.
99. Zhang, D.D., Lo, S.C., Sun, Z., Habib, G.M., Lieberman, M.W. and Hannink, M. (2005) Ubiquitination of Keap1, a BTB-Kelch substrate adaptor protein for Cul3, targets Keap1 for degradation by a proteasome-independent pathway. *J Biol Chem*, **280**, 30091-30099.
100. Camolotto, S., Racca, A., Rena, V., Nores, R., Patrito, L.C., Genti-Raimondi, S. and Panzetta-Dutari, G.M. (2010) Expression and transcriptional regulation of individual pregnancy-specific glycoprotein genes in differentiating trophoblast cells. *Placenta*, **31**, 312-319.
101. Tian, Y.M., Mole, D.R., Ratcliffe, P.J. and Gleadle, J.M. (2006) Characterization of different isoforms of the HIF prolyl hydroxylase PHD1 generated by alternative initiation. *Biochem J*, **397**, 179-186.
102. Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A. *et al.* (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*, **107**, 43-54.
103. Runyan, C.E., Schnaper, H.W. and Poncelet, A.C. (2005) The role of internalization in transforming growth factor beta1-induced Smad2 association with Smad anchor for receptor activation (SARA) and Smad2-dependent signaling in human mesangial cells. *J Biol Chem*, **280**, 8300-8308.
104. Liu, N., Li, Z., Pei, D. and Shu, X. (2013) Zfyve9a regulates the proliferation of hepatic cells during zebrafish embryogenesis. *Int J Dev Biol*, **57**, 773-778.

105. Rodrigo-Brenni, M.C., Gutierrez, E. and Hegde, R.S. (2014) Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol Cell*, **55**, 227-237.
106. Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J. and Hegde, R.S. (2010) A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature*, **466**, 1120-1124.
107. Sasaki, T., Marcon, E., McQuire, T., Arai, Y., Moens, P.B. and Okada, H. (2008) Bat3 deficiency accelerates the degradation of Hsp70-2/HspA2 during spermatogenesis. *J Cell Biol*, **182**, 449-458.
108. Tsai, T.C., Lee, Y.L., Hsiao, W.C., Tsao, Y.P. and Chen, S.L. (2005) NRIP, a novel nuclear receptor interaction protein, enhances the transcriptional activity of nuclear receptors. *J Biol Chem*, **280**, 20000-20009.
109. Angers, S., Li, T., Yi, X., MacCoss, M.J., Moon, R.T. and Zheng, N. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature*, **443**, 590-593.
110. Amann, J., Kidd, V.J. and Lahti, J.M. (1997) Characterization of putative human homologues of the yeast chromosome transmission fidelity gene, CHL1. *J Biol Chem*, **272**, 3823-3832.
111. Hirota, Y. and Lahti, J.M. (2000) Characterization of the enzymatic activity of hChIR1, a novel human DNA helicase. *Nucleic Acids Res*, **28**, 917-924.
112. Parish, J.L., Rosa, J., Wang, X., Lahti, J.M., Doxsey, S.J. and Androphy, E.J. (2006) The DNA helicase ChIR1 is required for sister chromatid cohesion in mammalian cells. *J Cell Sci*, **119**, 4857-4865.
113. Parish, J.L., Bean, A.M., Park, R.B. and Androphy, E.J. (2006) ChIR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. *Mol Cell*, **24**, 867-876.
114. Villa, N., Do, A., Hershey, J.W. and Fraser, C.S. (2013) Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J Biol Chem*, **288**, 32932-32940.
115. Sun, F., Palmer, K. and Handel, M.A. (2010) Mutation of Eif4g3, encoding a eukaryotic translation initiation factor, causes male infertility and meiotic arrest of mouse spermatocytes. *Development*, **137**, 1699-1707.
116. Hauf, S., Waizenegger, I.C. and Peters, J.M. (2001) Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. *Science*, **293**, 1320-1323.
117. Obholz, K.L., Akopyan, A., Waymire, K.G. and MacGregor, G.R. (2006) FNDC3A is required for adhesion between spermatids and Sertoli cells. *Dev Biol*, **298**, 498-513.
118. Kishimoto, K., Kato, A., Osada, S., Nishizuka, M. and Imagawa, M. (2010) Fad104, a positive regulator of adipogenesis, negatively regulates osteoblast differentiation. *Biochem Biophys Res Commun*, **397**, 187-191.
119. Maharzi, N., Parietti, V., Nelson, E., Denti, S., Robledo-Sarmiento, M., Setterblad, N., Parcelier, A., Pla, M., Sigaux, F., Gluckman, J.C. *et al.* (2013) Identification of TMEM131L as a novel regulator of thymocyte proliferation in humans. *J Immunol*, **190**, 6187-6197.
120. Zhou, H.J., Wong, C.M., Chen, J.H., Qiang, B.Q., Yuan, J.G. and Jin, D.Y. (2001) Inhibition of LZIP-mediated transcription through direct interaction with a novel host cell factor-like protein. *J Biol Chem*, **276**, 28933-28938.

