

AID, CHROMATIN, AND UNDERSTANDING THE ROLE OF SPLICING IN SHM

Ph.D. Thesis

By
ANKIT JAISWAL



**DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL
ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JAN 2023**

AID, CHROMATIN, AND UNDERSTANDING THE ROLE OF SPLICING IN SHM

A THESIS

*Submitted in partial fulfillment of the
requirements for the award of the degree
of*
DOCTOR OF PHILOSOPHY

by
ANKIT JAISWAL



**DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL
ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JAN 2023**



INDIAN INSTITUTE OF TECHNOLOGY INDORE

I hereby certify that the work which is being presented in the thesis entitled **AID, CHROMATIN, AND UNDERSTANDING THE ROLE OF SPLICING IN SHM** in the partial fulfillment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **DEPARTMENT/SCHOOL OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2016 to Dec 2022 under the supervision of **Dr. Prashant Kodgire**, Professor, IIT Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Ankit Jaiswal
25/01/23

signature of the student with date
(ANKIT JAISWAL)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

P.V. Kodgire
25/01/2023

Signature of Thesis Supervisor with date
(Prof. PRASHANT KODGIRE)

ANKIT JAISWAL has successfully given his Ph.D. Oral Examination held on
04/12/2023

P.V. Kodgire
04/12/2023

Signature of Thesis Supervisor with date
(Prof. PRASHANT KODGIRE)

ACKNOWLEDGEMENTS

With immense pleasure, I acknowledge the individuals who have supported me in my Ph.D. journey. I would like to begin by expressing my sincere gratitude and appreciation to my thesis supervisor and mentor, Dr. Prashant Kodgire. His unwavering support, motivation, and scientific contribution have been invaluable to this thesis work. His rigorous training and supervision have given me the confidence to work in any research lab and workspace in the future. Under his expertise and supervision, I had the opportunity to identify research problems, plan experiments, collect scientific data, and finally publish the complete scientific story. His lab opened an entire window of opportunities that allowed me to explore and work with the latest scientific techniques, and handle state-of-the-art instruments and equipment. At a personal level, Dr. Prashant Kodgire, an excellent teacher and public speaker, boosted my confidence and taught me the art of delivering a good research presentation.

I would like to extend gratitude to my PSpC members, Dr. Abhijeet Joshi and Dr. Sanjay Kumar Singh, for their valuable guidance and suggestions, which improved the quality of my research immensely. I would also like to thank Dr. Debasis Nayak for his extremely valuable discussion on science and technology, which always inspires and motivates me. I am also thankful to the Head, Department of Biosciences and Biomedical Engineering) and the Discipline Post-Graduate Committee (DPGC) Convener, Department of Biosciences and Biomedical Engineering) for their constant reinforcement and motivation during my research. I express my sincere thanks to all the BSBE department faculty for their continual support. I am also grateful to the Department of Biosciences and Biomedical Engineering for providing me with excellent infrastructural support that contributed to my thesis work.

I am also especially extending my vote of thanks to all the non-teaching BSBE department and support staff, Mr. Arif Patel, Mr. Murphy B. and

Mr. Gaurav Singh, for their extreme support in maintaining and smooth running of instrument facilities as well as raising and processing invoices of research items.

I would like to express my immense gratitude to the Director, Indian Institute of Technology Indore, and I appreciate his vision to make IIT Indore one of the best research institutions in the country. I extend my warm thanks to the Dean of Academic Affairs (DOAA), Indian Institute of Technology Indore, for being the structural and organizational backbone of the research program. I express deep gratitude to the Indian Institute of Technology Indore for providing me with world-class research facilities, comfortable hostel accommodations, and providing a platform for various conferences, lectures, and events that have greatly enriched my Ph.D. journey.

The Sophisticated Instrument Centre (SIC) has helped produce and improve the quality of my research data enormously. I would also like to extend my appreciation for the professional support of SIC's technical staff: Dr. Ravinder Kumar. I also extend my sincere thanks to the academic office staff: Mr. Tapesh Parihar, Mr. Kailash Jamra, Mr. Nitin Parashar, Mr. Anup Sharma, and Ms. Rinki Kukreja for their continuous help and support.

Further, I would like to thank the Council of Scientific and Industrial Research (CSIR), Ministry of Science and Technology, Government of India, New Delhi, for my research fellowship and a stipend during the tenure of my Ph.D. and Indian Council of Medical Research for fellowship post-July 2021 till date.

I extend my deep appreciation to my group members for their support and motivation and for creating an intellectually rich and stimulating research environment.

I am grateful to Dr. Amit Kumar Singh, arguably the best senior and researcher that I have ever known. Dr. Amit was always ready to help and assist me in research and experiments. He guided me from holding

pipette, plasmid isolation, and confocal microscopy to numerous molecular biology techniques. I would also like to thank Ms. Meenal Chaudhary and Dr. Monika Jain for always helping sort research problems. I would also like to thank Tuhin Sarkar for always fruitful discussions related to science and technology. I would also thank my former labmates Ms. Priyanka, Ms. Sakshi, Mr. Vinay, and current lab members Mr. Rahul Chaudhuri, Mr. Brijeshwar Singh, Ms. Surbhi Jaiswal, Ms. Kritika Yadav, Ms. Sushma, Ms. Sayali, Mr. Junaid and Ms. Isha for making conducive research environment.

I extend my special thanks to Ms. Kanika Singh for assisting and performing experiments related to the AID splicing project and the SRSF1-3 project.

I extend my special thanks to Dr. Venkata Narayana Are for providing the MBP expression vector. I am also thankful to him for guiding and suggesting a lot of issue related to protein expression, solubilization and refolding.

As the Ph.D. journey is a roller coaster ride and friends make this journey smooth. I am very lucky to have great friends Dr. Anubhav Tamrakar, Dr. Vishal Anand, Dr. Gaurav Pandey, and Dr. Pranav Tiwari. Dr. Anubhav & I joined IIT Indore in Dr. Prashant's Lab in 2016, and we spent a lot of time together. Anubhav has always motivated, inspired, and helped me in research and all aspects of life. Additionally, he is a very good researcher, and his mind is always filled and excited with millions of research ideas and solutions. I remember during COVID times, he was interested in designing a COVID detection kit (antigen detection kit), but due to a difficult situation, it didn't happen. Importantly, the COVID detection kit was launched after a few months, which was a great motivation for him. I am thankful to Vishal for always cooking delicious food and inviting me for lunch/dinner. His improvement and dedication to research are highly appreciated. Gaurav is a most dynamic student, and our last resort to find solutions to any problems. I have learned a lot of life skills from him. Pranav is a self-

motivated researcher, a great human being, and an extrovert guy. I enjoyed a lot with all of my friends during my Ph.D. journey.

I would also extend my thanks to our collaborators Dr. Rajarshi Roy, Dr. Nisha Jonnaiya, and Dr. Parimal Kar. Rajarshi was a good researcher and always ready to help. He was the reason why I ate a lot of fast food from Nafees and Noor.

I also extend my thanks to Dr. Mena Asha Krishnan for supporting and assisting me during my Ph.D. days. I have learned a lot of things, such as patience, problem-solving skills, and research aptitude, from her.

Most importantly, none of this would have been possible without the support of my family. I sincerely express my love and gratitude to my parents and my sister for their support of my Ph.D. and in my life.

ANKIT JAISWAL

Dedicated to my family and friends

SYNOPSIS

Introduction

Antibody diversity is a key event inside B-cells that limits/eliminates pathogens. B-cells produce antibody diversity via VDJ recombination, somatic hypermutation (SHM), and class switch recombination (CSR). Interestingly, activation-induced cytidine deaminase (AID) is the initiator of SHM and CSR. AID mediates its effect by deamination of cytosine to uracil in the variable region (in case of SHM) and the constant region (in case of CSR), resulting in nucleotide mismatch (**Fig. 1**). Subsequently, AID-created mismatches are repaired in error-prone fashion to give rise high-affinity antibody and isotypic switching. Usually, AID activity is confined to immunoglobulin (Ig) genes, but in many cases, it is reported to be mistargeted to non-Ig genes resulting in deleterious effects. AID mistargeting can lead to B-cell-related malignancy.

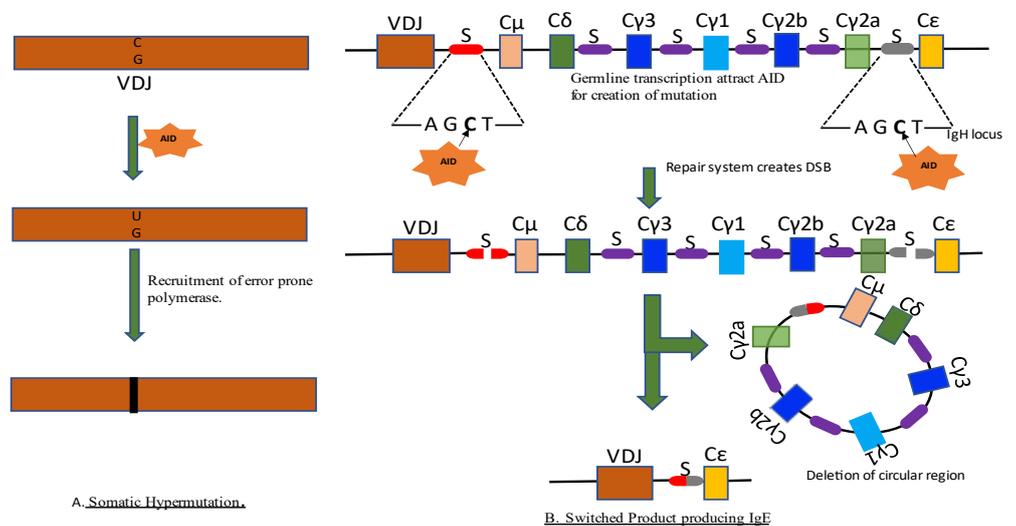


Fig.1. Molecular events that lead to antibody diversity. (A) The variable region of Ig genes is deaminated by AID that initiates SHM. (B) The switch regions are deaminated by AID which mediates CSR.

AID targeting Ig genes is a complex phenomenon that requires a lot of crosstalk between different proteins and cis-acting elements. After the

discovery of AID two decades ago, numerous AID-interacting factors have unfolded. However, the complete understanding of molecular mechanisms that result in AID targeting is still an enigma. Remarkably, AID targeting immunoglobulin (Ig) genes requires transcription [1] as it provides an opportunity for AID to deaminate the single-stranded DNA. Besides, transcription nucleosome restricts the AID accessibility to the DNA. Interestingly, immune cells recruit chromatin modifiers like the facilitates chromatin transcription (FACT) complex or histone cell cycle regulator (HIRA) chaperon complex to overcome the nucleosome barrier. Further, the HIRA chaperon complex is responsible for the deposition of histone variant H3.3 in a replication-independent manner at the transcription start site and gene regulatory regions that make chromatin accessible for AID-mediated mutation.

Transcription through Ig genes induces the supercoiling of DNA that is relaxed by topoisomerase 1 (TOP1). Interestingly, TOP1 inhibition promotes antibody diversity and vice-versa. AID was also found to interact with a splicing regulator SRSF1-3 that has a role in SHM. Moreover, SRSF1-3 was reported to interact with TOP1 [2].

Background of work and objective

AID was reported to interact with replication protein A (RPA) [3], and HIRA protein, a member of the HIRA chaperon complex, was found to interact with RPA [4]. Interestingly, HIRA protein is also required for the deposition of H3.3 at Ig genes, and it is indispensable for SHM as HIRA KO severely impaired somatic mutation. However, the interaction between AID and ubinuclein 1 (UBN1), a member HIRA chaperon complex, has not been reported yet. In this study, we explored the interaction of AID with UBN1 by *in-silico* as well as *in-vitro* methods, which may play a potential role in antibody diversification (the role in SHM or CSR still needs to be explored).

