Unraveling the rules of transcriptional repression through repressive domains

Loni Klaus[1,2], Anna Vlasova, Filip Nemcko[1,2] and Alexander Stark[1,3]
Research Institute of Molecular Pathology (IMP)[1], Campus Vienna Biocenter 1, 1030 Vienna, Austria
Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna[2]
Medical University of Vienna[3]

1. Abstract

Transcriptional programs regulate the expression of distinct subsets of genes in different cell types and under changing environmental conditions. While the mechanisms of transcriptional activation are the subject of many previous and ongoing studies, how transcription is repressed is poorly understood. However, the actions of both opponents – activators and repressors - are crucial for the maintenance of transcriptional programs.

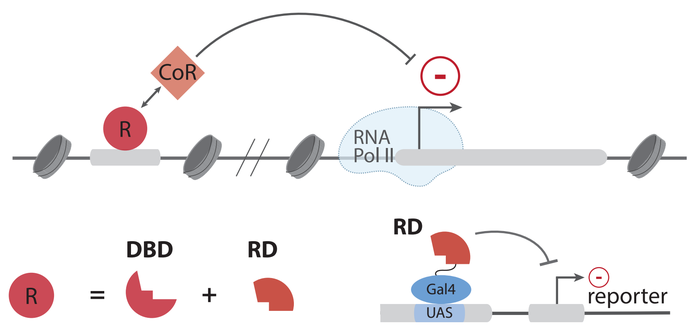
Transcriptional repression is mediated by DNA-binding repressors that contain **repressive domains (RDs)**, which recruit **co-repressors (CoRs)**, such as Groucho or CtBP and it has been shown that RDs alone are sufficient to mediate repression when tethered to a reporter. However, neither the identity nor the properties of many RDs are known, nor have the respective CoRs been identified.

In order to systematically identify RDs, we developed a high-throughput next-generation sequencing-based method called **Repressive Domain-sequencing (RD-seq)**. RD-seq identifies various RDs in a comprehensive pool of short candidate fragments. These domains are part of known repressors, but also proteins that have not been identified as transcriptional repressors so far.

To further dissect the ways by which different groups of RDs lead to transcriptional repression, we will analyze RD sequences and determine the CoRs they recruit.

Altogether, I expect my PhD project to reveal distinct types of RDs, their sequence characteristics and CoRs. This will not only improve our understanding of repression but also lead to a more comprehensive picture of gene regulation in general.