121. Li, X., Li, R., Lin, X. and Guan, M.X. (2002) Isolation and characterization of the putative nuclear modifier gene MTO1 involved in the pathogenesis of deafness-associated mitochondrial 12 S rRNA A1555G mutation. *J Biol Chem*, **277**, 27256-27264.
122. Tischner, C., Hofer, A., Wulff, V., Stepek, J., Dumitru, I., Becker, L., Haack, T., Kremer, L., Datta, A.N., Sperl, W. *et al.* (2015) MTO1 mediates tissue specificity of OXPHOS defects via tRNA modification and translation optimization, which can be bypassed by dietary intervention. *Hum Mol Genet*, **24**, 2247-2266.
123. Li, R., Li, X., Yan, Q., Qin Mo, J. and Guan, M.X. (2003) Identification and characterization of mouse MTO1 gene related to mitochondrial tRNA modification. *Biochim Biophys Acta*, **1629**, 53-59.
124. Henninger, E.E. and Pursell, Z.F. (2014) DNA polymerase epsilon and its roles in genome stability. *IUBMB Life*, **66**, 339-351.
125. Hubscher, U., Nasheuer, H.P. and Syvaola, J.E. (2000) Eukaryotic DNA polymerases, a growing family. *Trends Biochem Sci*, **25**, 143-147.
126. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. and Cramer, P. (2013) RNA polymerase I structure and transcription regulation. *Nature*, **502**, 650-655.
127. Laferte, A., Favry, E., Sentenac, A., Riva, M., Carles, C. and Chedin, S. (2006) The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev*, **20**, 2030-2040.
128. Grandemange, S., Schaller, S., Yamano, S., Du Manoir, S., Shpakovski, G.V., Mattei, M.G., Keding, C. and Vigneron, M. (2001) A human RNA polymerase II subunit is encoded by a recently generated multigene family. *BMC Mol Biol*, **2**, 14.
129. Feng, L., Wang, J.T., Jin, H., Qian, K. and Geng, J.G. (2011) SH3KBP1-binding protein 1 prevents epidermal growth factor receptor degradation by the interruption of c-Cbl-CIN85 complex. *Cell Biochem Funct*, **29**, 589-596.
130. Zhang, L., Zhou, F., Drabsch, Y., Gao, R., Snaar-Jagalska, B.E., Mickanin, C., Huang, H., Sheppard, K.A., Porter, J.A., Lu, C.X. *et al.* (2012) USP4 is regulated by AKT phosphorylation and directly deubiquitylates TGF-beta type I receptor. *Nat Cell Biol*, **14**, 717-726.
131. Song, E.J., Werner, S.L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J.W., Elledge, S.J., Kirschner, M.W. and Rape, M. (2010) The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. *Genes Dev*, **24**, 1434-1447.
132. Zhang, X., Berger, F.G., Yang, J. and Lu, X. (2011) USP4 inhibits p53 through deubiquitinating and stabilizing ARF-BP1. *EMBO J*, **30**, 2177-2189.
133. Nishi, R., Wijnhoven, P., le Sage, C., Tjeertes, J., Galanty, Y., Forment, J.V., Clague, M.J., Urbe, S. and Jackson, S.P. (2014) Systematic characterization of deubiquitylating enzymes for roles in maintaining genome integrity. *Nat Cell Biol*, **16**, 1016-1026.
134. Wijnhoven, P., Konietzny, R., Blackford, A.N., Travers, J., Kessler, B.M., Nishi, R. and Jackson, S.P. (2015) USP4 Auto-Deubiquitylation Promotes Homologous Recombination. *Mol Cell*, **60**, 362-373.
135. McInerney-Leo, A.M., Schmidts, M., Cortes, C.R., Leo, P.J., Gener, B., Courtney, A.D., Gardiner, B., Harris, J.A., Lu, Y., Marshall, M. *et al.* (2013) Short-rib polydactyly and Jeune syndromes are caused by mutations in WDR60. *Am J Hum Genet*, **93**, 515-523.
136. Yu, T.W., Mochida, G.H., Tischfield, D.J., Sgaier, S.K., Flores-Sarnat, L., Sergi, C.M., Topcu, M., McDonald, M.T., Barry, B.J., Felie, J.M. *et al.* (2010) Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat Genet*, **42**, 1015-1020.

137. Bilguvar, K., Ozturk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatli, B., Yalnizoglu, D., Tuysuz, B., Caglayan, A.O., Gokben, S. *et al.* (2010) Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*, **467**, 207-210.
138. Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P. and Artandi, S.E. (2009) A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science*, **323**, 644-648.
139. McKeegan, K.S., Debieux, C.M., Boulon, S., Bertrand, E. and Watkins, N.J. (2007) A dynamic scaffold of pre-snoRNP factors facilitates human box C/D snoRNP assembly. *Mol Cell Biol*, **27**, 6782-6793.
140. Kamachi, Y. and Kondoh, H. (2013) Sox proteins: regulators of cell fate specification and differentiation. *Development*, **140**, 4129-4144.
141. Lai, T., Jabaudon, D., Molyneaux, B.J., Azim, E., Arlotta, P., Menezes, J.R. and Macklis, J.D. (2008) SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. *Neuron*, **57**, 232-247.
142. Lefebvre, V., Behringer, R.R. and de Crombrughe, B. (2001) L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway. *Osteoarthritis Cartilage*, **9 Suppl A**, S69-S75.
143. Ikeda, T., Kawaguchi, H., Kamekura, S., Ogata, N., Mori, Y., Nakamura, K., Ikegawa, S. and Chung, U.I. (2005) Distinct roles of Sox5, Sox6, and Sox9 in different stages of chondrogenic differentiation. *J Bone Miner Metab*, **23**, 337-340.
144. Uusitalo, H., Hiltunen, A., Ahonen, M., Gao, T.J., Lefebvre, V., Harley, V., Kahari, V.M. and Vuorio, E. (2001) Accelerated up-regulation of L-Sox5, Sox6, and Sox9 by BMP-2 gene transfer during murine fracture healing. *J Bone Miner Res*, **16**, 1837-1845.
145. Nesbitt, A., Bhoj, E.J., McDonald Gibson, K., Yu, Z., Denenberg, E., Sarmady, M., Tischler, T., Cao, K., Dubbs, H., Zackai, E.H. *et al.* (2015) Exome sequencing expands the mechanism of SOX5-associated intellectual disability: A case presentation with review of sox-related disorders. *Am J Med Genet A*, **167**, 2548-2554.
146. Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M.S., Gramolini, A.O., Morris, Q., Hallett, M.T. *et al.* (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell*, **125**, 173-186.
147. Fu, H., Subramanian, R.R. and Masters, S.C. (2000) 14-3-3 proteins: structure, function, and regulation. *Annu Rev Pharmacol Toxicol*, **40**, 617-647.
148. Satoh, J., Kawana, N. and Yamamoto, Y. (2013) Pathway Analysis of ChIP-Seq-Based NRF1 Target Genes Suggests a Logical Hypothesis of their Involvement in the Pathogenesis of Neurodegenerative Diseases. *Gene Regul Syst Bio*, **7**, 139-152.
149. Bugno, M., Daniel, M., Chepelev, N.L. and Willmore, W.G. (2015) Changing gears in Nrf1 research, from mechanisms of regulation to its role in disease and prevention. *Biochim Biophys Acta*, **1849**, 1260-1276.
150. Willier, S., Butt, E., Richter, G.H., Burdach, S. and Grunewald, T.G. (2011) Defining the role of TRIP6 in cell physiology and cancer. *Biol Cell*, **103**, 573-591.
151. Arquint, C., Gabryjonczyk, A.M., Imseng, S., Bohm, R., Sauer, E., Hiller, S., Nigg, E.A. and Maier, T. (2015) STIL binding to Polo-box 3 of PLK4 regulates centriole duplication. *Elife*, **4**, e07888.
152. Habedanck, R., Stierhof, Y.D., Wilkinson, C.J. and Nigg, E.A. (2005) The Polo kinase Plk4 functions in centriole duplication. *Nat Cell Biol*, **7**, 1140-1146.