AID expressed in human cells is mostly full-length hAID (FL-hAID), but its splice forms are also reported in various lymphomas. Further, AID is purified from SF9 cells and *E. coli* cells for routine

biophysical, kinetics, and structural studies. Surprisingly, hAID expressed in *E. coli* showed two induced bands instead of a single band. In this study, we unfold that heterologous expression of hAID in *E. coli* cells produces a splice isoform of hAID δ C (hAID containing missing amino acids near the C-terminal domain).

SRSF1, a splicing regulator, interacts with TOP1, inhibiting the latter's activity [5]. Moreover, TOP1 inhibition promotes AID-mediated antibody diversity. Interestingly, SRSF1-3, a splice isoform of SRSF1, was also found to interact with TOP1 [2]. Nevertheless, SRSF1-3 inhibition of TOP1 activity is not reported. In this study, we identify that SRSF1-3 influences TOP1 activity, leading to promoting antibody diversity.

The main objectives of this study are as follows: -

1. To investigate the potential interaction between chromatin modifier UBN1 and antibody diversification enzyme AID in B-cells.
2. To study why the heterologous expression of hAID in *E. coli* cells results in the expression of two bands.
3. To explore the fate of SRSF1-3 interaction with TOP1 on the DNA nicking activity of the latter.

Activation-induced cytidine deaminase, an antibody diversification enzyme, interacts with chromatin modifier UBN1 in B-cells

AID targeting Ig genes is dependent on transcription as it provides single-stranded DNA. The DNA in eukaryotes is packed in the form of nucleosomes, restricting the AID accessibility to the DNA. To overcome this limitation eukaryotes cells are well equipped with chromatin modifiers or chaperon complexes like the HIRA chaperon complex, which results in an open chromatin state. HIRA chaperon complex consists of UBN1, HIRA, anti-silencing function 1A histone chaperone (ASF1a), and calcineurin binding protein 1 (CABIN1) that deposit H3.3 at the gene regulatory region. Remarkably, HIRA chaperon

complex specificity to H3.3 is imparted through UBN1 and is responsible for de novo deposition of H3.3 at the Ig locus, which is also the hallmark of SHM. As HIRA protein, a member of the HIRA chaperon complex interacts with RPA, and RPA interaction with AID is also well established, we explored the possibility of whether AID interacts with UBN1 or not, which may have a potential role in antibody diversification. Subsequently, we used computational studies, such as molecular docking as well as molecular dynamics, co-immunoprecipitation, *in-vitro* pull-down assay, double immunofluorescence assay as well as proximity ligation assay, to establish AID interaction with UBN1. The *in-silico* study showed that the N-terminal domain of UBN1 is mainly involved in the interaction with AID. Further, UBN1 and AID interactions are mainly driven by the hydrophobic region. As anticipated, co-immunoprecipitation data suggest that AID indeed interacts with UBN1 in DT40ΨV KO cells and Raji cells (**Fig. 2**). Subsequently, co-immunoprecipitation was also performed on other members of the HIRA chaperon complex such as HIRA and ASF1a against AID to rule out any possibility of AID interacting with other members of the HIRA complex. AID interaction with neither HIRA nor ASF1a was detected in DT40ΨV KO and Raji cell lines validating that AID interaction with UBN1 is specific. Eventually, co-immunoprecipitation of UBN1 and AID was also performed on DT40 AID KO (AID gene is knocked out in DT40 cells) control cells which yielded no band. Additionally, *in-vitro* pull-down with cell lysate and purified protein suggests that UBN1 interacts with AID. Besides AID interaction with UBN1 detected by CoIP, both AID as well as UBN1 should be present at the same location inside the cell. The double-immunofluorescence assay using fluorescently labeled antibodies showed that AID (primarily localized in the cytoplasm) and UBN1 (mostly present inside the nucleus) are co-localized inside the nucleus as represented by yellow or orange dots in DT40ΨV KO and Raji cell lines. Moreover, co-localization was also performed in DT40 AID KO control cell lines to eliminate any possibility of background fluorescence artifact, which showed no co-localized yellow dots. Thus,

co-localization of AID and UBN1 inside the nucleus may have a fate in SHM. Subsequently, a proximity ligation assay (PLA) was performed to determine whether the interaction of UBN1 and AID is direct or indirect. As anticipated, PLA spots were seen in Raji and DT40 ψ V KO cells that suggest interaction between UBN1 and AID is direct.

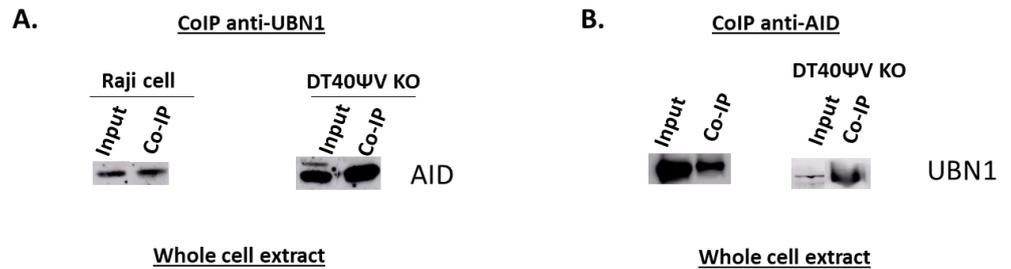


Fig. 2. AID interaction with UBN1 in Raji and DT40 ψ V KO cells. (A) Co-immunoprecipitation of UBN1 from whole cell lysate was performed on Raji and DT40 ψ VKO cells using anti-UBN1 pAb and proteinA/G beads, detection via WB using anti-AID mAb. (B) Co-immunoprecipitation of AID from whole cell lysate of Raji & DT40 ψ V KO cells using anti-AID mAb and protein A/G beads, detection via WB using anti-UBN1 mAb.

Our study is the first report that AID interacts with a member of the HIRA chaperon complex, UBN1. As HIRA chaperon complex is responsible for the deposition of H3.3 at the Ig locus and promotes SHM. AID interaction with UBN1 may have a role in antibody diversity that still needs to be explored (**Fig. 3**).

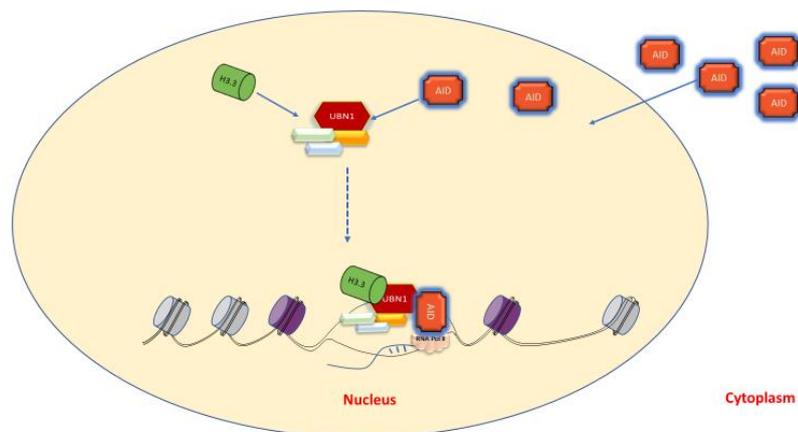


Fig. 3. The fate of AID-UBN1 interaction. UBN1 may bring AID to chromatin along with H3.3, which in turn makes AID more accessible to chromatin.

Heterologous expression of hAID in *E. coli* cells leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored.

AID enzyme is the crucial player that dictates the initiation of antibody diversification processes such as SHM and CSR. Moreover, AID stands at the crossroads between immunity and cancer, as optimal AID activity provides immunity, whereas even a minute fluctuation in its activity leads to cancer. AID is a key player in the immune system, and a lot of studies are going on to understand the structure, function, regulation, and targeting of AID. Moreover, for most studies, AID is frequently purified either from the SF9 insect cell line [6] or *E. coli* cells [7]. Interestingly, when we expressed AID in *E. coli* cells, it produced two distinct induced bands, the upper band belonging to full-length AID and a lower band (Fig. 4). We used techniques like protein purification, western blotting, peptide mapping, the mutation in AID gene and addition of solubilization tag like N utilization substance A (NusA) and maltose-binding protein (MBP) to explore the reason behind the appearance of two bands whenever hAID is expressed.

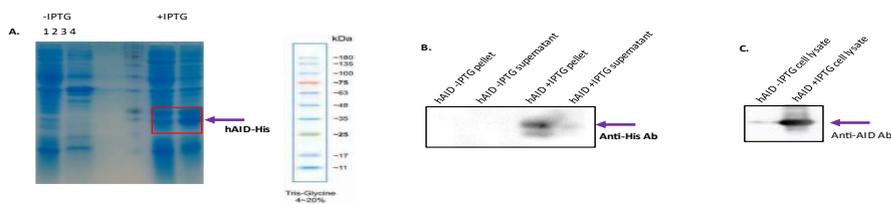


Fig. 4. (A) Induction of hAID-His in *E. coli* cells. lane 1: uninduced supernatant, lane 2: uninduced pellet, lane 3: induced supernatant, lane 4: induced pellet. (B) Western blot of uninduced and induced hAID with anti-His antibody. lane 1: uninduced pellet, lane 2: uninduced supernatant, lane 3: induced pellet, lane 4: induced supernatant. (C) Western blot of uninduced and induced hAID with anti-AID antibody. lane 1: uninduced cell lysate, lane 2: induced cell lysate

To know the identity of two expressed bands of hAID-His western blotting was performed. Surprisingly, the anti-His antibody picked both upper and lower bands, whereas the anti-AID antibody showed only a single band corresponding to full-length AID. So,

Western blotting of hAID-His tells that both bands belong to expressed protein. Subsequently, peptide mapping of upper and lower bands was performed, which showed that both bands belong to hAID. Unexpectedly, the lower band of hAID showed a few missing amino acids from exon 4, which in literature is known as hAID δ E4a, the AID isoform expressed in B-cells. Further, we asked whether the mutation in this putative splice site influences the expression of two bands of hAID or not. The putative splice mutant of hAID does not impact the expression of two bands.

To further investigate whether adding a solubilization tag to hAID-His still produced two expressed bands, we cloned AID along with NusA-tag. As anticipated, Nus-hAID-His produced two induced bands like hAID-His only, one band corresponding to full-length Nus-hAID-His, whereas another band is slightly lower than full-length hAID. Subsequently, purification, as well as western blotting of Nus-hAID-His, showed the two bands. Moreover, we switched the solubilization tag from NusA to MBP to rule out any ambiguity that the NusA tag is prone to proteolysis which may produce two induced bands of hAID. We performed His-MBP-hAID expression, purification as well as western blotting that showed that indeed two bands were produced like hAID-His and Nus-hAID-His. Moreover, western blotting using an anti-AID antibody showed only a single band corresponding to full-length hAID in each case, whereas the lower band that may have some missing amino acids was not picked by the anti-AID antibody. Finally, we performed peptide mapping of Nus-hAID-His and His-MBP-hAID, which clearly showed that the upper band in each case is full-length hAID whereas the lower band is also hAID but contains a few missing amino acids located near the C-terminal end.

In this study, we conclude that heterologous expression of hAID in *E. coli* cells produced a splice isoform hAID δ C beside the full-length hAID. Further, the possibility of proteolysis behind the appearance of the lower band of hAID is ruled out as the lower band was able to bind to his affinity column, picked by the western blot using an anti-His

antibody as well as peptide mapping. Furthermore, a study needs to be done to identify whether the appearance of two bands of hAID is due to protein splicing phenomena or RNA splicing.

SRSF1-3 interaction with TOP1 results in the inhibition of the latter activity, which may promote SHM.