2. Introduction

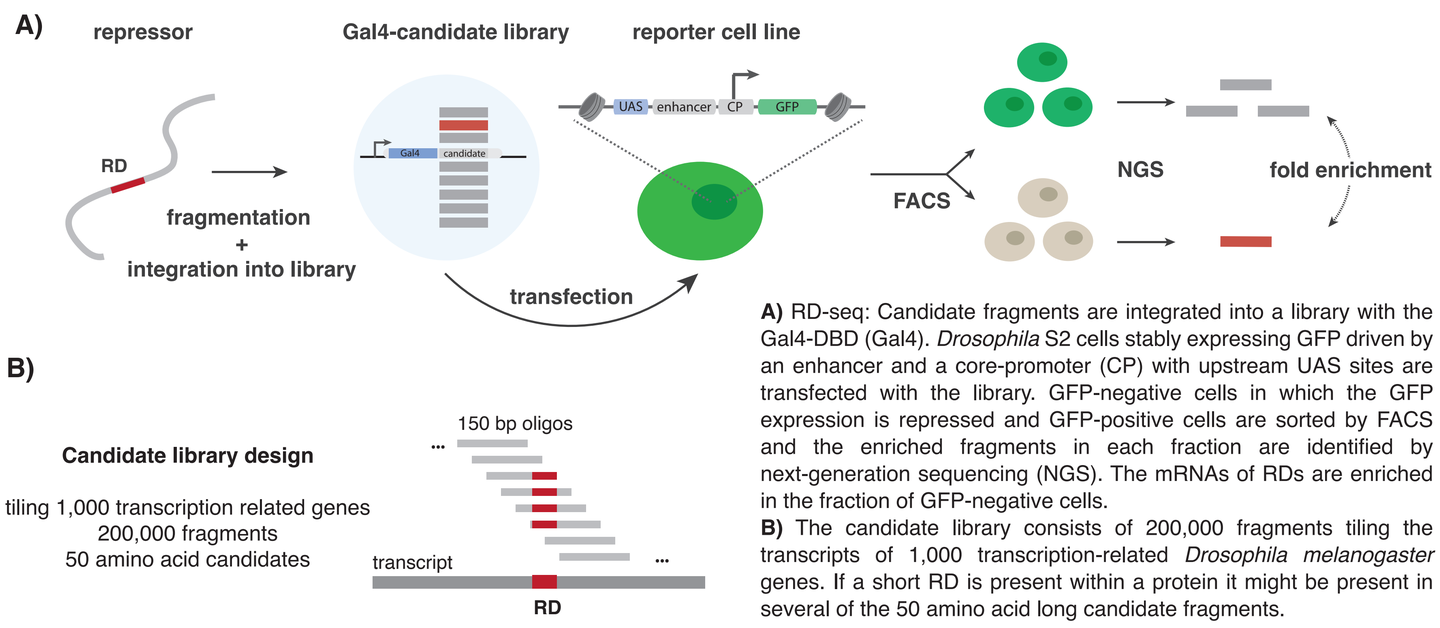


Transcriptional repression is mediated by DNA-binding repressors (R) that recruit non-DNA-binding co-repressors (CoR).
Interestingly, repressors are modular and consist of a DNA-binding domain (DBD) and a repressive domain (RD). Different studies have shown that RDs alone tethered to a transcriptional reporter for example through the Gal4-UAS system are sufficient to mediate the repression of this reporter.

Aims of the PhD project:

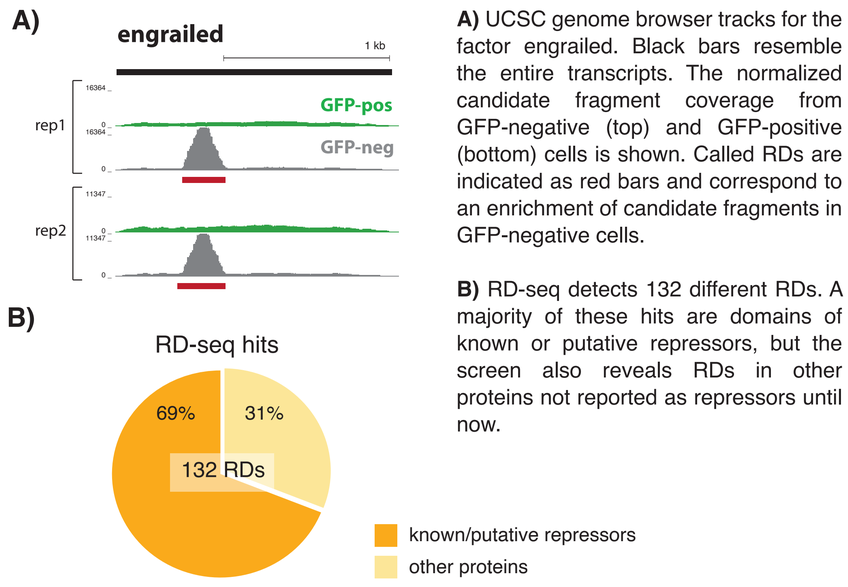
1) **Identify and classify RDs.**
2) **Explore their repressive mechanisms by identifying RD interaction partners.**

3. Approach: Repressive Domain-sequencing (RD-seq)



A) RD-seq: Candidate fragments are integrated into a library with the Gal4-DBD (Gal4). *Drosophila* S2 cells stably expressing GFP driven by an enhancer and a core-promoter (CP) with upstream UAS sites are transfected with the library. GFP-negative cells in which the GFP expression is repressed and GFP-positive cells are sorted by FACS and the enriched fragments in each fraction are identified by next-generation sequencing (NGS). The mRNAs of RDs are enriched in the fraction of GFP-negative cells.
B) The candidate library consists of 200,000 fragments tiling the transcripts of 1,000 transcription-related *Drosophila melanogaster* genes. If a short RD is present within a protein it might be present in several of the 50 amino acid long candidate fragments.

