Structural and functional characterisation of the protein-RNA complexes in the epithelial-mesenchymal transition

K 125340 PROJECT – FINAL REPORT

Project start: 2017. 09. 01.
Closing date: 2022. 08. 31.

Summary of the performed work

In our proposal we aimed at understanding the molecular background of the interaction of long non-coding RNA HOTAIR with its protein partners and the involvement of the HOTAIR-Ezh2 pathway in the epithelial mesenchymal transition induced by the absence of Tks4. We planned the work in three directions: i) determination of the binding affinities and specificity of the RNA recognition of the Ezh2 loop region, ii) structural characterization of the HOTAIR-Ezh2 loop complex and iii) detailed study of the involvement of HOTAIR and Ezh2 in the Tks4 KO-induced epithelial-mesenchymal transition (EMT).

The execution of the experimental work was somewhat hindered by the COVID-19 pandemic, but we managed to achieve significant results in all three directions.

1. *Determination of the binding affinities and specificity of the RNA recognition of the Ezh2 loop region*

In order to understand better the RNA recognition specificity of the disordered Ezh2 loop region, we expressed four different HOTAIR constructs:

$HOTAIR_{50}$: sequence between 20-60 nt of HOTAIR
$HOTAIR_{140}$: sequence between 1-140 nt of HOTAIR
$HOTAIR_{300}$: sequence between 140-440 nt of HOTAIR (originally annotated as the Ezh2-interacting region)
$HOTAIR_{440}$: sequence between 1-440 nt of HOTAIR

We also tested a 50 nt-long random sequence RNA (R50) and a different lncRNA partner of Ezh2: MEG3.

We performed electrophoretic mobility shift assay (EMSA) and microscale thermophoresis (MST) measurements to characterize the binding of the Ezh2 loop to the different RNA species.

While EMSA is a quantitative measurement, supplying information only on the existence of the interaction, MST is a qualitative method that can be used to determine the exact binding affinities of the different ligands.

EMSA results confirmed that all of the tested RNA species were capable of binding to the Ezh2 loop, with only the R50 showing lower binding capacity.

With the MST measurements we could determine that the $HOTAIR_{140}$ has the highest affinity to the Ezh2 loop (0.3 μM), which is twice as strong as the affinity of $HOTAIR_{300}$ (0.69 μM), that was originally determined as the Ezh2-binding segment of HOTAIR. MEG3 bound with similar K_d as $HOTAIR_{300}$ (0.65 μM), while $HOTAIR_{440}$ only showed a K_d of 0,75 μM. Interestingly, from the two short RNAs, the R50 was bound stronger than $HOTAIR_{50}$ (1.79 μM vs 3.55 μM, respectively).

Since earlier observations published in the literature suggested that a phosphorylation event on Tyr345 residue in the Ezh2 loop increased the binding affinity of Ezh2 to RNAs, we tested the binding of a phosphomimetic mutant (Ezh2-TD) to the same RNAs.

Contrary to the published results and the expectations, the phosphomimetic mutation did not result in an overall affinity increase, but it had varying effects on the interaction with the specific RNAs. The affinity to the short RNAs was increased for both the R50 and the $HOTAIR_{50}$ (1.26 and 2.75 μM, respectively), while it was decreased for MEG3 (0.91 μM). The binding strength to the longer HOTAIR constructs did not change markedly ($HOTAIR_{140}$: 0.29 μM, $HOTAIR_{300}$: 0.81 μM, $HOTAIR_{400}$: 0.65 μM). This indicates that the phosphorylation induces changes in the RNA recognition that are affected by the specific sequence and/or structure of the RNA.

In all our experiments, we used thymosin beta 4 (TB4), a disordered protein as negative control. TB4 did not show measurable binding to any RNAs that we tested.

2. Structural characterization of the HOTAIR-Ezh2 loop complex

To better understand the structural background of the binding, we applied circular dichroism (CD) and nuclear magnetic resonance (NMR) measurements combined with molecular modeling.

Based on literature data, we hypothesized that the RNA recognition might be directed by the G-quadruplex structures found in the RNA molecules.