SRSF1-3 is the splice isoform of SRSF1, which is essential for SHM and promotes the accumulation of AID inside the nucleus. Human SRSF1-3 consists of 244 amino acids comprising two RNA binding domains, namely RRM1 and RRM2, separated by a spacer, and a c-terminal domain. SRSF1-3, like SRSF1, was also reported to interact with TOP1 [5]. However, unlike SRSF1, the fate of the interaction of SRSF1-3 on the activity of TOP1 in the context of SHM is largely unknown. In this study, we unravel that SRSF1-3 influences the TOP1 DNA nicking activity locally, which may be one of the mechanisms mediated by SRSF1-3 to promote SHM.

SRSF1-3 structure predicted by alphafold2 was docked with the TOP1 structure available on the protein data bank by Cluspro software. As anticipated, docked model 1 with the highest free energy of -815 Kcal/Mol showed a clear interaction of SRSF1-3 with TOP1. In the absence of a complex structure of TOP1 with SRSF1-3, and to shed insights into their binding mechanism and conformational changes in TOP1 upon its interaction with the SRSF1-3, we investigated the overall conformational dynamics and stability of the complex using the MD simulations. We investigated the overall structural stability and calculated the probability density of each protein as well as the apo TOP1 throughout the MD simulation, as shown (**Fig. 5A**). These results reveal that compared to the apo TOP1, the interaction with the SRSF1-3 stabilizes the topoisomerase as seen by the sharp peak of the complex. In addition, it shows that TOP1 is quite stable compared to the SRSF1-3 in the complex, and a clear picture of the stabilization of TOP1 can be observed by the sharp peak in contrast to the flattened peak of apo TOP1. Also, the different regions of the TOP1 in its complex system were

compared with the apo TOP1 and shown (**Fig. 5B**). It suggests that overall, the structure got stabilized upon its interaction with the SRSF1-3 as can be seen from the shift in the peak of different regions of TOP1 in the complex compared with the apo system. Additionally, RRM1 and RRM2 of the SRSF1-3 are more stable than the spacer and c-terminal domain favouring the interaction with TOP1.

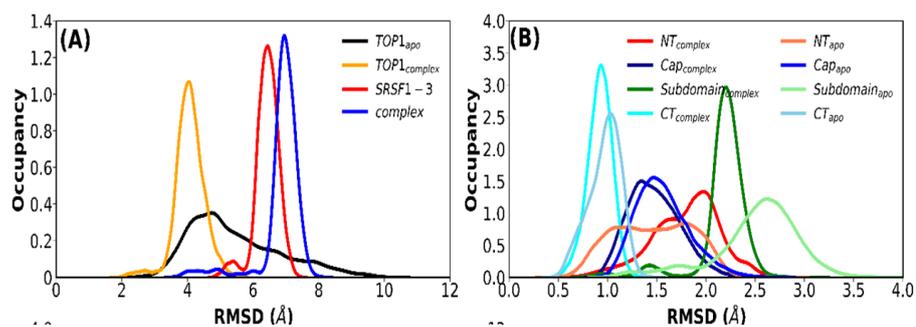


Fig. 5. (A) The root-mean-square deviations (RMSD) of the backbone of TOP1 complex and apo (B) The different regions of the TOP1 RMSD of both apo and complex.

Further, to know the impact of SRSF1-3 on the DNA-nicking activity of TOP1, we expressed SRSF1-3 in bacterial cells. As the expression of SRSF1-3 was in inclusion bodies, we solubilized, refolded, buffer exchanged, and concentrated SRSF1-3. Subsequently, we performed TOP1 inhibition activity in the presence of an increasing concentration of SRSF1-3. Interestingly, we found that SRSF1-3 impacts the TOP1 activity. Further, the impact of SRSF1-3 on TOP1 activity may play a crucial role in SHM.

Scope of the thesis

The thesis is broadly divided into six chapters. The introduction and an appropriate review of literature heading towards the objective make up chapter 1 of the thesis. Chapter 2 extensively mentions the experimental protocols that are used in the achievement of the objective. Subsequently, chapter 3 demonstrates the results to establish the interaction of AID with UBN1. Chapter 4 talks about the results obtained when hAID genes are overexpressed in *E. coli* cells and discusses the occurrence of splice isoform of hAID δ C. Chapter 5 covers

the results of the impact of SRSF1-3 on the activity of TOP1. Finally, chapter 6 concludes the Ph.D. work by providing a summary as well as prospects that give a better understanding of AID-mediated antibody diversification.

In a nutshell, our work identifies one important AID interacting partner, UBN1. However, the fate of this interaction in terms of SHM still needs to be explored. Interestingly, our study on AID expression in *E. coli* cells produced two bands: FL-hAID and splice isoform hAID δ C. Subsequently, it provides that heterologous expression of the foreign gene in *E. coli* cells may be spliced out either at RNA or protein level to produce FL as well as splice isoform. Lastly, our results also highlight how SRSF1-3 affects the activity of TOP1 and may promote SHM.

References

1. Peters A., Storb U. (1996), Somatic hypermutation of immunoglobulin genes is linked to transcription initiation, *Immunity*, 4, 57–65 ([https://doi.org/10.1016/s1074-7613\(00\)80298-8](https://doi.org/10.1016/s1074-7613(00)80298-8))
2. Singh A.K., Tamrakar A., Jaiswal A., Kanayama N., Agarwal A., Tripathi P., et al. (2019), Splicing regulator SRSF1-3 that controls somatic hypermutation of IgV genes interacts with topoisomerase 1 and AID, *Mol Immunol*, 116, 63–72 (<https://doi.org/10.1016/j.molimm.2019.10.002>)
3. Chaudhuri J., Khuong C., Alt FW. (2004), Replication protein A interacts with AID to promote deamination of somatic hypermutation targets, *Nature*, 430, 992–8 (<https://doi.org/10.1038/nature02821>)
4. Zhang H., Gan H., Wang Z., Lee J.H., Zhou H., Ordog T., et al. (2015), RPA Interacts with HIRA and Regulates H3.3 Deposition at Gene Regulatory Elements in Mammalian Cells. *Mol Cell*, 65, 272–84 (<https://doi.org/10.1016/j.molcel.2016.11.030>)
5. Ishikawa T., Krzysko K.A., Kowalska-Loth B., Skrajna A.M., Czuby A., Girstun A., et al. 2012, Activities of topoisomerase I in its complex with SRSF1, *Biochemistry*, 51, 1803–16. (<https://doi.org/10.1021/bi300043t>)
6. Qiao Q., Wang L., Meng F.L., Hwang J.K., Alt FW., Wu H. (2017), AID Recognizes Structured DNA for Class Switch Recombination, *Mol Cell*, 67, 361-373.e4. (<https://doi.org/10.1016/j.molcel.2017.06.034>)
7. Zheng S., Vuong B.Q., Vaidyanathan B., Lin J.Y., Huang F.T., Chaudhuri J. (2015), Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA, *Cell*, 161, 762–73. (<https://doi.org/10.1016/j.cell.2015.03.0>)

List of Publications

(A) Publications from PhD thesis work:

1. **Jaiswal A.**, Singh A.K., Tamrakar A., Kodgire P. (2021), Unfolding the role of splicing factors and RNA debranching in AID mediated antibody diversification, *Int Rev Immunol*, 14, 1-18 (DOI: 10.1080/08830185.2020.1815725) Impact factor: 5.08.
2. **Jaiswal A.**, Jain M., Singh K., Kodgire P. (2022), Heterologous expression of hAID in *E. coli* leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored, *Protein Expr Purif*, 199:106149. (DOI: 10.1016/j.pep.2022.106149) Impact factor: 2.6
3. **Jaiswal A.**, Roy R., Tamrakar A., Singh A.K., Kar P., Kodgire P. (2023), Activation-induced cytidine deaminase an antibody diversification enzyme interacts with chromatin modifier UBN1 in B-cells, *Nat. Sci. Rep*, 13:19615 (DOI: 10.1038/s41598-023-46448-7) Impact factor: 4.6.

(B) Manuscript under preparation from PhD thesis work:

1. **Jaiswal A.**, Singh K., Jonniya N.A., Arya V.N., Kar P., Kodgire P. SRSF1-3 interaction with TOP1 inhibits TOP1 DNA nicking activity.

(C) Publications apart from PhD thesis work:

1. Singh A.K., Tamrakar A., **Jaiswal A.**, Kanayama N., Kodgire P. (2020), SRSF1-3, a splicing and somatic hypermutation regulator, controls transcription of IgV genes via chromatin regulators SATB2, UBN1 and histone variant H3.3, *Mol Immunol*, 119, 69-82 (DOI: 10.1016/j.molimm.2020.01.005) Impact factor: 4.17.
2. Jain N., Mishra S.K., Shankar U., **Jaiswal A.**, Sharma T.K., Kodgire P., Kumar A. (2020), G-quadruplex stabilization in the ions and maltose transporters gene inhibits salmonella enterica growth and virulence, *Genomics*, 112(6), 4863-4874 (DOI: 10.1016/j.ygeno.2020.09.010) Impact factor: 6.2

3. Singh A.K., **Jaiswal A.**, Kodgire P. (2019), AID preferentially targets the top strand in nucleosome sequences, *Mol Immunol*, 112, 198-205 (DOI: 10.1016/j.molimm.2019.05.015) Impact factor: 4.17.

4. Singh A.K., Tamrakar A., **Jaiswal A.**, Kanayama N., Agarwal A., Tripathi P., Kodgire P. (2019), Splicing regulator SRSF1-3 that controls somatic hypermutation of IgV genes interacts with topoisomerase 1 and AID, *Mol Immunol*, 116, 63-72 (DOI: 10.1016/j.molimm.2019.10.002) Impact factor: - 4.17.

5. Choudhary M., Tamrakar A., Singh A.K., Jain M., **Jaiswal A.**, Kodgire P. (2017), AID Biology: A pathological and clinical perspective, *Int Rev Immunol*, 37(1), 37-56 (DOI: 10.1080/08830185.2017.1369980) Impact factor: 5.08.

(D) List of international conferences

1. **Jaiswal A** and Kodgire P. (2018) "Identifying the off-targets of Activation-induced cytidine deaminase in B-cell lymphoma" in a symposium on Emerging areas in Biosciences and Biomedical Technology held on 5-6 Jan at Indian Institute of Technology Indore.

TABLE OF CONTENTS

CONTENTS		PAGE NO.
1.	LIST OF FIGURES	XXVII
2.	LIST OF TABLES	XXXI
3.	Abbreviations	XXXIII
 Chapter 1: Introduction and Literature Review		
1.1	Immunoglobulin diversification	1
1.2	Somatic hypermutation (SHM)	1-2
1.3	Class switch recombination (CSR)	2-4
1.4	AID and its splice isoforms	4-5
1.5	Regulation of AID	5-7
1.6	Role of transcription in AID-mediated antibody diversity	7-8
1.7	Effect of a nucleosome on AID activity	8
1.8	Role of chromatin modifiers or histone chaperon in antibody diversity	9
1.9	Role of AID interacting partner in antibody diversity	9-10
1.10	Role of SRSF1-3 in antibody diversity	10
1.11	Background of work and objective	10-11

Chapter 2: Materials, methods, and Instrumentation

2.1	Materials	
2.1.1	Cell lines	15
2.1.2	Chemicals	15
2.1.3	Reagents and Kits	16
2.2	Methods and Instrumentation	
2.2.1	<i>E. coli</i> Competent cell preparation	16
2.2.2	Transformation of plasmid in <i>E. coli</i> competent cells	16
2.2.3	Isolation of plasmid DNA	16-17
2.2.4	Polymerase chain reaction	17
2.2.5	Agarose gel electrophoresis	17
2.2.6	Gel Extraction	17-18
2.2.7	PCR Purification	18
2.2.8	Restriction digestion	18
2.2.9	Ligation	18
2.2.10	B-Cell culture	18
2.2.11	Structural modelling of UBN1 and docking analysis	18-19
2.2.12	Molecular Dynamics Simulation of UBN1-AID complex	19