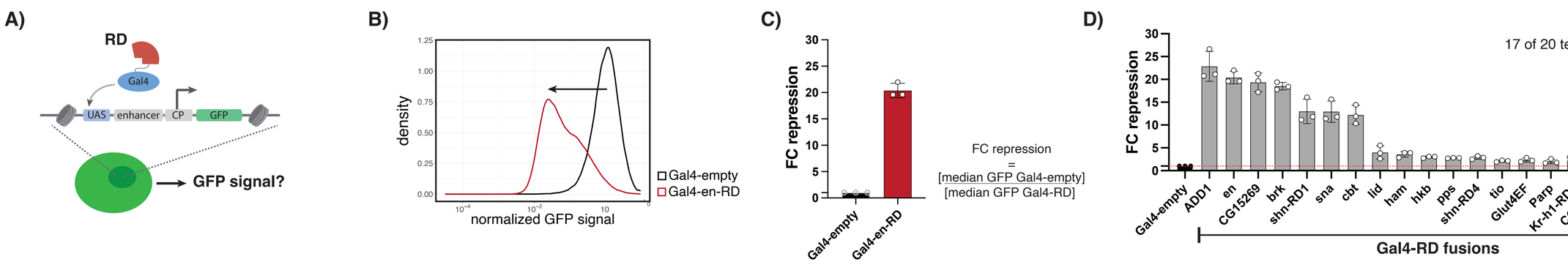
4. Hits of RD-seq screens



A) UCSC genome browser tracks for the factor engrailed. Black bars resemble the entire transcripts. The normalized candidate fragment coverage from GFP-negative (top) and GFP-positive (bottom) cells is shown. Called RDs are indicated as red bars and correspond to an enrichment of candidate fragments in GFP-negative cells.

B) RD-seq detects 132 different RDs. A majority of these hits are domains of known or putative repressors, but the screen also reveals RDs in other proteins not reported as repressors until now.