In silico prediction methods indicated the presence of 5 G-quadruplex structures in the sequence of $HOTAIR_{440}$ (1 in $HOTAIR_{50}$, 2 in $HOTAIR_{140}$, 4 in $HOTAIR_{300}$), 31 in MEG3 and 1 in R50.

Molecular modeling simulations also confirmed the possible formation of the G-quadruplexes in $HOTAIR_{440}$.

Nevertheless, no stable G-quadruplex structures were observed in the RNAs using Thioflavin T assay, CD or NMR measurements. The fact that we could still measure relatively strong binding indicates that the studied region of Ezh2 does not specifically recognize quadruplex structures. This is in line with literature data, where the deletion of the Ezh2 loop did not significantly reduce binding of Ezh2 to model quadruplex RNAs.

Regarding the structure of the Ezh2 loop, both NMR and CD measurements confirmed that it is highly disordered and that the phosphomimetic mutation does not induce measurable changes in the structure of the protein. With the help of NMR measurements, we were able to identify the region within the Ezh2 loop that participates in RNA binding. This short segment between Thr367–Ser375 is a unique RNA binding motif, as it does not resemble the canonical RNA recognition elements found in other disordered proteins. The interaction of the Ezh2 loop and HOTAIR was determined to belong to the fuzzy complexes, since no stable structure was

formed in the protein upon binding. This was also the case in the TD mutant loop region, indicating that phosphorylation does not alter the mode of the interaction on the protein's side.

Due to technical limitations of the NMR, the detailed structural study of the different RNAs was not possible with this method. Therefore, we approached the question of the complex formation on the RNA side using CD measurements.

With this method we were able to follow structural changes of the RNAs that occurred upon binding to the Ezh2 loop and its phosphomimetic mutant version. In line with the NMR results, CD also showed that the protein is disordered and remains disordered even in the complex, but the structure of the RNA partner was more affected. Detailed analysis of the CD spectra revealed that the RNAs contained a significant amount of double stranded segments, which were markedly reduced when binding to the Ezh2 loop. This decrease in the double stranded RNA regions was more pronounced upon binding to the TD mutant version of the loop, revealing the importance of the phosphorylation. As the actual binding localized to a shorter segment of the RNAs, the structural changes in the CD spectra were less pronounced, but still detectable.

Summarizing our results on the *in vitro* binding and structural studies we formed a model of the role of the loop region of Ezh2 in the RNA recognition of the protein. Since the loop is highly disordered and is localized on the surface of Ezh2, it is ideally placed to serve as a first contact point for RNAs. Given its lack of specificity towards G-quadruplexes, the binding is rather nonspecific, allowing for interaction with a wide range of possible partners. The "unwinding" effect on the RNA structure may induce specific structural reorganization of the RNA species that can facilitate the formation of more specific, G-quadruplex-driven interactions with the more distant interaction surfaces of Ezh2. Since the affinity of the loop towards the RNAs is not very high, it allows for the dissociation of the RNA if the specific interactions are not formed within a reasonable timeframe. The phosphorylation of the 345th Threonine results in a stronger unwinding effect, probably increasing the chances of the RNA to reach the other binding regions, achieving an overall stronger RNA binding in the context of the whole protein.

3. *Detailed study of the involvement of HOTAIR and Ezh2 in the EMT induced by the deletion of Tks4*

Studies aimed at understanding the role of Tks4 in cancer cell behavior revealed that HCT116 colorectal carcinoma cells undergo an epithelial-mesenchymal transition (EMT) like process in the absence of Tks4. Since the HOTAIR-Ezh2 interaction is heavily involved in the regulation of EMT through the recruitment of the Polycomb Repressive Complex 2 (PRC2) to induce repression of gene expression, we aimed at uncovering if the HOTAIR-Ezh2 axis is connected to Tks4. In order to gather detailed information on the gene expression changes elicited by the knockout of Tks4, we performed whole transcriptome sequencing of the wild type and the Tks4 KO HCT116 cells and confirmed the variations in gene expression using quantitative PCR (qPCR) of the selected genes. In specific cases we also checked the changes in protein levels, to further corroborate the validity of our observations.

