Disordered TAD Prediction Data Analysis of *Homo sapiens* Transcription Factors

Margaret Campbell, Samantha Renard, Shih-Ting Huang, Dr. Komives
Division of Biological Sciences, University of California, San Diego, La Jolla, CA

Abstract

p53 and NFkB have been shown to contain a disorder transactivation domain (TAD). TADs are a transcription factor scaffold domain which contains binding sites for other proteins, they are part of a transcription factor that is responsible for binding other proteins to activate transcription. To[1] further research and understand this, a database has been created, looking at around 100 transcription factors from homo sapiens from transcriptionfactor.org to compare to the results found within PONDR and 9aa TAD prediction tool to predict which of these transcription factors have a 9 amino acid long disordered domain. After obtaining this information, we will be expressing/producing P50 and P65 in a lab to use as an example of what would be done with the rest of the proteins from the database. The database created is a list of transcription factors with the 9 amino acid long domain from the 100 transcription factors screened.

Introduction to Database/Database Procedure

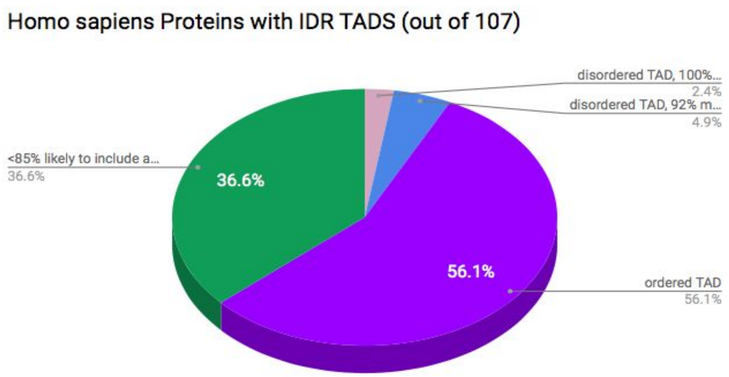
When folded correctly proteins perform their regulatory tasks, however many proteins are disordered, meaning that the protein or part of the protein is missing a set 3D structure, however a disordered protein can be totally functional performing the task it is meant to do with no problem. Proteins can have multiple types of disorders including loops/coils, hot loops, and missing coordinates.[2] These intrinsically disordered proteins (IDPs) are partially or completely unstructured. IDPs are proteins that lack a fixed or ordered three-dimensional structure. [3]

In understanding how the transactivation domain (TAD) works, first the ID and sequence was taken of each amino acid from homo sapiens, then domain range within the sequence of where possible disorders were noted. Once these possible domains were established the intrinsically disordered protein had a range of possible TADs to investigate.

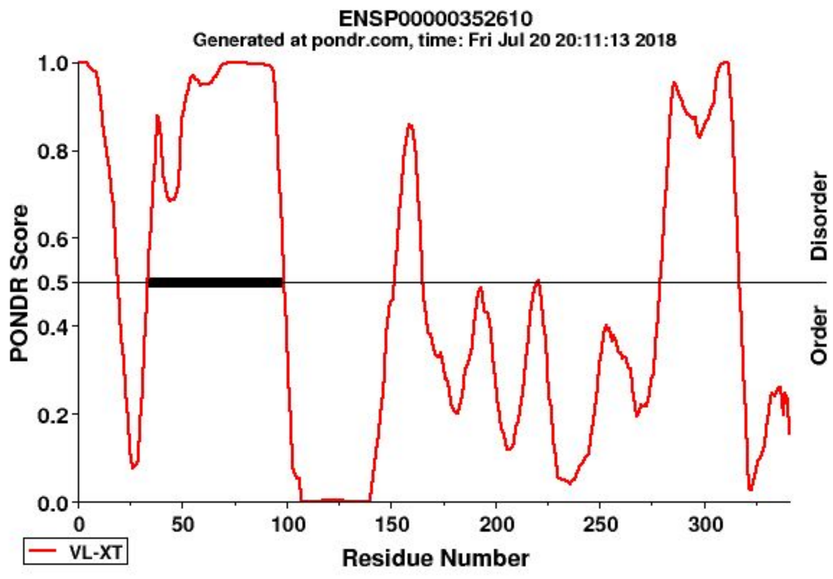
To complete the database, three websites were used, transcriptionfactor.org, PONDR, and 9aa TAD prediction tool. Between these three websites the areas of the disordered proteins were located both visually in a graph and numerically on the sequence. Specifically we were looking for 9 amino acid long transactivation domain (TAD) in these transcription factors. Both p53 and NF-kappa-B have disordered 9 amino acid long TAD and is the reasoning behind looking for 9 amino acid long TADs in other transcription factors.