153. Bettencourt-Dias, M. and Glover, D.M. (2007) Centrosome biogenesis and function: centrosomics brings new understanding. *Nat Rev Mol Cell Biol*, **8**, 451-463.
154. Corominas, R., Yang, X., Lin, G.N., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S.A. *et al.* (2014) Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat Commun*, **5**, 3650.
155. Kedersha, N.L. and Rome, L.H. (1986) Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J Cell Biol*, **103**, 699-709.
156. Kedersha, N.L., Heuser, J.E., Chugani, D.C. and Rome, L.H. (1991) Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. *J Cell Biol*, **112**, 225-235.
157. Lara, P.C., Pruschy, M., Zimmermann, M. and Henriquez-Hernandez, L.A. (2011) MVP and vaults: a role in the radiation response. *Radiat Oncol*, **6**, 148.
158. Kickhoefer, V.A., Vasu, S.K. and Rome, L.H. (1996) Vaults are the answer, what is the question? *Trends Cell Biol*, **6**, 174-178.
159. Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A. and Rovira, C. (2009) The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat Cell Biol*, **11**, 1268-1271.
160. Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K. *et al.* (2000) Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat Genet*, **25**, 302-305.
161. Xiong, H., Wang, D., Chen, L., Choo, Y.S., Ma, H., Tang, C., Xia, K., Jiang, W., Ronai, Z., Zhuang, X. *et al.* (2009) Parkin, PINK1, and DJ-1 form a ubiquitin E3 ligase complex promoting unfolded protein degradation. *J Clin Invest*, **119**, 650-660.
162. Chen, D., Gao, F., Li, B., Wang, H., Xu, Y., Zhu, C. and Wang, G. (2010) Parkin mono-ubiquitinates Bcl-2 and regulates autophagy. *J Biol Chem*, **285**, 38214-38223.
163. da Costa, C.A., Sunyach, C., Giaime, E., West, A., Corti, O., Brice, A., Safe, S., Abou-Sleiman, P.M., Wood, N.W., Takahashi, H. *et al.* (2009) Transcriptional repression of p53 by parkin and impairment by mutations associated with autosomal recessive juvenile Parkinson's disease. *Nat Cell Biol*, **11**, 1370-1375.
164. Kuroda, Y., Sako, W., Goto, S., Sawada, T., Uchida, D., Izumi, Y., Takahashi, T., Kagawa, N., Matsumoto, M., Matsumoto, M. *et al.* (2012) Parkin interacts with Klok1 for mitochondrial import and maintenance of membrane potential. *Hum Mol Genet*, **21**, 991-1003.
165. Emanuele, M.J., Elia, A.E., Xu, Q., Thoma, C.R., Izhar, L., Leng, Y., Guo, A., Chen, Y.N., Rush, J., Hsu, P.W. *et al.* (2011) Global identification of modular cullin-RING ligase substrates. *Cell*, **147**, 459-474.
166. Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J. *et al.* (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell*, **44**, 325-340.
167. Tahmasebi, S., Ghorbani, M., Savage, P., Gocevski, G. and Yang, X.J. (2014) The SUMO conjugating enzyme Ubc9 is required for inducing and maintaining stem cell pluripotency. *Stem Cells*, **32**, 1012-1020.
168. Dasso, M. (2008) Emerging roles of the SUMO pathway in mitosis. *Cell Div*, **3**, 5.

169. Jackson, S.P. and Durocher, D. (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol Cell*, **49**, 795-807.
170. Melchior, F., Schergaut, M. and Pichler, A. (2003) SUMO: ligases, isopeptidases and nuclear pores. *Trends Biochem Sci*, **28**, 612-618.
171. Eifler, K. and Vertegaal, A.C. (2015) Mapping the SUMOylated landscape. *FEBS J*, **282**, 3669-3680.
172. Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P.P. and Dejean, A. (2005) The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev Cell*, **9**, 769-779.
173. Driescher, M., Chaugule, V.K. and Pichler, A. (2013) SUMO rules: regulatory concepts and their implication in neurologic functions. *Neuromolecular Med*, **15**, 639-660.
174. Tirard, M., Hsiao, H.H., Nikolov, M., Urlaub, H., Melchior, F. and Brose, N. (2012) In vivo localization and identification of SUMOylated proteins in the brain of His6-HA-SUMO1 knock-in mice. *Proc Natl Acad Sci U S A*, **109**, 21122-21127.
175. Kho, C., Lee, A., Jeong, D., Oh, J.G., Chaurine, A.H., Kizana, E., Park, W.J. and Hajjar, R.J. (2011) SUMO1-dependent modulation of SERCA2a in heart failure. *Nature*, **477**, 601-605.
176. Bawa-Khalfe, T. and Yeh, E.T. (2010) SUMO Losing Balance: SUMO Proteases Disrupt SUMO Homeostasis to Facilitate Cancer Development and Progression. *Genes Cancer*, **1**, 748-752.
177. Collins, B.M., McCoy, A.J., Kent, H.M., Evans, P.R. and Owen, D.J. (2002) Molecular architecture and functional model of the endocytic AP2 complex. *Cell*, **109**, 523-535.
178. Ma, H., Ng, H.M., Teh, X., Li, H., Lee, Y.H., Chong, Y.M., Loh, Y.H., Collins, J.J., Feng, B., Yang, H. *et al.* (2014) Zfp322a Regulates mouse ES cell pluripotency and enhances reprogramming efficiency. *PLoS Genet*, **10**, e1004038.
179. Li, Y., Wang, Y., Zhang, C., Yuan, W., Wang, J., Zhu, C., Chen, L., Huang, W., Zeng, W., Wu, X. *et al.* (2004) ZNF322, a novel human C2H2 Kruppel-like zinc-finger protein, regulates transcriptional activation in MAPK signaling pathways. *Biochem Biophys Res Commun*, **325**, 1383-1392.
180. Bingol, B., Wang, C.F., Arnott, D., Cheng, D., Peng, J. and Sheng, M. (2010) Autophosphorylated CaMKIIalpha acts as a scaffold to recruit proteasomes to dendritic spines. *Cell*, **140**, 567-578.
181. Liang, R.Y., Chen, L., Ko, B.T., Shen, Y.H., Li, Y.T., Chen, B.R., Lin, K.T., Madura, K. and Chuang, S.M. (2014) Rad23 interaction with the proteasome is regulated by phosphorylation of its ubiquitin-like (Ubl) domain. *J Mol Biol*, **426**, 4049-4060.
182. Voges, D., Zwickl, P. and Baumeister, W. (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem*, **68**, 1015-1068.
183. Lee, D., Ezhkova, E., Li, B., Pattenden, S.G., Tansey, W.P. and Workman, J.L. (2005) The proteasome regulatory particle alters the SAGA coactivator to enhance its interactions with transcriptional activators. *Cell*, **123**, 423-436.
184. Russell, S.J., Reed, S.H., Huang, W., Friedberg, E.C. and Johnston, S.A. (1999) The 19S regulatory complex of the proteasome functions independently of proteolysis in nucleotide excision repair. *Mol Cell*, **3**, 687-695.