2.2.13 Binding free energy calculation using MM/GBSA scheme	19-20
2.2.14 Co-Immunoprecipitation assay	20-21
2.2.15 Western immunoblotting	21
2.2.16 Double immunofluorescence microscopy	21
2.2.17 Cloning & Expression of various proteins used in pull down	22
2.2.18 Pull-down assay	22-23
2.2.19 Proximity Ligation Assay	23-24
2.2.20 Cloning of hAID-His, Nus-hAID-His and His-MBP-hAID	24
2.2.21 Expression of hAID-His	24
2.2.22 Mass spectrometry sample preparation of hAID-His, Nus-hAID-His and His-MBP-hAID	25
2.2.23 cDNA preparation and quantitative PCR	25-26
2.2.24 Mutation of putative splice junctions of human AID	26
2.2.25 Purification and western blotting of Nus-hAID-His, His-MBP-hAID	26-27
2.2.26 Structural modelling of SRSF1-3 and docking with Topoisomerase1	27
2.2.27 Molecular dynamics simulation of SRSF1-3-TOP1 complex	28
2.2.28 Cloning and expression of His-SRSF1-3 and His-SRSF1-3ΔC	28-29

2.2.29 Cloning, expression and purification of His-MBP-SRSF1-3	29
2.2.30 Topoisomerase 1 inhibition assay	29
2.2.31 Solubilization, purification, refolding and dialysis of His-SRSF1-3	30

Chapter 3: Activation-induced cytidine deaminase an antibody diversification enzyme interacts with chromatin modifier UBN1 in B-cells

3.1 Introduction	35-36
3.2 Results	
3.2.1 <i>In-silico</i> estimation of UBN1-AID interaction	37-43
3.2.2 Interaction of AID with UBN1	44-52
3.2.3 Co-localization studies for AID and UBN1	52-58
3.2.4 Proximity ligation assay for direct interaction between AID and UBN1	58-61
3.3 Summary	62

Chapter 4: Heterologous expression of hAID in *E. coli* leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored.

4.1 Introduction	67-68
4.2 Results	
4.2.1 Expression of WT hAID-His	69
4.2.2 Mass spectral analysis of hAID-His	69-72

4.2.3	Expression of Nus-hAID-His and His-MBP-hAID	72-75
4.2.4	hAID and hAID δ E4a	75
4.2.5	Expression of splice site mutant of Nus-hAID-His	75-79
4.2.6	PCR from cDNA of induced and uninduced Nus-hAID-His	79
4.2.7	Mass spectral analysis of Nus-hAID-His and His-MBP-hAID	79-81
4.2.8	RNA structure prediction of AID mRNA	81-84
4.3	Summary	85
Chapter 5: SRSF1-3 interaction with TOP1 likely inhibits TOP1 DNA nicking activity		
5.1	Introduction	89-90
5.2	Results	
5.2.1	<i>In-silico</i> estimation of SRSF1-3-AID interaction	91-95
5.2.2	Expression of His-SRSF1-3 and His-SRSF1-3 Δ C	95
5.2.3	Expression and purification of His-MBP-SRSF1-3	95-98
5.2.4	Topoisomerase 1 inhibition assay	98
5.2.5	Solubilization, purification, and refolding of His-SRSF1-3	98-102
5.3	Summary	103

Chapter 6: Conclusions and scope for future work

6.1	Introduction	107-108
6.2	Activation-induced cytidine deaminase an antibody diversification enzyme interacts with chromatin modifier UBN1 in B-cells	109
6.3	Heterologous expression of hAID in <i>E. coli</i> leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored.	110-112
6.4	SRSF1-3 interaction with TOP1 likely inhibits TOP1 DNA-nicking activity	112-114
6.5	Conclusion	114-115
6.6	Future directions	115-116
	Appendix-A Vector maps	119-121
	Appendix-B List of primers used in PCR	122
	Appendix-C Antibodies used for co-immunoprecipitation, western blotting, and immunofluorescence microscopy	123
	Appendix-D Peptide mapping files	124-132
	References	133-148

LIST OF FIGURES

FIGURES	PAGE NO.
Chapter 1: Introduction and review literature	
Figure 1.1 Molecular mechanism of SHM and CSR.	3
Figure 1.2 The functional domain organization of AID & its splice isoforms	6
Chapter3: Activation-induced cytidine deaminase an antibody diversification enzyme interacts with chromatin modifier UBN1 in B-cells	
Figure 3.1 <i>In-silico</i> study on the interaction of UBN1 and AID	38
Figure 3.2 RMSD value of AID & UBN1 atoms	39
Figure 3.3 Energy components for binding of AID to UBN1	42
Figure 3.4 The protein-protein interaction profile between UBN1 and AID was constructed using Ligplot+	43
Figure 3.5 AID interaction with UBN1 in Raji cells	45
Figure 3.6 AID interaction with UBN1 in DT40 ψ V KO cells	46

Figure 3.7	Co-immunoprecipitation of UBN1 and AID in DT40 AID KO cells	47
Figure 3.8	AID interaction with ASF1a in Raji cells	48
Figure 3.9	AID interaction with ASF1a in DT40 ψ V KO cells	49
Figure 3.10	AID interaction with HIRA in Raji and DT40 ψ V KO cells	50
Figure 3.11	His-tagged pull-down assay for AID interaction with UBN1 in Raji cells and DT40 ψ V KO cells.	51
Figure 3.12	His-tagged pull-down assay for AID interaction with UBN1 in Raji cells and DT40 ψ V KO cells	53
Figure 3.13	Co-localization of AID with UBN1 in DT40 Ψ V KO cells	54
Figure 3.14	Merged zoomed image of AID and UBN1 co-localized in DT40 ψ V KO cells	55
Figure 3.15	Co-localization of AID with UBN1 in Raji cells	56
Figure 3.16	Merged zoomed image of AID and UBN1 co-localized in Raji cells.	57
Figure 3.17	Co-localization of AID with UBN1 in DT40 AID KO cells	59
Figure 3.18	Proximity ligation assay for AID and UBN1 in Raji cells	60
Figure 3.19	Proximity ligation assay for AID and UBN1 in DT40 Ψ V KO cells	61

Chapter 4: Heterologous expression of hAID in *E. coli* leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored

Figure 4.1	Expression and western blotting of hAID-His.	70
Figure 4.2	Peptide mapping and trypsin digestion map of hAID	71
Figure 4.3	Expression and purification of Nus-hAID-His from <i>E. coli</i> cells	73
Figure 4.4	Expression and purification of His-MBP-hAID	74
Figure 4.5	Western blot of Nus-hAID-His using antibody	76
Figure 4.6	Western blot of His-MBP-hAID	77
Figure 4.7	Identification of putative hAID splice site	78
Figure 4.8	Expression of splice mutant and PCR from cDNA	80
Figure 4.9	Peptide mapping of Nus-hAID-His in which upper and lower bands are aligned with hAID	82
Figure 4.10	Peptide mapping of His-MBP-hAID in which upper and lower bands are aligned with hAID	83
Figure 4.11	mRNA structure of hAID as predicted by RNA fold server	84

Chapter 5: SRSF1-3 interaction with TOP1 likely inhibits TOP1 DNA nicking activity

Figure 5.1	Domain organization of TOP1 and SRSF1-3	92
Figure 5.2	Structural dynamics conformation of TOP1 bound and unbound to SRSF1-3	94
Figure 5.3	Compactness and solvent accessibility analysis of TOP1 bound and unbound to SRSF1-3	96
Figure 5.4	The per-residual decomposition of binding free energy of the TOP1 with SRSF1-3 complex obtained from MMGBSA	97
Figure 5.5	Expression of His-SRSF1-3	99
Figure 5.6	Expression of His-SRSF1-3 Δ C	99
Figure 5.7	Purification of His-MBP-SRSF1-3 by using Ni-NTA beads	100
Figure 5.8	Removal of solubilization tag by treatment with TEV protease.	100
Figure 5.9	Investigation of consequences of SRSF1-3 on TOP1 plasmid relaxation activity.	101
Figure 5.10	Purification of solubilized His-SRSF1-3	102

LIST OF TABLES

Chapter3: Activation-induced cytidine deaminase an antibody diversification enzyme interacts with chromatin modifier UBN1 in B-cells

Table 3.1	Binding free energy components (KCAL/MOL) for the binding of UBN1-AID complex	41
------------------	---	----

Chapter 5: SRSF1-3 interaction with TOP1 likely inhibits TOP1 DNA nicking activity

Table 5.1	The different components of binding free energy obtained from MMPBSA for the TOP1 with SRSF1-3 complex	97
------------------	--	----

Appendix B

	Lists of primers	122
--	------------------	-----

Appendix C

	List of antibodies used for co-immunoprecipitation, western blotting and immunofluorescence	123
--	---	-----

ABBREVIATIONS

AID	activation-induced cytidine deaminase
AP	site abasic site
BER	base excision repair
cDNA	complementary DNA
CSR	class switch recombination
DAPI	4',6-diamidino-2-phenylindole
DZ	dark zone
LZ	light zone
FACT	facilitates chromatin transcription
HIRA	histone regulator A
UTR	untranslated region
TOP1	topoisomerase 1
UBN1	ubiquitin 1
UNG	uracil DNA glycosylase
RPA	replication protein A
SATB2	special AT-rich sequence binding protein 2
SHM	somatic hypermutation
SR proteins	serine and arginine-rich proteins
RNA polII	RNA polymerase II
PTBP2	polypyrimidine tract binding protein 2

NES

nuclear export signal

NLS

nuclear localization signal

Chapter 1

Chapter 1

Introduction and literature review

1.1 Immunoglobulin diversification

Adaptive immunity acts as a shield against invading pathogens and plays a pivotal role in the flourishing of human civilization on earth. A remarkable feature of the adaptive immune system is the production of millions of diverse antibodies, a phenomenon widely known as antibody diversity. Antibody diversification is restricted to the B-cells of the immune system. The B-cells diversify their antibody archive in antigen-independent as well as antigen-dependent manners [1]. The Variable (V), Diversity (D) and Joining (J) gene segments of immunoglobulin genes are rearranged during the development of B-cell and the process is known as VDJ recombination. VDJ recombination antibody diversification is antigen-independent and carried out by the recombination activating gene (RAG) recombinase [2]. The antigen-dependent antibody diversification is somatic hypermutation (SHM) and class switch recombination (CSR) in activated B-cells. SHM and CSR are mediated by a key genome mutator enzyme widely known as activation-induced cytidine deaminase (AID) encoded by the AICDA locus [3–5]. SHM is confined to the variable regions of the light and heavy chain, where AID-induced point mutations are unfaithfully repaired giving rise to antibodies, having either higher or lower affinity against an antigen, and subsequently, the higher affinity antibodies are selected in the process of clonal selection. CSR is a DNA deletion event taking place in the constant region of immunoglobulin heavy chain (IgH), leading to isotype switching.

1.2 Somatic hypermutation (SHM)

SHM takes place in the variable regions of immunoglobulin (Ig) light and heavy chains in B-cells, upon antigenic stimulation [6]. SHM introduces point mutation in the variable region of Ig genes which produces high as well as low affinity antibodies.

Further, B-cell expressing antibodies expand into plasma B-cell and memory B-cells, whereas B-cells expressing low affinity are cleared by apoptosis. SHM, together with clonal selection, is known as affinity maturation [7]. SHM takes place in the dark zone (DZ) of the germinal centre. Subsequently, SHM-generated antibody moves to the light zone (LZ) of the germinal centre where the clonal selection of high-affinity antibodies and CSR takes place. SHM is mediated by genome mutator enzyme AID, which induces a point mutation via the deamination of cytosine (C) into uracil (U), C to U conversion leads to the creation of a mismatch that can have various fates inside the B-cell [8]. If a U: G mismatch is unrepaired, and replication occurs then it leads to the creation of a transition mutation from C to thymine T [9]. Moreover, if Uracil DNA glycosylase (UNG) recognizes a U: guanine (G) mismatch, it removes the uracil base, leading to the creation of an abasic site (AP) [10]. In the context of Ig genes, the AP site is repaired by the recruitment of error-prone DNA polymerase resulting in a mutation. Additionally, if U: G mismatch is processed by base excision repair (BER) or mismatch repair system (MMR) it results in insertion, deletion as well as substitution mutation [11] (**Fig. 1.1 A**). Thus, AID-initiated mutations in Ig genes are unfaithfully repaired resulting in SHM.