Results of Database

HOMO SAPIENS: 107 proteins were placed in DBD/PONDR
1 p53, 2 nfkb
79 of the 107 proteins were actually more than 85% likely to include 9 aa TAD regions
45 didn't have any disordered 9 aa TAD, 24 of them had partially disordered 9 aa TAD regions
9 of them have actual disordered 9 aa TAD
Of the 9 actual disordered regions, 3 were perfect matches



1 Transcription Factor.
https://en.wikipedia.org/wiki/AP-1_(transcription_factor) (accessed July 25, 2018).
2 Linding, R., Jensing, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B., Implications for structural proemics.
http://dis.embl.de/html/paper.html (accessed July 25, 2018
3 Intrinsically disordered proteins.
https://en.wikipedia.org/wiki/Intrinsically_disordered_proteins (accessed July 25, 2018)



Database Information

Homo sapiens :ENSP00000352610 (P53)
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMPEAAP
PVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCP
VQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNT
FRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDR
RTEEENLRKKGEPHHELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQDQTSFQKENC
List of domain: p53-like transcription factors 97-287
P53 DNA-binding domain 95-289
IDR:PLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPS
RDRRTEEENLRKKGEPHHELPPGSTKRALPNNTSSSPQ
Predicted IDR: 34-99, 280-317
Relevant IDR: 34-99
Disordered TAD: 38-46

Sequence	Start	End	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12		
ETFSDLWKL	17	25	-	+	+	-	+	+	+	+	+	+	+	+	83% match	Graph
QAMDDLMLS	38	46	+	+	+	+	+	+	+	+	+	+	+	+	Perfect match	Graph
DDIEQWFTE	48	56	-	+	+	-	-	+	+	+	+	+	+	+	75% match	Graph

Conclusion

In conclusion most transcription factors from homo sapiens looked at, did not have the disorder 9 amino acid long TAD domain. Based upon prediction, disordered TAD domains are not common and therefore this makes pp50 and NFKB special. In the future this procedure could redone to test other proteins.

Citations

Derek Wilson, Varodom Charoensawan, Sarah K. Kummerfeld and Sarah A. Teichmann, DBD - taxonomically broad transcription factor predictions: new content and functionality Nucleic Acids Research 2008 doi: 10.1093/nar/gkm964.
Intrinsically disordered proteins. (2018, May 25). Retrieved July 25, 2018, from https://en.wikipedia.org/wiki/Intrinsically_disordered_proteins
Linding, R., Jensing, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction: Implications for structural proteomics. Retrieved July 25, 2018, from http://dis.embl.de/html/paper.html
Molecular Kinetics, Inc., Washington State University and the WSU Research Foundation
RELA. (2018, June 30). Retrieved July 25, 2018, from https://en.wikipedia.org/wiki/RELA
Transcription Factor. (2018, July 10). Retrieved July 25, 2018, from https://en.wikipedia.org/wiki/AP-1_(transcription_factor)
Piskacek, Martin. 9aaTAD Prediction result (2006). Available from Nature Precedings <http://dx.doi.org/10.1038/npre.2009.3984.1> (2009)
Willson, R., & Ladisch, M. R. (1990). Large-Scale Protein Purification. Retrieved July 25, 2018, from https://pubs.acs.org/doi/pdf/10.1021/bk-1990-0427.ch001

Introduction to Lab

The purpose of this lab was to use an expressive vector designed for gene expression to target p50FL and p65FL.[4] The vector used in the lab contains regulatory sequences to become both the enhancer and the promoter regions thereby leading to transcription of the gene carried on the expression vector. In doing this lab p50FL and p65FL will be expressed for further experiment in attempt to understand the TAD.