185. Chen, Y., Sharp, Z.D. and Lee, W.H. (1997) HEC binds to the seventh regulatory subunit of the 26 S proteasome and modulates the proteolysis of mitotic cyclins. *J Biol Chem*, **272**, 24081-24087.
186. Zhang, M., Wang, R., Wang, Y., Diao, F., Lu, F., Gao, D., Chen, D., Zhai, Z. and Shu, H. (2009) The CXXC finger 5 protein is required for DNA damage-induced p53 activation. *Sci China C Life Sci*, **52**, 528-538.
187. Kim, H.Y., Yoon, J.Y., Yun, J.H., Cho, K.W., Lee, S.H., Rhee, Y.M., Jung, H.S., Lim, H.J., Lee, H., Choi, J. *et al.* (2015) CXXC5 is a negative-feedback regulator of the Wnt/beta-catenin pathway involved in osteoblast differentiation. *Cell Death Differ*, **22**, 912-920.
188. Kim, M.Y., Kim, H.Y., Hong, J., Kim, D., Lee, H., Cheong, E., Lee, Y., Roth, J., Kim, D.G., Min, D.S. *et al.* (2016) CXXC5 plays a role as a transcription activator for myelin genes on oligodendrocyte differentiation. *Glia*, **64**, 350-362.
189. Litman, G.W., Rast, J.P., Shamblott, M.J., Haire, R.N., Hulst, M., Roess, W., Litman, R.T., Hinds-Frey, K.R., Zilch, A. and Amemiya, C.T. (1993) Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol Biol Evol*, **10**, 60-72.
190. Kuhn, U. and Wahle, E. (2004) Structure and function of poly(A) binding proteins. *Biochim Biophys Acta*, **1678**, 67-84.
191. Otero, L.J., Ashe, M.P. and Sachs, A.B. (1999) The yeast poly(A)-binding protein Pab1p stimulates in vitro poly(A)-dependent and cap-dependent translation by distinct mechanisms. *EMBO J*, **18**, 3153-3163.
192. Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J. and Kim, V.N. (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell*, **159**, 1365-1376.
193. Zhang, H., Sheng, C., Yin, Y., Wen, S., Yang, G., Cheng, Z. and Zhu, Q. (2015) PABPC1 interacts with AGO2 and is responsible for the microRNA mediated gene silencing in high grade hepatocellular carcinoma. *Cancer Lett*, **367**, 49-57.
194. Ozturk, S., Guzeloglu-Kayisli, O., Demir, N., Sozen, B., Ilbay, O., Lalioti, M.D. and Seli, E. (2012) Epab and Pabpc1 are differentially expressed during male germ cell development. *Reprod Sci*, **19**, 911-922.
195. Kini, H.K., Kong, J. and Liebhaber, S.A. (2014) Cytoplasmic poly(A) binding protein C4 serves a critical role in erythroid differentiation. *Mol Cell Biol*, **34**, 1300-1309.

Supplementary Table S4. Mutations detected in human FAM46 genes in cancer patients. Only missense mutations are listed. JSD (Jensen-Shannon divergence) was calculated for the respective positions in multiple sequence alignment of the FAM46 protein family. Mutations of highly conserved residues with a JSD score higher than 0.65 (the average JSD for five FAM46 active site residues: glutamic/aspartic acids, glycine and serine) are highlighted in orange.

Gene	Mutation	Cancer subtype	Reference	JSD
FAM46A	p.L23R	liver hepatocellular carcinoma	(1)	-
FAM46A	p.W68C	kidney renal papillary cell carcinoma	(2,3)	0.35
FAM46A	p.W68L	lung adenocarcinoma	(4)	0.35
FAM46A	p.G86R	lung adenocarcinoma	(5)	0.55
FAM46A	p.F90L	stomach adenocarcinoma	(6)	0.53
FAM46A	p.P97T	lung adenocarcinoma	(7,8)	0.45
FAM46A	p.E110K	kidney renal papillary cell carcinoma	(2,3)	0.30
FAM46A	p.K138N	breast invasive carcinoma	(9)	0.64
FAM46A	p.D159N	uterine corpus endometrial carcinoma	(10)	0.48
FAM46A	p.L165Q	ductal adenocarcinoma	(11)	0.47
FAM46A	p.E171K	lung adenocarcinoma	(5)	0.38
FAM46A	p.E248D	uterine carcinosarcoma	(12)	0.25
FAM46A	p.M251T	kidney renal papillary cell carcinoma	(2,3)	0.41
FAM46A	p.T252N	colorectal adenocarcinoma	(13)	0.43
FAM46A	p.E253K	bladder urothelial carcinoma	(14)	0.40
FAM46A	p.I259L	uterine corpus endometrial carcinoma	(1)	0.51
FAM46A	p.V264I	esophageal adenocarcinoma	(15)	0.56
FAM46A	p.D267Y	lung adenocarcinoma	(4)	0.58
FAM46A	p.D273H	bladder urothelial carcinoma	(14)	0.42
FAM46A	p.L275H	lung squamous cell carcinoma	(16)	0.52
FAM46A	p.E286K	small cell lung carcinoma	(17)	0.54
FAM46A	p.R289Q	colorectal adenocarcinoma	(1)	0.69
FAM46A	p.L294H	no data	(7,8)	0.68
FAM46A	p.R302G	skin cutaneous melanoma	(1)	0.50
FAM46A	p.R302M	skin cutaneous melanoma	(18)	0.50
FAM46A	p.A307T	cervical squamous cell carcinoma	(1)	0.34
FAM46A	p.D309G	lung adenocarcinoma	(4)	0.29
FAM46A	p.D309N	breast invasive carcinoma	(19)	0.29
FAM46A	p.Q315E	lung adenocarcinoma	(7,8)	0.52
FAM46A	p.S320F	cervical squamous cell carcinoma. endocervical adenocarcinoma	(2,3)	0.61
FAM46A	p.R321G	skin cutaneous melanoma	(1)	0.63
FAM46A	p.R321M	no data	(7,8)	0.63
FAM46A	p.Q333H	lung adenocarcinoma	(4)	0.34
FAM46A	p.E348G	liver hepatocellular carcinoma	(1)	0.38
FAM46A	p.R350H	stomach adenocarcinoma	(6)	0.31
FAM46A	p.R387Q	prostate carcinoma	(7,8)	0.45