Our results revealed that the deletion of Tks4 results in large-scale reorganization of the gene expression of the cells, with many different signal transduction pathways affected. Among the most affected pathways are the PI3K/Akt cascade, the Jak-Stat and the calcium

signaling, but others, like the mTOR, TNF, TGF-β and NF-kappa B networks also show alterations.

Above the mRNAs, our dataset also contains information on the non-coding RNA expressions and the analysis of these data revealed further details on the alterations that the Tks4 KO induced. Several genes, both coding and non-coding, known to be involved in EMT showed altered expression levels in the Tks4 KO cells, providing a broad genetic background to the observed effects.

Based on these results we were able to outline a map of the different signaling pathways that rely on Tks4 in their correct functioning, complete with the involvement of several non-coding RNAs.

Regarding the role of Ezh2 and HOTAIR in this process, we could show that the Tks4 KO cells have an overall higher level of H3K37me3 histone mark, that is characteristic of the activity of the PRC2. Since the expression level of Ezh2 did not change significantly in the KO cells, we concluded that this elevation is due to the increased activity of the complex. Remarkably, the application of a specific Ezh2 inhibitor, DZnep was able to revert the observed EMT-like changes and the DZnep treated Tks4 KO cells behaved like the untreated wild type controls, indicating that the effects of the Tks4 removal are conveyed through the PRC2. We performed the same analysis of the transcriptome of the DZnep treated cells as previously and we could see that the majority of the gene expression changes observed in the Tks4 KO cells were eliminated by the application of the Ezh2 inhibitor.

To determine if HOTAIR plays a role in this process, we silenced HOTAIR expression using RNA silencing. Interestingly, while the reduction in the HOTAIR expression exceeded 90 %, it had no effect on the behavior of the Tks4 KO cells. This was in line with the expression data from the RNA sequencing and the qPCR, which did not show an extreme overexpression of HOTAIR in the Tks4 KO cells.

Taken together, these results suggest that the signaling pathways involving Tks4 are indeed connected to the PRC2, but not through the HOTAIR-PRC2 interaction. Since the PRC2 is a promiscuous RNA binder and it has other known lncRNA interactors involved in the regulation of EMT, we checked the expression changes of those too, revealing two other candidate lncRNAs that show strong expression changes in the Tks4 KO cells: MALAT1 and NEAT1. As a continuation of the work, we will test the effect of the silencing of these two lncRNAs.

Chromatin-immunoprecipitation sequencing (Chip-Seq) was performed on the wild type, Tks4 KO and DZnep treated cells to identify the PRC2 target genes in HCT116 cells. This revealed several protein coding and non-coding genes that are regulated by PRC2 in colorectal cancer cells, with possible further relevance in the deeper understanding of the molecular background of the disease. An important new result is a comprehensive list of lncRNAs whose expression is under the control of PRC2, because previous publications only describe lncRNAs regulating the activity of PRC2, but no information is available on the role of PRC2 in the control of lncRNA expression.

Published papers

1. Piovesan, D. *et al.* DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Research* vol. 45 D219–D227 (2017).

2. Szabó, B. *et al.* Disordered Regions of Mixed Lineage Leukemia 4 (MLL4) Protein Are Capable of RNA Binding. *Int. J. Mol. Sci.* **19**, (2018).

3. Pancsa, R., Schad, E., Tantos, A. & Tompa, P. Emergent functions of proteins in non-stoichiometric supramolecular assemblies. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* vol. 1867 970–979 (2019).

4. Szabo, B. *et al.* Intrinsically Disordered Linkers Impart Processivity on Enzymes by Spatial Confinement of Binding Domains. *Int. J. Mol. Sci.* **20**, (2019).

5. Bokor, M., Tantos, Á., Tompa, P., Han, K.-H. & Tompa, K. WT and A53T α-Synuclein Systems: Melting Diagram and Its New Interpretation. *Int. J. Mol. Sci.* **21**, (2020).

6. Hatos, A. *et al.* DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.* **48**, D269–D276 (2020).

7. Mészáros, B. *et al.* PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.* **48**, D360–D367 (2020).

8. Murvai, N. *et al.* Interplay of Structural Disorder and Short Binding Elements in the Cellular Chaperone Function of Plant Dehydrin ERD14. *Cells* **9**, (2020).