1.3 Class switch recombination (CSR)

CSR occurs at the constant regions of IgH chains and is crucial for switching from IgM to IgG, IgE or IgA, having more effector function. Each IgH chain constant region gene is flanked by switch regions, except IgD, that are enriched in C residues [12]. AID creates a U: G mismatch at the switch region, which recruits BER and MMR. In the case of BER, UNG removes U from U: G mismatch that results in an abasic site. Subsequently, AP endonuclease cleaves the abasic site to create a single-strand break (SSB), and if another SSB is present in the vicinity, it results in a double-stranded break (DSB) (**Fig. 1.1 B**). Additionally, MSH2-MSH6 proteins of the MMR single-strand break (SSB), and if another SSB is present in the vicinity, it results in a double-stranded break (DSB) [10,13,14] (**Fig. 1.1B**).

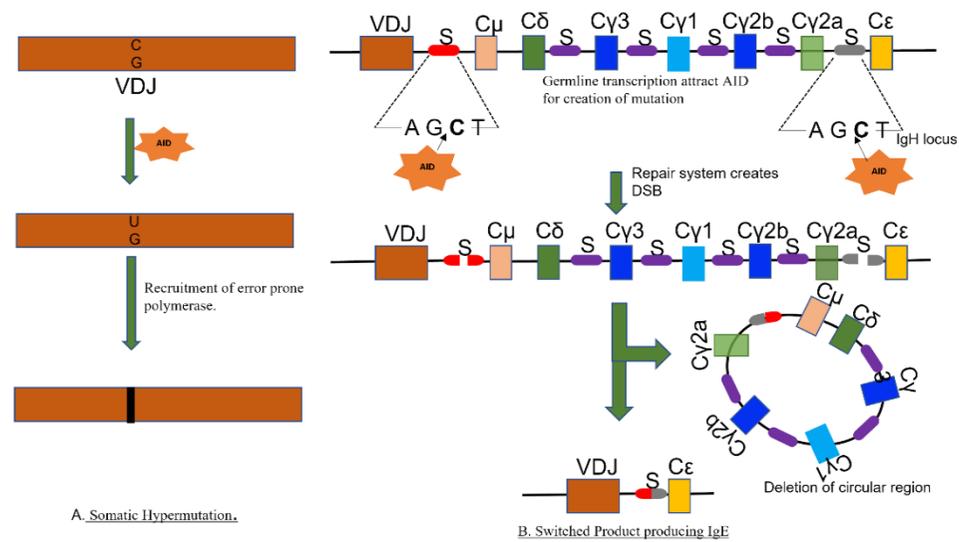


Fig. 1.1 Molecular mechanism of SHM and CSR. (A) Somatic Hypermutation (SHM): AID deaminates cytosine into uracil in the variable region of Ig genes. U:G mismatch recruits error-prone DNA polymerase that repairs U unfaithfully, creating a mutation. (B) Class Switch Recombination (CSR): The constant regions of IgH genes are flanked by the switch regions that are deaminated by AID. Repair pathway in the course of action creates double-stranded breaks, which lead to the deletion of intervening sequence, results in the class switching of antibody.

Additionally, MSH2-MSH6 proteins of the MMR pathway identify the U: G mismatch and recruit a complex of mutL homolog 1 (MLH1)/ PMS1 homolog 2 (PMS2)/ Exonuclease 1 (EXO1) [13]. Further, PMS2 creates SSB in the proximity of U: G mismatch, and EXO1 converts SSB into single strand gaps by its exonuclease activity. Should the PMS2 generate SSB on another strand of DNA, then EXO1 activity leads to DSB necessary for CSR [15–17]. DSBs of the switch regions are resolved by non-homologous end joining (NHEJ) [18]. Ku70/Ku86 protein binds to DSB, leading to the activation of DNA-dependent protein kinase (DNA-PKCs). Eventually, XRCC4/XRCC4-like factor/Ligase IV (Lig4) is recruited, which joins the nonhomologous end and leads to NHEJ. This NHEJ is responsible for joining two switch regions, and deletion of intervening sequence results in isotype switching of antibodies [19,20].

1.4 AID and its splice isoforms

Until the mid-90s, it was unclear how a limited set of Ig genes produced such a large number of antibodies. The breakthrough came when AID was discovered in 1999 by the Honjo group via the subtractive hybridization technique, a method used to detect differences between the RNA/cDNA of two cells [4]. It led to the quest for finding the AID structure, its function, as well as targeting. AID was originally thought to be an RNA editing enzyme due to its similarity with the APOBEC family [21]. AID consists of four distinct parts, namely, nuclear localization signal (NLS) sequence, AID catalytic domain, APOBEC-like domain, and nuclear export signal sequence. The N-terminal and C-terminal domains of AID are indispensable for SHM and CSR, respectively [22]. Besides AID-FL (full-length hAID), other splice isoforms of hAID are also reported in B-cells. As AID-FL is composed of five exons, their reported splice isoforms are namely AID δ E4a (a 30 bp from the start of exon 4 is missing), AID δ E4 (exon4 is completely absent), AID δ E3E4 (both exons 3 and 4 are absent) and AID-ivs3 (retention of intron 3) (**Fig. 1.2**).

Interestingly, in B-cell malignancies along with AID-FL, all other splice variants of AID such as AID δ E4a, AID δ E4, AID δ E3E4, and AID-ivs3 are reported. Furthermore, the splice variants of AID display a wide spectrum of activities, including inactivation of CSR (AID δ E4a and AID δ E4), hyper SHM (AID δ E4a and AID δ E4), inadequate in SHM (AID δ E3E4) [23,24]. Normal B-cell suppresses the generation of an alternative splice variant of AID, indicating a crucial regulation that limits the AID mutagenic activity [25].

In vitro and cell-free studies have shown that AID preferentially deaminates cytosine in the WRC (W = A/T and R = A/G) region, also known as a hot spot, compared to a cold spot in SYC (S = G/C and Y = C/T) [26–28]. Furthermore, AID preference for WRC inside the cell at Ig genes is also reported [29]. The AID-mediated mutation frequency at hotspots is 4.6 times higher than at a cold spot [28]. Nevertheless, why AID preferentially deaminates C in WRC hotspots is still to be explored.

1.5 Regulation of AID

Physiologically, AID activity is limited to Ig genes but its off-target activity on non-Ig genes, including some oncogenes, results in deregulated expression of these genes which is the leading cause of B-cell lymphomas [30]. AID being the mutator enzyme, its expression and activity are precisely regulated at the transcriptional, post-transcriptional, translational, and post-translational levels. In mice, the AICDA genomic locus is enriched in binding sites for numerous transcription factors, such as NF- κ B, STAT6, C/EBP, Smad3/4, HoxC4, Pax5, and E protein, which positively regulate the expression of AID upon B-cell activation [31–35]. On the contrary, ID proteins are the negative regulator of AID expression as this protein inhibits the binding of transcription activators Pax5 and E47 to the AID locus [31]. At mRNA levels, AID is regulated by microRNAs; for example, miR-155 binds to the 3' UTR of AID and represses its translation. Thus, SHM and CSR are reduced by the expression of miR-155 [36–38].

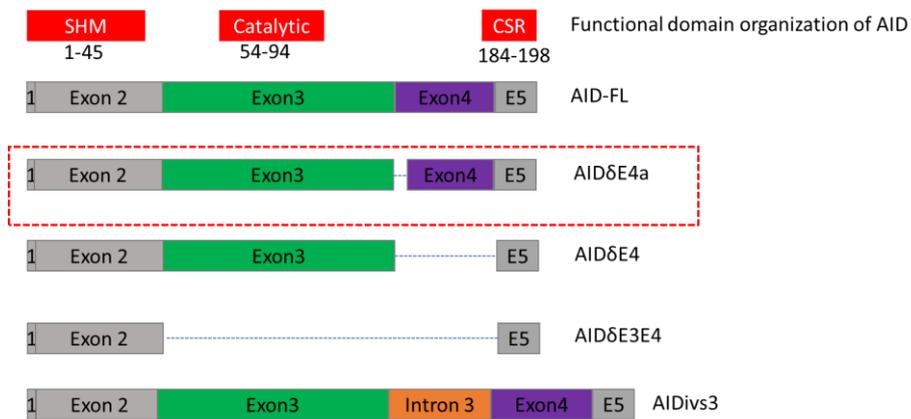


Fig. 1.2 The functional domain organization of AID & its splice isoforms. The splice isoform of AID along with its exons reported in various types of cancers.

As expected, miR-155 acts as a tumour suppressor, as the activated B-cell deficient in miR155 showed a higher level of translocation compared to the wild type [39]. Similarly, miR-181b was reported to bind to multiple repetitive sequences in the 3' UTR of AID, leading to diminished CSR. Thus, miR-181b can also limit the activity of AID and inhibit the oncogenic progression of B-cell [40]. Interestingly, when AID was expressed at low levels, miR-93 and miR-155 led to the inhibition of AID translation, thereby maintaining a low level of AID-mediated mutation leading to genomic stability. On the other hand, the disruption of miR-93 and miR-155 results in a high AID level and thereby drives oncogenesis [41]. In the cytosol, AID is anchored with eEF1A1, which controls its movement inside the nucleus [42]. The half-life of AID inside the nucleus is at least three times shorter, as it is readily poly-ubiquitinated and degraded [43]. Furthermore, the C-terminal NES signal moves out of AID from the nucleus, thus maintaining a low AID level in the nucleus [44–46].

1.6 Role of transcription in AID-mediated antibody diversity

Betz et al., for the first time, showed that the intronic enhancer such as κ 3' enhancer that assists in sustaining the high levels of mRNA is critically required for SHM [47]. Furthermore, researchers have established that transcription initiation is directly linked with SHM [48]. Transcription generates negative supercoiling of DNA that provides single-strand patches of DNA for AID action. Topoisomerase 1 (TOP1) relieves negative supercoiling and therefore results in removing single-strand DNA patches, and thus limiting AID mutator activity [49]. Further, AID activity in the light of TOP1 was studied and this study revealed that AID activity is conversely related to TOP1. Moreover, transcription elongation and pausing are crucial for AID action and were seen in the variable region of Ig genes in DT40 cells. The AID has better accessibility to DNA during transcription elongation rather than termination. Further, PolIII pausing near the termination site exposes the single-stranded DNA and the persistence of negative supercoiling assists in AID-mediated SHM [50] Interestingly, a convergent transcription

that occurs at intragenic super-enhancers assist in targeting AID. Notably, in convergent transcription, besides the normal sense RNA transcription, the small amount of antisense RNA transcription also arises from intragenic super-enhancers, which leads to the encounter of the RNA PolIII complexes transcribing on opposite strands to stall. Subsequently, stalled RNA PolIII complexes on both strands are a suitable target for AID [51]. Moreover, convergent transcription and super-enhancers have emerged as the key features of most non-Ig genes mistargeted by AID. Qian et al. demonstrated that super-enhancers and regulatory clusters present in human and mouse B-cells are crucial for AID targeting [52]. Additionally, it was illustrated that divergent transcription of antisense noncoding RNA assists in the targeting of AID. Divergent transcription favours transcription stalling and RNA loop formation, which in turn recruits AID to the target gene [53].