FAM46A	p.D391N	bladder urothelial carcinoma	(14)	0.32
FAM46A	p.V398A	multiple myeloma	(20)	0.27
FAM46A	p.Y420C	uterine corpus endometrial carcinoma	(10)	0.37
FAM46A	p.Y420H	serous cystadenocarcinoma	(1)	0.37
FAM46B	p.T20K	adrenocortical carcinoma, liver hepatocellular carcinoma	(1-3)	-
FAM46B	p.P45L	skin cutaneous melanoma	(2,3)	0.24
FAM46B	p.V57G	breast invasive carcinoma	(19)	0.43
FAM46B	p.S65N	lung adenocarcinoma	(5)	0.40
FAM46B	p.G72R	colorectal adenocarcinoma	(21)	0.55
FAM46B	p.R73H	colorectal adenocarcinoma	(1)	0.43
FAM46B	p.P77L	skin cutaneous melanoma	(2,3)	0.59
FAM46B	p.P83L	colorectal adenocarcinoma	(21)	0.45
FAM46B	p.R91C	uterine carcinosarcoma	(12)	0.40
FAM46B	p.R105W	colorectal adenocarcinoma	(22)	0.63
FAM46B	p.E118D	liver hepatocellular carcinoma	(23)	0.33
FAM46B	p.R135H	uterine corpus endometrial carcinoma	(10)	0.38
FAM46B	p.V146A	pancreatic adenocarcinoma	(2,3)	0.45
FAM46B	p.E171Q	colorectal adenocarcinoma	(1)	0.45
FAM46B	p.C181R	uterine corpus endometrial carcinoma	(10)	0.44
FAM46B	p.D183N	colorectal adenocarcinoma	(21)	0.43
FAM46B	p.D206H	breast invasive carcinoma	(9)	0.45
FAM46B	p.F212Y	kidney renal papillary cell carcinoma	(2,3)	0.67
FAM46B	p.Q220H	lung adenocarcinoma	(2,3)	0.69
FAM46B	p.D224N	uterine corpus endometrial carcinoma	(10)	0.56
FAM46B	p.C232F	lung adenocarcinoma	(5)	0.26
FAM46B	p.E248K	bladder urothelial carcinoma, lung adenocarcinoma	(5,14)	0.58
FAM46B	p.Y251C	uterine corpus endometrial carcinoma	(10)	0.61
FAM46B	p.R262Q	colorectal adenocarcinoma, stomach adenocarcinoma	(1,22)	0.46
FAM46B	p.R264H	colorectal adenocarcinoma	(22)	0.55
FAM46B	p.R269H	stomach adenocarcinoma, uterine corpus endometrial carcinoma	(1,6)	0.47
FAM46B	p.G277C	bladder urothelial carcinoma	(24)	0.64
FAM46B	p.R291W	bladder urothelial carcinoma, cervical squamous cell carcinoma, endocervical adenocarcinoma	(2,3,14)	0.44
FAM46B	p.R293Q	uterine corpus endometrial carcinoma	(10)	0.34
FAM46B	p.R299C	colorectal adenocarcinoma, prostate carcinoma	(7,8,22)	0.38
FAM46B	p.R299H	multiple myeloma	(20)	0.38
FAM46B	p.R303C	esophageal adenocarcinoma	(15)	0.54
FAM46B	p.R325C	uterine carcinosarcoma	(12)	0.43
FAM46B	p.G332S	multiple myeloma	(20)	0.41
FAM46B	p.E352G	esophageal adenocarcinoma	(15)	0.36
FAM46B	p.T365M	colorectal adenocarcinoma	(22)	0.50
FAM46B	p.D396N	colorectal adenocarcinoma	(1)	-

FAM46B	p.P419S	skin cutaneous melanoma	(18)	0.31
FAM46C	p.S5G	skin cutaneous melanoma	(2,3)	-
FAM46C	p.D10N	skin cutaneous melanoma	(2,3)	0.24
FAM46C	p.V16M	uterine carcinosarcoma	(12)	0.42
FAM46C	p.R24Q	acute lymphoblastic leukemia	(1)	0.37
FAM46C	p.P34H	uterine corpus endometrial carcinoma	(10)	0.39
FAM46C	p.G37R	skin cutaneous melanoma	(7,8)	0.55
FAM46C	p.P42S	skin cutaneous melanoma	(2,3)	0.59
FAM46C	p.T43A	multiple myeloma	(20)	0.55
FAM46C	p.V52I	glioblastoma multiforme, uterine corpus endometrial carcinoma	(10,25)	0.45
FAM46C	p.R56H	skin cutaneous melanoma	(1)	0.40
FAM46C	p.V69I	uterine corpus endometrial carcinoma	(10)	0.59
FAM46C	p.R70W	stomach adenocarcinoma	(6)	0.63
FAM46C	p.D90G	multiple myeloma	(20)	0.72
FAM46C	p.D90H	multiple myeloma	(20)	0.72
FAM46C	p.L93V	multiple myeloma	(26)	0.53
FAM46C	p.Q106L	serous cystadenocarcinoma	(27)	0.43
FAM46C	p.L116F	cutaneous squamous cell carcinoma	(28)	0.47
FAM46C	p.F119L	multiple myeloma	(29)	0.42
FAM46C	p.P121A	kidney renal papillary cell carcinoma	(2,3)	0.50
FAM46C	p.P121S	ductal adenocarcinoma, stomach adenocarcinoma	(1,6)	0.50
FAM46C	p.L134V	multiple myeloma	(26)	0.36
FAM46C	p.Y138C	multiple myeloma	(30)	0.63
FAM46C	p.K141T	multiple myeloma	(26)	0.60
FAM46C	p.K144E	multiple myeloma	(30)	0.59
FAM46C	p.V145D	multiple myeloma	(20)	0.55
FAM46C	p.V145G	multiple myeloma	(30)	0.55
FAM46C	p.D148N	liver hepatocellular carcinoma	(23)	0.43
FAM46C	p.I155L	multiple myeloma	(29)	0.67
FAM46C	p.S156F	primary plasma cell leukemia, skin cutaneous melanoma	(2,3,26)	0.65
FAM46C	p.N164T	secondary plasma cell leukemia	(26)	0.55
FAM46C	p.V165G	multiple myeloma	(26)	0.56
FAM46C	p.R175P	lung adenocarcinoma	(2,3)	0.63
FAM46C	p.E178A	multiple myeloma	(31)	0.61
FAM46C	p.F179C	multiple myeloma	(29)	0.67
FAM46C	p.D182Y	multiple myeloma	(31)	0.69
FAM46C	p.S183C	multiple myeloma	(20)	0.70
FAM46C	p.F184L	multiple myeloma	(20)	0.67
FAM46C	p.S190Y	multiple myeloma	(20)	0.49
FAM46C	p.P201F	skin cutaneous melanoma	(2,3)	0.24
FAM46C	p.S203C	multiple myeloma	(20)	0.43