9. Bokor, M. & Tantos, Á. Secondary Structures of Proteins: A Comparison of Models and Experimental Results. *J. Proteome Res.* **20**, 1802–1808 (2021).

10. Avramov, M. *et al.* Identification of intrinsically disordered proteins and regions in a non-model insect species Ostrinia nubilalis (hbn.). *Biomolecules* **12**, 592 (2022).

11. Micsonai, A. *et al.* BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res.* (2022)

12. Micsonai, A. *et al.* Disordered–Ordered Protein Binary Classification by Circular Dichroism Spectroscopy. *Frontiers in Molecular Biosciences* vol. 9. (2022).

13. Zeke, A. *et al.* Deep structural insights into RNA -binding disordered protein regions. *WIREs RNA* vol. 13 Preprint at https://doi.org/10.1002/wrna.1714 (2022).

14. Szabó, C. L. *et al.* The Disordered EZH2 Loop: Atomic Level Characterization by H- and H-Detected NMR Approaches, Interaction with the Long Noncoding HOTAIR RNA. *Int. J. Mol. Sci.* **23**, (2022).

Papers under submission

1. Mevan Jacksi, Eva Schad, Laszlo Buday, Agnes Tantos: Tks4 at the crossroads of signaling pathways
   Submitted to International Journal of Molecular Sciences

2. Beáta Szabó, András Micsonai, József Kardos, Ágnes Tantos: Detailed investigation of the RNA binding of the intrinsically disordered loop in Ezh2
   Manuscript ready for submission to Protein Science

3. Mevan Jacksi, Eva Schad, Balint Szeder, Laszlo Buday, Agnes Tantos: The connection between Tks4 and Ezh2 in the regulation of epithelial mesenchymal transition
   Manuscript under preparation

4. Mevan Jacksi, Eva Schad, Agnes Tantos: LncRNA expression controlled by Ezh2 in colorectal cancer cells
   Manuscript under preparation

The results of the work were presented at the following conferences

1. FEBS3+ Meeting, Hungary: From molecules to living systems, Siófok, Hungary, September 2-5, 2018
   Oral presentation: RNA-binding of the disordered regions in histone lysine methyltransferases

2. 43rd FEBS Congress, Biochemistry Forever, Prague, Czech Republic, July 7-12, 2018
   Poster presentation: Intrinsic Protein Disorder in Histone Lysine Methylation

3. 3rd Danube Conference on Epigenetics, Budapest, Hungary, October 9-12, 2018
   Poster presentation: RNA-Binding Of The Disordered Regions In Histone Lysine Methyltransferases

4. 2nd Workshop on Intrinsically Disordered Proteins in Core Data Resources, March 13-14, 2019, Prague
   Oral presentation: Intrinsic Protein Disorder in Histone Lysine Methylation

5. Hungarian Molecular Life Sciences Conference, Eger, Hungary, March 29-31, 2019
   Oral presentation: Disordered Regions of Mixed Lineage Leukemia 4 (MLL4) Protein Are Capable of RNA Binding

6. International Symposium on Disordered Proteins, Protein Folding, and Disease-causing Aggregation. KRIBB, Daejeon, Korea, October 22-26, 2019
   Oral presentation: Intrinsically disordered regions of KMT2D harbor RNA interacting segments

7. Onkoplatform, Mátraháza, Hungary, November 20-22, 2019
   Oral presentation: Interaction of LncrRNA And HKMTs In The Regulation of Cellular Processes

8. 25th Annual Meeting of the RNA Society, Online event, May 26-31, 2020
   Poster Presentation: Interaction of lncRNAs and disordered protein regions in histone lysine methyltransferases

9. Hungarian Molecular Life Sciences Conference, Eger, Hungary, November 5-7, 2021
   Poster Presentation: The role of Ezh2 and Tks4 in EMT-related processes

10. Annual Meeting of the Hungarian Biochemical Society, Pécs, Hungary, August 25-27, 2022*
    Oral presentation: Connection between HOTAIR, Ezh2 and Tks4 in Epithelial Mesenchymal Transition