1.7 Effect of a nucleosome on AID activity

Transcription progression in higher eukaryotes by RNA polymerases routinely encounters a roadblock known as nucleosome, which impedes as well as limits the accessibility of transcriptional machinery to the DNA. Nucleosomes are the structure formed when DNA is bound to the histone octamers (consisting of H2A, H2B, H3 as H4) that assist in packaging large DNA inside the small, confined space of the nucleus [54]. Further, the nucleosome assembly depends on the DNA sequences as some sequences are strong for nucleosome assembly, whereas others are weak nucleosome positioning sequences. Interestingly, Kodgire et al., unravel the impact of nucleosome positioning sequence on AID-mediated SHM [55]. Further, it was demonstrated that AID-induced somatic mutations are enriched in the variable regions of Ig genes of DT40 ψ V KO cells containing weak nucleosome sequences in contrast to strong nucleosome sequences. Thus, strong nucleosome placed at Ig genes restricts the AID-accessibility to Ig genes. Nevertheless, immune cells recruit various chromatin modifiers to overcome the nucleosome barrier [56,57].

1.8 Role of chromatin modifiers or histone chaperon in antibody diversity

FACT and HIRA chaperon complexes are chromatin remodelers and are involved in AID-mediated antibody diversity [56,57]. FACT complexes have a role in DNA repair, replication, and transcription [58,59]. FACT complex in humans is composed of hSpt16 and SSRP1. FACT mediates its function by binding to H2A-H2B histone in the nucleosome, which in turn results in the reconstitution of the nucleosome and smooth progression of RNA polymerase. Interestingly, the knockdown of the FACT complexes severely impaired SHM as well as CSR without affecting the level of transcription [56].

HIRA chaperon complex is responsible for the deposition of histone variant H3.3 in a replication-independent manner that is responsible for gene activation as well as silencing [60,61]. HIRA chaperon complex consists of UBN1, HIRA, ASF1a, and CABIN1. UBN1 plays a key role in the deposition of H3.3 at gene regulatory elements in ES cells of the mouse [62]. UBN1 domains comprised of N-terminal Hpc2-related domain (NHRD) domain, HRD domain, and middle domain. UBN1 interacts with HIRA through the NHRD domain and H3.3 through the HRD domain. Recently, it was demonstrated that HIRA knockdown dramatically reduced SHM in the Ramos cell line along with altered mutation patterns in the Ig variable regions [57].

1.9 Role of AID interacting partner in antibody diversity

Besides transcription and chromatin modifiers, AID requires numerous interacting partners for efficient targeting of Ig locus. Remarkably, the last two-decade studies have found numerous AID-associated proteins such as RPA, CTNNBL1, PTBP2, and SRSF1-3 [63–66]. RPA interaction with AID assists the latter in binding to single-stranded DNA [63]. CTNNBL1 interaction with AID is crucial for AID accumulation inside the nucleus. Interestingly, the disruption of the interaction between CTNNBL1 and AID is reported in clinical conditions with common variable immunodeficiency associated with

autoimmune cytopenias that showed diminished SHM [67]. PTBP2 interaction with AID affects the occupancy of the AID to the switch region of the Ig heavy chain as knock down of PTBP2 impairs CSR due to reduced occupancy of AID to the switch region [65].

1.10 Role of SRSF1-3 in antibody diversity

SRSF1-3 is the splice isoform of SRSF1. SRSF1 belongs to the family of serine/arginine-rich protein that plays a pivotal role in pre-mRNA splicing and regulates alternative splicing, as well as RNA metabolism [68,69]. In humans, there are nine different SR proteins each containing at least one N-terminal RNA binding domain and C-terminal RS domain. SRSF1-3 protein consists of two RNA binding domains, namely RRM1, RRM2, and the C-terminal domain [70]. Interestingly, Kanehiro et al., were the first to identify that DT40 cells lacking SRSF1-3, showed impaired SHM, whereas reconstitution with SRSF1-3 restored SHM [71]. Furthermore, SRSF1-3 was reported to be indispensable for SHM but the molecular mechanism of its action was unclear. Recently, RNA-seq analysis of SRSF1-3 reconstituted DT40 ASF cells showed the up-regulation of numerous genes such as SATB2 and UBN1 in SRSF1-3 [72]. SATB2 is an AT-rich sequence-binding protein that binds to the MAR region of Ig genes in turn enhances transcription [73]. Thus, SRSF1-3 enhances the expression of UBN1 that assists in the recruitment of the histone variant H3.3 to the promoter of IgL, whereas SATB2 binds to the MAR region of Ig genes, thereby indirectly regulating Ig transcription and thus SHM [72]. Besides indirectly regulating the expression of the IgL gene, SRSF1-3 is also interacting with several proteins such as TOP1 [66]. Interestingly, in the context of antibody diversity TOP1 activity is directly linked to SHM/CSR [74,75].

1.11 Background of work and objective

AID was reported to interact with RPA protein [63] and HIRA protein a member of the HIRA chaperon complex was found to interact

with RPA [76]. Interestingly, HIRA protein is also required for the deposition of H3.3 at Ig genes and it is indispensable for SHM as HIRA KO severely impaired somatic mutation [57]. However, UBN1 a member HIRA chaperon complex interaction with AID was not known. In this study, we explored the interaction of AID with UBN1 by *in-silico* as well as *in-vitro* methods, which may play a potential role in antibody diversification (the role in SHM or CSR still needs to be explored).

AID expressed in human cells are mostly FL-hAID but its splice forms are also reported in various lymphoma. Further, AID is purified from SF9 cells and *E. coli* cells for routine biophysical studies, kinetics studies, and structural studies. Surprisingly, hAID expressed in *E. coli* showed two induced bands instead of a single band. In this study, we want to examine the heterologous expression of hAID in *E. coli* cells that produce two induced bands.

SRSF1, a splicing regulator, interacts with TOP1, which inhibits the activity of the latter [77]. Moreover, TOP1 inhibition promotes AID-mediated antibody diversity. Interestingly, SRSF1-3, a splice isoform of SRSF1-3, was also found to interact with TOP1 [66]. Nevertheless, SRSF1-3 inhibition of TOP1 activity is not reported. In this study, we want to investigate the influence of SRSF1-3 on TOP1 activity.

The main objectives of my studies are mentioned below which are covered over three chapters.

1. To explore whether UBN1 interacts with antibody diversification enzyme AID that may have a role in SHM or CSR.
2. To explore why the heterologous expression of hAID in *E. coli* cells results in the expression of two bands.
3. To explore the fate of SRSF1-3 interaction with TOP1 on the DNA nicking activity of the latter.

Chapter 2

Chapter 2

Materials, Methods, and Instrumentation

2.1 Materials

2.1.1 Cell lines

DT40 ψ V knockout cells were a kind gift by H. Arakawa and J.M. Buerstedde (Institute of Molecular Radiology, Neuherberg, Germany). DT40 AID KO cell lines were a gift from Storb (University of Chicago, USA). Raji cells were procured from the National Centre for Cell Sciences, Pune, India.

2.1.2 Chemicals

The chemicals used in this research work are as follows: Triton X-100, Nonidet P-40 (tergitol), Tween-20, Chloroform, Isoamyl alcohol, Fetal Bovine Serum (FBS), Chicken serum, 2-mercaptoethanol, Penicillin-streptomycin-L-glutamine, RPMI-1640, Ampicillin, Chloramphenicol, Isopropyl- β -D-thiogalactopyranoside, Bovine serum albumin (BSA), Lysozyme, Glycerol, Luria-Bertani agar, Luria-Bertani broth, ethylenediaminetetraacetic acid (EDTA) tetrasodium salt dihydrate, Calcium Chloride, Methanol, Sodium dodecyl sulfate, Glacial acetic acid, Coomassie Brilliant Blue G-250 dye, Paraformaldehyde (PFA), Skim Milk Powder, Sodium Chloride (NaCl), Urea, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), Hydrochloric acid (HCl), Sodium hydroxide (NaOH), Tris base, Acrylamide, Bisacrylamide, Ammonium Persulfate (APS), N, N, N'-tetramethylene-1-2-diamine (TEMED), Dimethyl sulfoxide (DMSO), Bromophenol Blue, Xylene cyanol, Protease Inhibitor Cocktail, Phenylmethanesulphonyl fluoride (PMSF), Benzamidine Hydrochloride, RNase A, Proteinase K. All chemicals mentioned above belongs to molecular biology grade and procured from Sigma Aldrich Chemical Pvt, Invitrogen Pvt. Ltd, MP Biomedical USA, HiMedia Pvt, Alpha aesar Ltd, Sisco Research Lab Pvt.

2.1.3 Reagents and Kits

The plasmid isolation kit and PCR/Gel extraction Kit were procured from Thermo Fisher Scientific, USA. The human topoisomerase kit 1 assay kit was procured from TopoGEN, USA.

T4 DNA ligase, dNTPs, *Taq* and *Pfu* DNA polymerase, *Nde*I, *Xho*I, *Spe*I, *Bam*HI, *Hind*III, *Pst*I, *TEV* protease, 1kb DNA ladder procured from New England Biolabs, USA.

2.2 Methods and Instrumentation

2.2.1 *E. coli* Competent cell preparation

Competent cells of *E. coli* (DH5 α for cloning and Rosetta for protein expression) were prepared using the calcium chloride method. *E. coli* cells were streaked on the LB agar plate (overnight) followed by inoculation of a single colony into LB broth. Subsequently, overnight grown culture was inoculated at 1% in 50 ml LB broth and allowed to grow till the O.D. reached 0.5 and cells were harvested via centrifuged. Cells were kept for 45 min and 4 hrs in 0.1 M CaCl₂ and stored at -80°C with the addition of 15% glycerol.

2.2.2 Transformation of plasmid in *E. coli* competent cells

E. coli strain DH5 α was transformed with a plasmid vector. For transformation, the desired amount of plasmid was added to competent cells and kept on ice for 30 min. Subsequently, cells were given heat shock (42°C for 90 sec and 4°C for 120 sec) followed by the addition of 800 μ l of LB broth and kept at 37°C for 90 min. Finally, cells were spread on the LB agar plate supplemented with a suitable antibiotic.

2.2.3 Isolation of plasmid DNA

Isolation of plasmid DNA was performed as mentioned in the manufacture protocol using a plasmid DNA isolation kit (Thermo Scientific). For plasmid DNA extraction, a single colony was inoculated into 5 ml of LB broth supplemented with suitable antibiotics (depending on antibiotic resistance mentioned in the vector) and kept at 37°C for

12-16 hrs. Subsequently, cells were pelleted by centrifugation and sequentially resuspended in resuspension buffer, incubated in lysis buffer for 5 min neutralization buffer, and centrifuged at maximum speed for 10 min. The aqueous phase was transferred without disturbing the white pellet to the mini-spin column, centrifuged, followed by washing with wash buffer and eluted in 50 µl of elution buffer.

2.2.4 Polymerase chain reaction

The gene sequences were amplified using either genomic DNA or cDNA as a template by PCR. A typical PCR reaction of 50 µl contains reaction buffer (10 mM Tris-Cl, 50 mM KCl, 1.5 mM MgCl₂, pH 8.3 at 25°C), 200 µM of each dNTPs (HiMedia), 1 µM of each primer (forward and reverse), *Taq* or *Pfu* DNA polymerase. The PCR cycle for amplification was as follows: 95°C for 5 min for initial denaturation, followed by 30 cycles of 95°C for 30 sec, annealing for 30 sec (based on the melting temperature of primers), extension at 72°C (time is based on the length of the PCR product and polymerization speed of the polymerase) and final extension at 72°C for 10 min.

2.2.5 Agarose gel electrophoresis

The genes amplified from PCR and colony PCR for clones were confirmed by running on 0.8% agarose (supplemented with 100 µg/ml ethidium bromide) in 0.5X TAE buffer. Samples were loaded in the agarose gel with 6X DNA loading dye, and a standard 1 Kb DNA ladder was used for size confirmation. Finally, the gels were visualized in a UV fluorescence channel, and images were captured by Image Quant (GE Healthcare).