Lab Procedure

The database used, transcriptionfactor.org, predicted possible IDPs, but, we were unable to test each of the 107 proteins analyzed. Instead, p50FL and p65FL were examined, and the process that was used would be replicated for other proteins.

First, the E.coli BL21 cells were transformed, then plated onto AMP plates. This included placing cell culture samples into a plasmid solution containing our recombinant plasmid, which was genetically modified to encode the required proteins and enzymes that would protect the cells from antibiotics. To get the cells to transform, they were transitioned between heat and ice to put the cells under stress, therefore increasing their intake of plasmids. The cell cultures were then placed in the shaking incubator to receive necessary oxygen and to be moved around to obtain imperative nutrients.

When the cell cultures were placed onto AMP plates, the antibiotic ampicillin killed E.coli that had not taken in the recombinant plasmid, therefore insuring they would not be used. The cultures were inoculated and placed to grow again in the shaking incubator until the solution density reached the optimal density of 0.6 to 0.8.

When the cell cultures became large enough to express the protein, the cultures were removed from the shaker and placed on ice to stop further growth. With the new plasmid, which had modified the Lac Operon, the E.coli now produced the desired protein due to the presence of the inducer, IPTG. The Lac Operon on the plasmid is modified to code for the protein instead of lactose digesting enzymes. The Lac Repressor keeps the RNA polymerase from producing unnecessary lactose-digesting enzymes. Inducers like allolactose can temporarily bind to the repressor so this part of the plasmid is transcribed and translated. To make the E.coli constantly produce the protein, they were induced with IPTG, which would act similar to allolactose and bind to the Lac repressor. When IPTG binded to the Lac repressor, E.coli will continuously produced the desired proteins, p50FL and p65FL.

The E.coli culture was placed in the 18° C shaking incubator where it continued to produce the protein without culture growth or cell death. The broth with the E.coli was then spun down to separate the liquid (supernatant), which included E.coli environment-broth, from the pellet, which only featured the E.coli. The liquid was removed and the formed pellet was resuspended in buffer, then placed into a sonicator. The sonicator used sound waves to burst the E.coli and rip them into pieces so they released the protein they had been building. After another round of centrifuging, the dead E.coli became a pellet, while the supernatant included the proteins.

To purify it, the supernatant was ran through a column, which used three different buffers with different concentrations of Imidazole which resemble Histidine, an amino acid. The plasmid that we modified coded for our protein with poly-Histidine tail. Histein binds to the column, when we washed the column, Imidazole and Histidine both bind to column.When increasing the concentration of Imidazole in the buffer, more imidazole will bind to the column, meaning with enough Imidazole present, the column will release the protein as it is busy binding to the excess Imidazole. The protein is best released from the column with buffer B,which has the highest concentration of Imidazole, the column is busy binding to the excess Imidazole and therefore the protein is releasing from the column. Anything without Imidazole and Histidine will fall through the column as the flow-through, and will we have our protein. After, the protein was dialyzed and stored in SEC (size exclusion column) buffer. To ensure that the proteins were present, a gel was run with standard ladder, the distance traveled by each band indicated the size of the proteins, which can be used to ID the protein. The standard ladder gel was compared to that of the flow-through (what did not bind to the column), the wash (what bound to the column, but not as tightly as the protein), and the elute (the protein).

Results of Lab

The p50FL/65FL Co-expression vector expression and purification lab resulted in pure test quantities of the sample ready for analysis, characterization, and other future testing. After running a gel electrophoresis, it was clear that the column purification had efficiently worked and the cell culture grew to satisfy the target needed to examine the expression made by the gene.

Acknowledgements

I would like to thank everyone in Dr. Komives' lab for their guidance throughout this experience working in the lab. I would specifically like to thank Shih-Ting for being an amazing mentor. Finally, I would like to thank Dr. Komives for creating Research Scholars and for helping me through this unique experience.

RELA.
https://en.wikipedia.org/wiki/RELA (accessed July 25, 2018)