FAM46C	p.H207N	lung squamous cell carcinoma	(16)	0.48
FAM46C	p.I211T	stomach adenocarcinoma	(2,3)	0.42
FAM46C	p.Y216C	multiple myeloma	(32)	0.61
FAM46C	p.D218N	skin cutaneous melanoma	(2,3)	0.58
FAM46C	p.A232T	colorectal adenocarcinoma	(21)	0.45
FAM46C	p.G241C	multiple myeloma	(20)	0.64
FAM46C	p.L250R	multiple myeloma	(20)	0.62
FAM46C	p.R253Q	colorectal adenocarcinoma, glioblastoma multiforme	(1,25)	0.50
FAM46C	p.R253W	glioblastoma multiforme, lower grade glioma	(7,8)	0.50
FAM46C	p.D259H	bladder urothelial carcinoma	(14)	0.39
FAM46C	p.E261K	skin cutaneous melanoma	(33)	0.29
FAM46C	p.E262D	uterine corpus endometrial carcinoma	(10)	0.29
FAM46C	p.I263M	head and neck squamous cell carcinoma	(34)	0.34
FAM46C	p.L266V	stomach adenocarcinoma	(7,8)	0.50
FAM46C	p.R268C	colorectal adenocarcinoma	(22)	0.54
FAM46C	p.Y269C	multiple myeloma	(26)	0.59
FAM46C	p.M270T	liver hepatocellular carcinoma	(1)	0.67
FAM46C	p.M270V	multiple myeloma	(31)	0.67
FAM46C	p.F274L	uterine corpus endometrial carcinoma	(7,8)	0.70
FAM46C	p.F278L	multiple myeloma	(26)	0.59
FAM46C	p.L288S	colorectal adenocarcinoma	(13)	0.55
FAM46C	p.Y291C	multiple myeloma	(29)	0.62
FAM46C	p.A297T	breast invasive carcinoma, multiple myeloma	(9,26)	0.41
FAM46C	p.Y306C	lung adenocarcinoma	(5)	0.57
FAM46C	p.R311C	breast invasive carcinoma, skin cutaneous melanoma	(2,3,9)	0.48
FAM46C	p.R311H	stomach adenocarcinoma	(7,8)	0.48
FAM46C	p.R312M	skin cutaneous melanoma	(2,3)	0.46
FAM46C	p.V314G	multiple myeloma	(31)	0.48
FAM46C	p.T318I	skin cutaneous melanoma	(18)	0.54
FAM46C	p.V319L	head and neck squamous cell carcinoma	(34)	0.52
FAM46C	p.G323E	skin cutaneous melanoma	(2,3)	0.49
FAM46C	p.R326C	multiple myeloma	(26)	0.56
FAM46C	p.S334F	skin cutaneous melanoma	(2,3)	0.32
FAM46C	p.A337D	bladder urothelial carcinoma	(35)	0.40
FAM46C	p.P361S	skin cutaneous melanoma	(2,3)	-
FAM46C	p.V363I	glioblastoma multiforme	(25)	-
FAM46D	p.S2Y	uterine corpus endometrial carcinoma	(10)	0.24
FAM46D	p.R5K	liver hepatocellular carcinoma	(2,3)	0.40
FAM46D	p.D12N	skin cutaneous melanoma	(2,3)	0.43
FAM46D	p.D18V	lung adenocarcinoma	(7,8)	0.46
FAM46D	p.V20A	renal non-clear cell carcinoma	(36)	0.48

FAM46D	p.H28N	lung adenocarcinoma	(5)	0.55
FAM46D	p.G29E	skin cutaneous melanoma	(18)	0.55
FAM46D	p.N32I	liver hepatocellular carcinoma	(2,3)	0.41
FAM46D	p.D42Y	breast invasive carcinoma	(19)	0.40
FAM46D	p.V46F	liver hepatocellular carcinoma	(7,8)	0.35
FAM46D	p.G53R	lung adenocarcinoma	(7,8)	0.30
FAM46D	p.G55E	skin cutaneous melanoma	(33)	0.43
FAM46D	p.A61P	head and neck squamous cell carcinoma	(34)	0.59
FAM46D	p.G65D	lung adenocarcinoma	(4)	0.65
FAM46D	p.S69G	esophageal adenocarcinoma, stomach adenocarcinoma	(6,15)	0.57
FAM46D	p.S69R	stomach adenocarcinoma	(6)	0.57
FAM46D	p.D82N	multiple myeloma, skin cutaneous melanoma	(2,3,20)	0.72
FAM46D	p.L83P	ductal adenocarcinoma	(1)	0.66
FAM46D	p.I86N	skin cutaneous melanoma	(18)	0.49
FAM46D	p.G88S	skin cutaneous melanoma	(33)	0.44
FAM46D	p.L91H	stomach adenocarcinoma	(6)	0.38
FAM46D	p.E95K	uterine corpus endometrial carcinoma	(10)	0.21
FAM46D	p.E96K	skin cutaneous melanoma	(18)	0.22
FAM46D	p.V99I	uterine corpus endometrial carcinoma	(10)	0.39
FAM46D	p.D102H	esophageal adenocarcinoma	(15)	0.48
FAM46D	p.D102Y	colorectal adenocarcinoma, uterine carcinosarcoma	(1,12)	0.48
FAM46D	p.P113S	liver hepatocellular carcinoma	(2,3)	0.50
FAM46D	p.D115N	colorectal adenocarcinoma, uterine corpus endometrial carcinoma	(1,10)	0.42
FAM46D	p.D124E	colorectal adenocarcinoma	(1)	0.32
FAM46D	p.I125F	kidney renal clear cell carcinoma	(37)	0.36
FAM46D	p.D128G	lung adenocarcinoma	(4)	0.45
FAM46D	p.D128Y	rectum adenocarcinoma	(1)	0.45
FAM46D	p.V137F	breast invasive carcinoma	(19)	0.55
FAM46D	p.C138Y	liver hepatocellular carcinoma	(2,3)	0.44
FAM46D	p.W144L	head and neck squamous cell carcinoma	(34)	0.69
FAM46D	p.K155N	lung adenocarcinoma	(4)	0.54
FAM46D	p.K160Q	colorectal adenocarcinoma	(21)	0.69
FAM46D	p.K160T	colorectal adenocarcinoma	(21)	0.69
FAM46D	p.S172I	liver hepatocellular carcinoma, lung adenocarcinoma	(1,5)	0.64
FAM46D	p.F186L	colorectal adenocarcinoma	(21)	0.40
FAM46D	p.A192V	stomach adenocarcinoma	(38)	0.24
FAM46D	p.K196I	skin cutaneous melanoma	(2,3)	0.40
FAM46D	p.P200L	skin cutaneous melanoma	(2,3)	0.61
FAM46D	p.E213K	skin cutaneous melanoma	(33)	0.46
FAM46D	p.L242F	small cell lung carcinoma	(39)	0.62
FAM46D	p.K256R	lung adenocarcinoma	(5)	0.38

FAM46D	p.F266L	uterine corpus endometrial carcinoma	(10)	0.70
FAM46D	p.I273M	skin cutaneous melanoma	(18)	0.46
FAM46D	p.K278N	rectum adenocarcinoma	(1)	0.35
FAM46D	p.N286S	uterine corpus endometrial carcinoma	(10)	0.54
FAM46D	p.H287Y	skin cutaneous melanoma	(2,3)	0.63
FAM46D	p.Y296N	lung adenocarcinoma	(4)	0.44
FAM46D	p.L302M	lung adenocarcinoma	(4)	0.57
FAM46D	p.E308K	pheochromocytoma, paraganglioma	(2,3)	0.36
FAM46D	p.R318K	colorectal adenocarcinoma	(22)	0.56
FAM46D	p.L322F	gingivobuccal	(1)	0.54
FAM46D	p.L322I	uterine corpus endometrial carcinoma	(10)	0.54
FAM46D	p.A329V	gallbladder carcinoma	(40)	0.40
FAM46D	p.P340Q	breast invasive carcinoma	(19)	0.32
FAM46D	p.F348L	colorectal adenocarcinoma	(21)	0.47
FAM46D	p.P353S	skin cutaneous melanoma	(18)	-
FAM46D	p.R360W	glioblastoma multiforme	(2,3)	-
FAM46D	p.P362H	lung adenocarcinoma	(5)	-
FAM46D	p.I366M	stomach adenocarcinoma	(41)	0.35
FAM46D	p.P369Q	small cell lung carcinoma	(18)	-
FAM46D	p.P371S	lung adenocarcinoma	(4)	-
FAM46D	p.V372I	colorectal adenocarcinoma, colorectal adenocarcinoma non-western	(1)	-
FAM46D	p.P379S	liver hepatocellular carcinoma, lung large cell carcinoma	(1,13)	0.31
FAM46D	p.R383C	lung adenocarcinoma	(5)	0.29
FAM46D	p.G387D	skin cutaneous melanoma	(2,3)	-
FAM46D	p.M388I	skin cutaneous melanoma	(18)	-
FAM46D	p.S389I	uterine corpus endometrial carcinoma	(10)	-