2.2.6 Gel Extraction

For cloning, the vector and insert were gel extracted using a gel extraction kit (Thermo Scientific). The vector and insert were run on 0.7% agarose gel and allowed to separate the band followed by excision of the desired portion of gel and eluting as per manufacture protocol. The eluted DNA fragments were again confirmed by running on 0.8%

agarose gel and stored at -20°C.

2.2.7 PCR Purification

Before proceeding to the restriction digestion, PCR products were purified according to the manufacturer protocol (Thermo Scientific). Finally, the PCR products were eluted in 50 µl volume, confirmed on agarose gel, and stored at -20°C till required.

2.2.8 Restriction digestion

The desired amount of plasmid DNA (Vector) or PCR products (Insert) were added to a 1.5 ml microcentrifuge tube, followed by the addition of sterile distilled water (if required to make up the final volume), restriction enzyme buffer (generally cut smart buffer to 1X) and 1 unit of enzyme per microgram of DNA. The restriction digestion mixture was incubated at a temperature recommended by the manufacturer for 2 hrs and digestion of plasmid DNA was confirmed by running agarose gel electrophoresis. Lastly, the vector was gel eluted and the insert was PCR purified.

2.2.9 Ligation

In a ligation reaction, the ratio of vector to insert is 1:3. A typical 10 µl ligation reaction consists of vector, insert, 1 µl of 10X ligation buffer and 1 µl of T4 DNA ligase volume making up to 10 µl and kept at 22°C for 2 hrs. Subsequently, 5 µl of the ligation mixture was transformed into competent cells.

2.2.10 B-Cell culture

Raji cells were cultured in RPMI 1640 medium containing 10% FBS, 2 mM glutamine, 100 µg/ml penicillin G, 50 µg/ml streptomycin and 50 µM β-mercaptoethanol at 37°C and 5% CO₂. DT40ψV knock-out cells were cultured in media like Raji cells and supplemented with 1% chicken serum grown at 39.5°C, and 5% CO₂.

2.2.11 Structural modelling of UBN1 and docking analysis

Homology modelling of UBN1 was performed using the webser

-ver I-TASSER [78,79] to obtain the 3D structure from the amino acid sequence retrieved from Uniport ID (Q9NPG3). Models were generated using the threading approach, and the best model was selected based on the C score. For AID, we downloaded the already predicted structure from the Alpha fold database (AF-Q9GZX7-F1) [80]. The UBN1-AID complex was formed by the docking protocol using the Rosetta online server [81]. Further, the docked complex was selected according to the docking score and further subjected to molecular dynamics (MD) simulation.

2.2.12 Molecular Dynamics Simulation of UBN1-AID Complex

Herein, the UBN1-AID complex was subjected to 150 ns MD simulation using the *pmemd.cuda* module of the AMBER18 suite [82]. Amber ff14SB forcefield was used to simulate the protein-protein complex [83]. The complex was neutralized with an appropriate number of ions and solvated into an octahedron water box with TIP3P water molecules [84]. SHAKE algorithm was used to restrain the bonds involving hydrogen atoms to keep the timestep of simulation 2 fs [85]. The long-range interaction was estimated using the particle mesh Ewald (PME) method with a non-bonded cut-off of 10 Å [86]. The temperature was kept constant at 300 K using a Langevin thermostat with a collision frequency of 2 ps⁻¹ [87]. Similarly, the pressure of the system was also controlled using Berendsen's Barostat. Complexes were subjected to two-step minimizations followed by a stepwise heating phase and equilibrium simulation before the final production run. We performed a 100 ns long equilibrium simulation to remove any structural clashes and stabilize the protein-protein complex. A detailed description of the simulation protocol was already discussed in our previous works [88]. Finally, we performed a 50 ns production run under the NPT ensemble. Trajectory analysis was conducted using the *cpptraj* module of AmberTools [89].

2.2.13 Binding free energy calculations using MM/PBSA scheme

The binding free energy was calculated using the molecular me-

chanics/Poisson Boltzmann surface area (MM/PBSA) scheme. The following equations were employed to estimate the binding free energy [90].

$$\Delta G_{binding} = G_{complex} - (G_{protein1} + G_{protein2}) \quad (1)$$

where the energy term further subdivided into

$$G = E_{vdW} + E_{ele} + G_{pol} + G_{np} \quad (2)$$

E_{vdW} and E_{ele} represent the van der Waals and electrostatic interaction energies. The solvation term, G_{pol} , indicates the electrostatic solvation free energy estimated using the Poisson Boltzmann equation and G_{np} represents the non-polar solvation. Configurational entropy was avoided in our current study as we are interested only in the relative binding free energy and due to the higher computational cost. We have used 2500 configurations obtained from the last 50 ns trajectory for the MM/PBSA calculation. Residue-wise contribution of binding free energy was also estimated using the molecular mechanics Born generalized surface area (MM/GBSA) scheme [91].

2.2.14 Co-Immunoprecipitation Assay

For CoIP from whole cell extract, 2×10^7 cells were washed three times with 1X PBS. Cells were lysed in a buffer containing 50 mM Tris-Cl, pH-8.0, 150 mM NaCl, 1.0% NP-40, and Protease inhibitor + 1 μ g/ml DNaseI. Further, cell lysate was passed through a 22-gauge needle multiple times to shear the DNA.

Lysate supernatants were incubated with either anti-AID mAB or anti-UBN1 pAB antibody (1:100 dilutions) (Appendix C) overnight at 4°C on a rota-spin. Afterwards, Protein-A agarose beads were washed and equilibrated in 1 ml of equilibration buffer (20 mM Tris-Cl pH-8.0, 5 mM MgCl₂, 5 mM MnCl₂, 150 mM NaCl, and 1.0% NP-40). Samples were incubated with a 20 μ l slurry of protein-A on a rota-spin for 2 hrs at 4°C and pelleted by centrifugation (2000g, 3min). Subsequently, pellets were washed three times in 1 ml of ice-cold washing buffer (20

mM Tris-Cl pH-8.0, 5 mM MgCl₂, 5 mM MnCl₂, 750 mM NaCl and 1.0% NP-40) with 5 min incubations between spins. Finally, the immunoprecipitated proteins were eluted with 100 µl of antibody strip buffer/elution buffer (120 mM Tris-Cl, pH 6.8, 1.0% SDS, at room temperature for 30 min on a shaker to elute the proteins) and boiled the samples at 95°C for 5 min.

2.2.15 Western immunoblotting

Western immunoblotting was performed using a mini protean tetra cell transfer unit (Bio-Rad) that results in the transfer of protein from gel to the nitrocellulose membrane (Amersham, GE Healthcare) for 12 hrs. Subsequently, blots were blocked in 20 mM Tris-Cl pH 7.5, 150 mM NaCl, 0.02% Tween 20 (Alfa Aesar) and 5% skimmed milk powder (Himedia). Western blotting was performed using antibodies (Appendix C), and blots were developed via chemiluminescence substrate ECL substrate (Bio-Rad). Lastly, chemiluminescence was detected by the Image Quant LAS 4000 Gel Doc system (GE Healthcare).

2.2.16 Double immunofluorescence microscopy

For co-localization studies, 5 X 10⁶ DT40 KO and Raji cells were washed 3 times in 1X PBS at 200 rcf, 5 min at 4°C. Subsequently, cells were fixed in 4% paraformaldehyde in 1X PBS for 20 mins. Cells were permeabilized in 0.1X triton X in 1X PBS for 20 min and blocked in 5% BSA in PBS-T for 30 min. Further, cells were incubated with anti-rabbit UBN1 and anti-mouse AID (1:100) overnight at 4°C in 1X PBS containing 2.5% BSA. Cells were washed 3 times with 1X PBS-T followed by incubation with anti-rabbit Alexa flour 594 and anti-mouse Alexa flour 488 (1:500) secondary antibodies in 2.5 % BSA for 60 min. Further, cells were washed 3 times and stained with DAPI for 15 min. Finally, cells were resuspended in 1X PBS and mounted on the slide, kept for drying at 37°C, and high-resolution images were captured via confocal microscopy (Olympus FV3000).

2.2.17 Cloning & Expression of various proteins used in pull-down

UBN1¹⁻¹⁷⁵ and hAID were PCR amplified from the cDNA of Raji cells using primers PK934 & PK935 and PK774 & PK910 (Appendix-B) and cloned at *Bam*HI/*Hind*III in His-MBP vector to create His-MBP-Flag-UBN1¹⁻¹⁷⁵ and His-MBP-hAID, respectively. Similarly, to create clones without His-tag, we removed His-MBP from His-MBP-Flag-UBN1 and His-MBP-hAID by restriction digestion with *Nde*I and *Nhe*I, and cloned only MBP (amplified using primers PK957 and PK960, (Appendix-B) at *Nde*I/*Nhe*I site to give rise to MBP-Flag-UBN1¹⁻¹⁷⁵ and MBP-hAID without His-tag, respectively. Subsequently, each protein was expressed in the Rosetta strain of *E. coli*. Proteins were induced by 0.5 mM IPTG at kept at 16 °C/18 hrs/180 rpm. His-tagged proteins were purified either on the HisTrap (Cytiva) column, whereas, proteins without His-tag were purified on the MBP-trap (Cytiva) column, according to the manufacturer's protocol. An empty vector was used to express His-MBP only.

2.2.18 Pull down Assay

His-MBP-UBN1¹⁻¹⁷⁵, His-MBP-AID, and His-MBP were expressed in *E. coli* and purified using standard Ni-NTA beads using manufacturer guidelines. Subsequently, all three proteins were buffer exchanged against 200 mM NaCl, 25 mM Tris-Cl pH 8.0, 10% glycerol, and 0.5 mM TCEP. Further, equimolar of His-MBP-UBN1¹⁻¹⁷⁵, His-MBP-AID, and His-MBP were separately incubated with 1.5 mg whole cell lysate of Raji or DT40ψV KO (lysis buffer contains 1X PBS, 1% NP-40, protease inhibitor cocktail and 1 mM beta-mercaptoethanol) in 1XPBS with mild agitation for 6 h/4 °C for binding. Further, 80 µl of pretreated Ni-NTA beads were added to each tube and incubated for 2 h/4 °C. Subsequently, beads were washed thrice with 1 ml wash buffer (200 mM NaCl, 10 mM Tris-Cl pH 8.0, 50 mM imidazole, 1 mM beta-mercaptoethanol) by centrifugation at 800 g. Protein-containing Ni-NTA beads were eluted in a wash buffer supplemented with 500 mM imidazole. Finally, samples were subjected to SDS-PAGE followed by

Western blotting and analyzed either by anti-AID mAb or anti-UBN1 pAb.

For *in-vitro* His-MBP-hAID pull-down experiments, a new clone MBP-Flag-UBN1¹⁻¹⁷⁵ (without His-tag) was created. Equimolar concentrations of His-MBP-hAID and His-MBP were mixed with MBP-Flag-UBN1¹⁻¹⁷⁵, separately, in the binding buffer (200 mM NaCl, 10 mM Tris-Cl pH 8.0, 10% glycerol, 0.1 Tween 20, 1 mM β ME supplemented with 20 mM imidazole), kept for 6 h/4 °C mild agitation. Subsequently, 40 μ l Ni-NTA beads were added in each tube and kept for 2 h/4 °C, washed thrice with binding buffer supplemented with 50 mM imidazole and eluted with binding buffer supplemented with 500 mM imidazole, detected with anti-flag antibody. Further, for pull-down studies with His-MBP-Flag-UBN1¹⁻¹⁷⁵, a new clone of MBP-hAID (without His-tag) was created. Similarly, in His-MBPFlag-UBN1¹⁻¹⁷⁵ pull-down experiments, equimolar quantities of His-MBP-Flag-UBN1¹⁻¹⁷⁵ and His-MBP were mixed with MBP-hAID and proceeded as mentioned for His-MBP-hAID pull-down protocol mentioned above. Finally, western blotting was performed using anti-AID antibody.