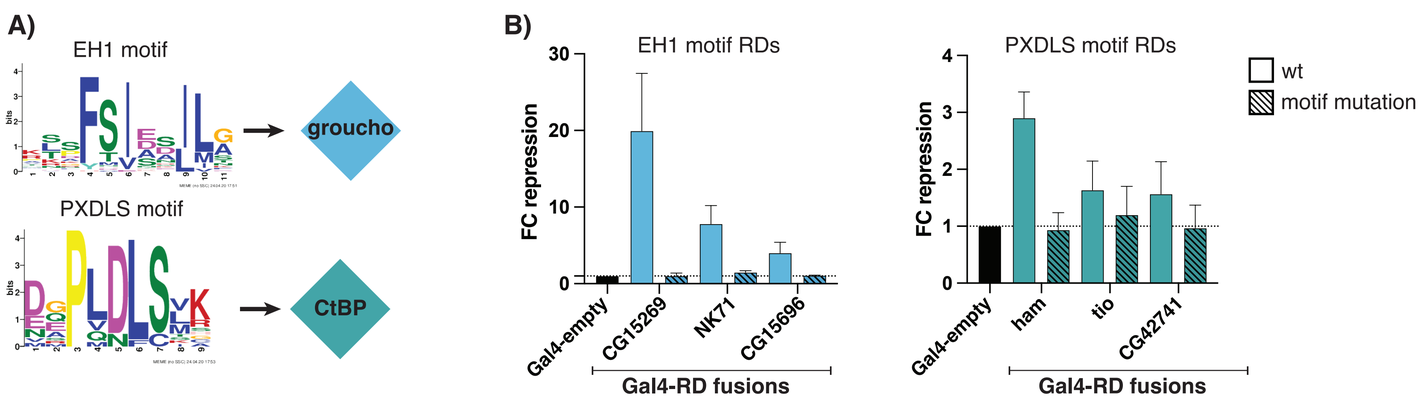
5. Validations of RD-sequencing hits



A) RD-seq hits are validated by expressing Gal4-RD fusions in the same reporter cell line used for the screen and subsequently monitoring changes in the GFP signal by flow cytometry.
B) Using flow cytometry we can observe a decrease of the cell population's GFP signal when the RD of engrailed (en) is tethered to the reporter in comparison to a control cell population expressing an empty Gal4 construct.

C) The strength of an RD is shown as the Fold change (FC) of repression calculated as the offset of the medians of the GFP signal between the RD and the control.
D) With this validation strategy we validated 17 out of 20 tested RDs that show diffferent repressive strengths. Two validated RDs lie within uncharacterized *Drosophila* proteins.

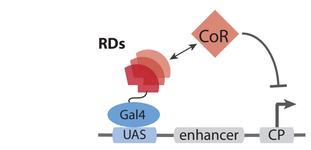
6. Discovery of conserved peptide motif within RDs



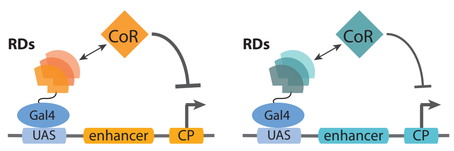
A) Using the MEME motif discovery tool, we found short peptide motifs enriched in the RD-seq hits. The EH1 and the PXDLS motifs have been reported in the literature to recruit the CoRs groucho and CtBP, respectively.
B) Validation experiments for the wild type RDs an RDs with mutated EH1 or PXDLS motifs confirm the requirement of these motifs for the activity of the respective RDs.

7. Outlook

global RDs and CoRs



specific RDs and CoRs



In the future, we will perform RD-seq with reporter cell lines utilizing different enhancer promoter combinations driving GFP expression.
We expect this to reveal global as well as context specific RDs that differ in their sequence features, for example regarding conserved peptide motifs, and in the CoRs they recruit.
We plan to identify these interaction partners of RDs using immunoprecipitation followed by mass spectrometry analysis.

References
Arnold, C. D., Nemcko, F., Woodfin, A. R., Wienerroither, S., Vlasova, A., Schleiffer, A., Pagani, M., Rath, M., Stark, A. (2018). A high-throughput method to identify trans-activation domains within transcription factor sequences. EMBO J. e98896
Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V., Stieger, C. E., Novatchkova, M., Lampersberger, L., Kauneckaite, K., Xie, W., Mechtler, K., Patel, D. J., Brennecke, J. (2019). The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. Nat. Struct. Mol. Biol. 26, 720-731
Cai, Y., Laughon, A. (2009). The Drosophila Smad cofactor Schnurri engages in redundant and synergistic interactions with multiple corepressors. BBA 1789, 232-245
Han, K., Manley, J.L. (1993). Functional domains of the Drosophila Engrailed protein. EMBO J. 12, 2723–2733
Hanna-Rose, W., Licht, J. D. & Hansen, U. (1997). Two evolutionarily conserved repression domains in the Drosophila Kruppel protein differ in activator specificity. Mol Cell Biol 17, 4820–9.
Quinlan, K. G. R., Verger, A., Kwok, A., Lee, S. H. Y., Perdomo, J., Nardini, M., Bolognesi, M., Crossley, M. (2006). Role of the C-Terminal Binding Protein PXDLS Motif Binding Cleft in Protein Interactions and Transcriptional Repression. Mol. Cell. Biol. 26, 8202–8213
Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D. & Jaynes, J. B. (1998). Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes.

Funding

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN
Boehringer Ingelheim
FFG Promoting Innovation.