References

1. Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabe, R.R., Bhan, M.K., Calvo, F., Eerola, I., Gerhard, D.S. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993-998.
2. Seedat, Y.K. (1990) Hypertension and vascular disease in India and migrant Indian populations in the world. *J Hum Hypertens*, **4**, 421-424.
3. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, **6**, pl1.
4. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A. *et al.* (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*, **150**, 1107-1120.
5. Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**, 543-550.
6. Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, **513**, 202-209.

7. Anokina, I.P., Kogan, B.M., Nickel, B., Schulz, R. and Kemper, A. (1989) Basal activities of adenylate and guanylate cyclase in lymphocytes and platelets of alcoholics. *Biomed Biochim Acta*, **48**, 593-596.
8. Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A. and Lopez-Bigas, N. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods*, **10**, 1081-1082.
9. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61-70.
10. Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., Benz, C.C. *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**, 67-73.
11. Scarlett, C.J., Salisbury, E.L., Biankin, A.V. and Kench, J. (2011) Precursor lesions in pancreatic cancer: morphological and molecular pathology. *Pathology*, **43**, 183-200.
12. Jones, S., Stransky, N., McCord, C.L., Cerami, E., Lagowski, J., Kelly, D., Angiuoli, S.V., Sausen, M., Kann, L., Shukla, M. *et al.* (2014) Genomic analyses of gynaecologic carcinosarcomas reveal frequent mutations in chromatin remodelling genes. *Nat Commun*, **5**, 5006.
13. Reinhold, W.C., Sunshine, M., Liu, H., Varma, S., Kohn, K.W., Morris, J., Doroshow, J. and Pommier, Y. (2012) CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res*, **72**, 3499-3511.
14. Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, **507**, 315-322.
15. Dulak, A.M., Stojanov, P., Peng, S., Lawrence, M.S., Fox, C., Stewart, C., Bandla, S., Imamura, Y., Schumacher, S.E., Shefler, E. *et al.* (2013) Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet*, **45**, 478-486.
16. Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, **489**, 519-525.
17. George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretic, L., Kong, G., Leenders, F., Lu, X., Fernandez-Cuesta, L., Bosco, G. *et al.* (2015) Comprehensive genomic profiles of small cell lung cancer. *Nature*, **524**, 47-53.
18. Hodis, E., Watson, I.R., Kryukov, G.V., Arold, S.T., Imielinski, M., Theurillat, J.P., Nickerson, E., Auclair, D., Li, L., Place, C. *et al.* (2012) A landscape of driver mutations in melanoma. *Cell*, **150**, 251-263.
19. Ciriello, G., Gatza, M.L., Beck, A.H., Wilkerson, M.D., Rhie, S.K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C. *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*, **163**, 506-519.
20. Lohr, J.G., Stojanov, P., Carter, S.L., Cruz-Gordillo, P., Lawrence, M.S., Auclair, D., Sougnez, C., Knoechel, B., Gould, J., Saksena, G. *et al.* (2014) Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*, **25**, 91-101.
21. Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**, 330-337.
22. Seshagiri, S., Stawiski, E.W., Durinck, S., Modrusan, Z., Storm, E.E., Conboy, C.B., Chaudhuri, S., Guan, Y., Janakiraman, V., Jaiswal, B.S. *et al.* (2012) Recurrent R-spondin fusions in colon cancer. *Nature*, **488**, 660-664.
23. Ahn, S.M., Jang, S.J., Shim, J.H., Kim, D., Hong, S.M., Sung, C.O., Baek, D., Haq, F., Ansari, A.A., Lee, S.Y. *et al.* (2014) Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology*, **60**, 1972-1982.
24. Guo, G., Sun, X., Chen, C., Wu, S., Huang, P., Li, Z., Dean, M., Huang, Y., Jia, W., Zhou, Q. *et al.* (2013) Whole-genome and whole-exome sequencing of bladder cancer identifies frequent

- alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet*, **45**, 1459-1463.
25. Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H. *et al.* (2013) The somatic genomic landscape of glioblastoma. *Cell*, **155**, 462-477.
 26. Barbieri, M., Manzoni, M., Fabris, S., Ciceri, G., Todoerti, K., Simeon, V., Musto, P., Cortelezzi, A., Baldini, L., Neri, A. *et al.* (2015) Compendium of FAM46C gene mutations in plasma cell dyscrasias. *Br J Haematol*.
 27. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609-615.
 28. Li, Y.Y., Hanna, G.J., Laga, A.C., Haddad, R.I., Lorch, J.H. and Hammerman, P.S. (2015) Genomic analysis of metastatic cutaneous squamous cell carcinoma. *Clin Cancer Res*, **21**, 1447-1456.
 29. Boyd, K.D., Ross, F.M., Walker, B.A., Wardell, C.P., Tapper, W.J., Chiecchio, L., Dagrada, G., Konn, Z.J., Gregory, W.M., Jackson, G.H. *et al.* (2011) Mapping of chromosome 1p deletions in myeloma identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as being genes in regions associated with adverse survival. *Clin Cancer Res*, **17**, 7776-7784.
 30. Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R. *et al.* (2014) Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun*, **5**, 2997.
 31. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467-472.
 32. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415-421.
 33. Krauthammer, M., Kong, Y., Ha, B.H., Evans, P., Bacchiocchi, A., McCusker, J.P., Cheng, E., Davis, M.J., Goh, G., Choi, M. *et al.* (2012) Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat Genet*, **44**, 1006-1014.
 34. Cancer Genome Atlas Network (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, **517**, 576-582.
 35. Kim, P.H., Cha, E.K., Sfakianos, J.P., Iyer, G., Zabor, E.C., Scott, S.N., Ostrovnaya, I., Ramirez, R., Sun, A., Shah, R. *et al.* (2015) Genomic predictors of survival in patients with high-grade urothelial carcinoma of the bladder. *Eur Urol*, **67**, 198-201.
 36. Durinck, S., Stawiski, E.W., Pavia-Jimenez, A., Modrusan, Z., Kapur, P., Jaiswal, B.S., Zhang, N., Toffessi-Tcheuyap, V., Nguyen, T.T., Pahuja, K.B. *et al.* (2015) Spectrum of diverse genomic alterations define non-clear cell renal carcinoma subtypes. *Nat Genet*, **47**, 13-21.
 37. Cancer Genome Atlas Research Network (2013) Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, **499**, 43-49.
 38. Kakiuchi, M., Nishizawa, T., Ueda, H., Gotoh, K., Tanaka, A., Hayashi, A., Yamamoto, S., Tatsuno, K., Katoh, H., Watanabe, Y. *et al.* (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet*, **46**, 583-587.
 39. Peifer, M., Fernandez-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T. *et al.* (2012) Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet*, **44**, 1104-1110.
 40. Li, M., Zhang, Z., Li, X., Ye, J., Wu, X., Tan, Z., Liu, C., Shen, B., Wang, X.A., Wu, W. *et al.* (2014) Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. *Nat Genet*, **46**, 872-876.

41. Wang, K., Kan, J., Yuen, S.T., Shi, S.T., Chu, K.M., Law, S., Chan, T.L., Kan, Z., Chan, A.S., Tsui, W.Y. *et al.* (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet*, **43**, 1219-1223.

Warszawa, 10 sierpnia 2019

.....
Miejscowość i data

Krzysztof Kuchta
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K, Rychlewski L, Ginalski K. FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.

mój udział polegał na:

- analizie funkcjonalnej, sekwencyjnej oraz strukturalnej białek posiadających strukturę nukleotydotransferaz (NTaz),
- modelowaniu struktury przestrzennej białka ludzkiego FAM46C,
- analizie, opisanych w literaturze, makromolekuł oddziaływujących z białkami należącymi do rodziny FAM46 (tabela S3),
- analizie struktury domenowej NTaz (rysunek S3) oraz białek należący do rodziny FAM46 (rysunek S4),
- przygotowaniu rysunku 2, rysunku 3 rysunku 4, rysunku 5, rysunku S3, rysunku S4, rysunku S5,
- przygotowaniu tabeli S3,
- opisanu w manuskrypcie przeprowadzonych przeze mnie badań.

.....
Krzysztof Kuchta
Podpis

Warszawa, 18 lipca 2017 roku
Miejscowość i data

Anna Muszewska
Instytut Biochemii i Biofizyki PAN
ul. Pawińskiego 5a
02-106 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS,
Pawlowski K, Rychlewski L, Ginalski K. FAM46 proteins are novel
eukaryotic non-canonical poly(A) polymerases. 2016 Nucleic Acids
Research 44(8):3534-48.*

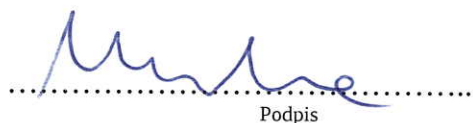
mój udział polegał na:

Analizie sekwencji białkowych i przeszukiwaniu baz danych o białkach.

Przeprowadziłam analizę filogenetyczną rodziny FAM46.

Współtworzyłam manuskrypt.

Uczestniczyłam w interpretacji uzyskanych wyników.



Podpis

Warszawa 19.07.2017
.....
Miejscowość i data

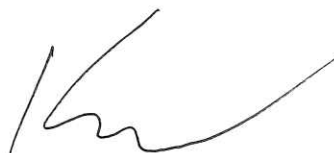
Łukasz Knizewski
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K,
Rychlewski L, Ginalski K. FAM46 proteins are novel eukaryotic non-canonical
poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.*

mój udział polegał na: analizach sekwencyjno-strukturalnych białek i analizie znanych mutacji zaobserwowanych w komórkach rakowych w kontekstach umiejscowienia w MSA rodziny FAM46 i struktur 3D białek FAM46A/B/C/D.


.....
Podpis

Warszawa 17.07.2017
.....
Miejscowość i data

Kamil Steczkiewicz
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K, Rychlewski L, Ginalski K. FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.

mój udział polegał na analizie sekwencyjno-strukturalnej białek należących do nadrodziny NTaz, na modelowaniu struktury przestrzennej białek FAM46, oraz interpretacji wyników.

Kamil Steczkiewicz
.....
Podpis

7 sierpień 2017

.....
Miejscowość i data

Krzysztof Kuchta
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawłowski K,
Rychlewski L, Ginalski K. FAM46 proteins are novel eukaryotic non-canonical
poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.*

udział **dr hab. n. med. Lucjana Wyrwicza (Laboratorium Bioinformatyki i
Biostatystyki, Centrum Onkologii – Instytut Im. Marii Skłodowskiej-Curie)**
polegał na zrecenzowaniu oraz redagowaniu manuskryptu.

.....
Krzysztof Kuchta
Podpis

Warszawa, 17.07.2017
Miejscowość i data

Krzysztof Pawłowski
Katedra Doświadczalnictwa i Bioinformatyki
Szkoła Główna Gospodarstwa Wiejskiego,
Nowoursynowska 166,
02-787 Warszawa

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawłowski K,
Rychlewski L, Ginalski K. FAM46 proteins are novel eukaryotic non-canonical
poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.*

mój udział polegał na:

zapropozowaniu tematu pracy, udziale w planowaniu analiz, udziale w pisaniu
manuskryptu.


.....
Podpis

18 lipca 2017

.....
Miejscowość i data

Leszek Rychlewski
Instytut BioInfoBank,
Św. Marcin 80/82 lok. 355
61-809 Poznań

OŚWIADCZENIE

Oświadczam, że w pracy:

*Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K,
Rychlewski L, Ginalska K. FAM46 proteins are novel eukaryotic non-canonical
poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.*

mój udział polegał na zrecenzowaniu oraz redagowaniu manuskryptu.

.....


Podpis

7 sierpień 2017

.....
Miejscowość i data

Krzysztof Ginalski
Laboratorium Bioinformatyki i Biologii Systemów,
Centrum Nowych Technologii,
Uniwersytet Warszawski,
Żwirki i Wigury 93,
02-089 Warszawa

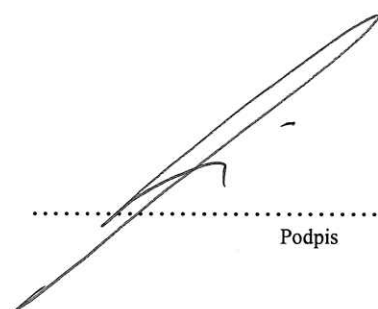
OŚWIADCZENIE

Oświadczam, że w pracy:

Kuchta K, Muszewska A, Knizewski L, Steczkiewicz K, Wyrwicz LS, Pawlowski K, Rychlewski L, Ginalski K. FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. 2016 Nucleic Acids Research 44(8):3534-48.

mój udział polegał na:

- kierowaniu projektem naukowym obejmującym badania opisane w tej pracy,
- zaplanowaniu badań bioinformatycznych,
- pomocy w pisaniu manuskryptu.


.....
Podpis