2.2.19 Proximity Ligation Assay

PLA was performed using reagents supplied in Duolink in situ Red PLA kit Mouse/Rabbit (Sigma-Aldrich, DUO92101) (Appendix C), following the manufacturer's instructions. The protocol was first standardized to use a minimal volume of the probes and enzymes required for the reaction. Briefly, 3×10^6 DT40 cells were taken per well of a 96-well U-bottom plate, and the protocol carried out for the co-localization studies, described previously, was followed. After the primary antibody's incubation, overnight at 4°C, cells were washed with wash buffer A, twice at 220 RCF, 4°C for 5 min each, followed by incubation with secondary antibodies conjugated to oligonucleotides (PLA Probes anti-mouse MINUS and PLA Probes anti-rabbit PLUS) (Appendix C), diluted in antibody diluent (1:5) and incubated for 60 min at 37°C in a humidity chamber. Cells were then washed twice with A

(washing buffer A), followed by centrifugation at 220 rcf at 4°C for 5 min. The 5X ligation buffer was diluted in high-purity water to a final concentration of 1X and the ligase (1U/μl) was diluted at 1:40 in the 1X ligation buffer. 20 μl of ligation mixture was added to each well and incubated for 30 min at 37°C in the humidity chamber. Again, the cells were washed twice with washing buffer A, followed by centrifugation at 220 rcf at 4°C for 5 min. Further, 20 μl of amplification polymerase mixture (0.5 μl polymerase in 39.5 μl 1X amplification buffer, 1:80 dilution) was added to each well and incubated for 120 min at 37°C in the humidity chamber. Cells were then washed twice with washing buffer-B at 220 rcf, 4°C for 10 min and the final wash was with 0.01X washing buffer-B for 1 min. DAPI with mounting medium was added to the wells and the cells were mounted onto the slides. Slides were left to dry overnight at 4°C, scanned and images were acquired using a confocal laser scanning microscope (Olympus, FV1200MPE).

2.2.20 Cloning of hAID-His, Nus-hAID-His and His-MBP-hAID

The human AID gene was amplified using PK515 and PK516, PK510 and PK516, PK774 and PK775 (Appendix B) from Raji cDNA as a template. AID gene products were digested with *Nde*I and *Xho*I, *Spe*I and *Xho*I, *Bam*HI and *Hind*III, respectively, and cloned at the appropriate sites to give rise to AID-His, Nus-hAID-His, and His-MBP-hAID.

2.2.21 Expression of hAID-His

hAID-His protein expression was induced at 0.5 mM IPTG at 16°C for 18 hrs. Cells were pelleted washed with 1XPBS, resuspended in lysis buffer (50 mM Tris-Cl pH 8.0, 15 mM NaCl, 10 mM ZnCl₂, 10% glycerol, 0.1% tween-20, 5 mM b-mercaptoethanol, 1 mM PMSF, 1 mM benzamidine hydrochloride, 0.5 mg/ml lysozyme and protease inhibitor cocktail) and kept at 4°C for half an hour. Subsequently, the cells were lysed via sonication and centrifuged to obtain clear lysate and loaded on 12% SDS-PAGE.

2.2.22 Mass spectrometry sample preparation of hAID-His, Nus-hAID-His and His-MBP-hAID

The hAID-His protein was run on the SDS-PAGE till the upper band and lower band were completely resolved. Subsequently, the gel was stained overnight, followed by destaining, and upper and lower bands were cut precisely. The gel was excised and stored in nuclease-free water frozen at -80°C degrees and shipped on dry ice to CCAMP (India) for peptide mapping. Proteins were isolated and digested with trypsin to generate the peptides that were separated and analyzed by LC-MS. The peptide mixture was separated on an Easy spray PepMap RSLC C18 column (2µm; 50cm x 75µm; 100Å). Solvent A (0.1% formic acid in HPLC grade water) and solvent B (80% acetonitrile + 0.1% formic acid) were used. A linear gradient of 5% to 95% solvent B in 60 min at a flow rate of 300 µL/min was applied. The peptide sequences of proteins were recognized by searching spectra in the Swiss-Prot database using MASCOT.

The Nus-hAID-His and His-MBP-hAID proteins were run onto 8% SDS-PAGE and cut precisely. The sample preparation and peptide running on MS/MS were performed as mentioned [92]. The raw files obtained were analyzed by proteome discoverer software 2.2.

2.2.23 cDNA preparation and quantitative PCR

The Nus-hAID-His containing *E. coli* was induced as mentioned above. The pellet was harvested by centrifugation at 14000 rpm for 5 min and resuspended in 1 ml of Trizol reagent and kept at room temperature for 5 min and 200 µL of chloroform was added. Following incubation for 3 min, the sample was centrifuged at 14000 rpm for 15 min. The aqueous phase formed was transferred to another tube containing 0.5 ml isopropanol and centrifuged for 5 min. Subsequently, the pellet of RNA was washed with 75% ethanol and air-dried, then resuspended in nuclease-free water. Further, isolated RNA was used for cDNA synthesis, as mentioned by the manufacturer's protocol (Takara).

Finally, cDNA from induced and uninduced cells was used to amplify AID using primers PK719 and PK720 (Appendix B).

2.2.24 Mutation of putative splice junctions of human AID

hAID is 597 bp long, and a stretch of 427 to 456 nucleotides likely acts as a putative intron. For creating, a 5' splice site mutant (5'SSM) and 3' SSM, the original sequence was substituted with the *Bam*HI and *Pst*I sites, respectively. The intermediate cloning for the generation of 5', 3', and 5'3' SSM of hAID was performed by sequential cloning of PCR product in pBSK and confirmed by sequencing. Subsequently, all three SSM clones were PCR amplified using primers PK578F/PK516R (Appendix B) with corresponding templates in pBSK, cloned into the pET43a vector at *Spe*I/ *Xho*I to generate 5'SSM Nus-hAID-His, 3' SSM Nus-hAID-His and 5'3'SSM Nus-hAID-His, respectively.

2.2.25 Purification and western blotting of Nus-hAID-His, His-MBP-hAID

Induced Nus-hAID-His cell pellet was dissolved in 50 mM Tris-Cl pH 7.5, 200 mM NaCl, 10% glycerol, and 5 mM imidazole (equilibration buffer) and lysed by sonication. The cell lysate was cleared by centrifugation at 16000g for 30 min at 4 °C. The cell lysate was incubated with Ni-NTA beads pre-equilibrated with equilibration buffer in a gravity column and kept at 4 °C shaking for 1 h. Subsequently, the column was washed with an equilibration buffer containing 20 mM, 40 mM, 60 mM, and 80 mM imidazole. Finally, the Nus-hAID-His protein was eluted in an equilibration buffer containing varying concentrations of imidazole. Induced His-MBP-hAID cell pellets were dissolved in 50 mM Tris-Cl pH 8.0, 15 mM NaCl, 10 μ M ZnCl₂, 10% glycerol, 0.1% tween-20, 1 mM PMSF, 1 mM benzamidinium hydrochloride, 20 mM imidazole and 5 mM β -mercaptoethanol (equilibration buffer) and lysed by sonication (adapted with modification from Yewdell W.T et. al., 2020, Immunity) [93]. Subsequently, the cell lysate was centrifuged at 16000 g for 30 min at 4°C. Ni-NTA beads were equilibrated with 50 mM Tris-Cl pH 8.0, 15 mM NaCl, μ M ZnCl₂, 10% glycerol, 0.1% tween

20, 20 mM imidazole and 5 mM β -mercaptoethanol. Equilibrated beads were mixed with cell lysate and loaded onto the gravity column and incubated at 4°C for 4 hrs on end-end to orbital shaker. Subsequently, beads were washed with 40, 60 and 100 mM imidazole and eluted with 100, 200 and 400 mM imidazole. For western blotting, induced as well as uninduced hAID-His pellet and supernatant or uninduced or induced cell lysate were separated on 12% SDS-PAGE whereas for Nus-hAID-His and His-MBP-hAID protein were separated on 8% SDS-PAGE and transferred to nitrocellulose membrane overnight. Subsequently, the blot was blocked in 20 mM Tris-Cl pH 7.5 and 150 mM NaCl supplemented with 0.02% tween 20 (TBST) in 5% skimmed milk for 4 hrs at 4°C followed by overnight incubation with either anti-His antibody (1:2000) or anti-AID antibody (1:1000) (Appendix C) at 4°C. Further, the membrane was washed in TBST, incubated with anti-mouse-HRP antibody (Appendix C) in TBST (5% skimmed milk) for 3 hrs at 4°C, and washed with TBST. Finally, the blot was developed with chemiluminescence substrate ECL substrate and detected by using Image Quant LAS 4000 gel doc system (GE Healthcare).

2.2.26 Structural modelling of SRSF1-3 and docking with Topoisomerase 1

Homology modelling of SRSF1-3 was performed using the Alphafold server by retrieving the sequence from the Naoki Kanayama (Okayama University, Tsushima-Naka, Kita-Ku, Okayama, Japan). The best model was selected based on the C score. For topoisomerase 1 (TOP1), we obtained the structure from the PDB database. The missing residues of TOP1 were modelled in the alpha fold server. Furthermore, the docking of SRSF1-3 with TOP1 was performed in the ClusPro server. The best-docked complex pose was taken based on prior studies of SRSF1-3 with the TOP1. The docked complex was then subjected to the molecular dynamics study to investigate the dynamics and stability of the complex.

2.2.27 Molecular dynamics simulation protocols for SRSF1-3 and TOP1

In addition to the SRSF1-3 with TOP1 complex, the apo form of TOP1 (without SRSF1-3) was evaluated in this study using Molecular dynamics (MD) simulations to compare the dynamics and conformational changes in TOP1 after its complexation with the SRSF1-3. Therefore, two systems, TOP1_{complex} and TOP1_{apo} were taken into consideration for our study and subjected to 600 ns of MD simulation runs. All systems were subjected to the MD by using the *pmemd.cuda* module of the AMBER18 [82]. All analyses of the MD trajectories were achieved using the *Cpptraj* module [89]. The AMBER ff14SB force was used for assigning parameters to proteins. The system was neutralized with an appropriate number of counter ions and solvated in an octahedron periodic box with the TIP3P water model [84]. The Langevin thermostat [87] and Berendsen barostat [94] were used for controlling the temperature at 300 K and pressure at 1 bar, respectively. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms [85]. A time step of 2 fs was used in simulations. The long-range electrostatic interactions were considered by using the Particle mesh Ewald (PME) scheme, and the non-bonded cut-off was taken as 10 Å [86]. Each system was subjected to two steps of minimization followed by stepwise heating from 0 K to 300 K and equilibrium simulation before the final production run. Finally, production simulations of 600 ns were performed at the NPT ensemble for each system.

2.2.28 Cloning and expression of His-SRSF1-3 and His-SRSF1-3ΔC

Human SRSF1-3 was PCR amplified using primers PK879 and PK951 (Appendix B), digested with *NheI* and *HindIII*, and cloned in a vector at *NheI/HindIII* to give rise to His-SRSF1-3. SRSF1-3ΔC was PCR amplified using primers PK879 and PK911 (Appendix B) and proceeded similarly as for SRSF1-3 to give rise to His-SRSF1-3ΔRS. Subsequently, His-SRSF1-3 and His-SRSF1-3ΔC clones were used for

transformation into rosetta and induced at 0.5 mM IPTG for 37°C/4hrs or 16°C/18hrs. Finally, cells were pelleted down by centrifugation, washed once with 1XPBS and resuspended in lysis buffer (100 mM HEPES, 10 mM Tris-Cl pH 8.0, 0.3 M NaCl, 10 % glycerol and supplemented with 1 mM phenylmethylsulfonyl fluoride) and sonicated (pulse 2 on 2 off, amplitude 65%) on ice till the cells were lysed completely. Finally, uninduced and induced samples of supernatant and pellets were loaded onto SDS-PAGE.

2.2.29 Cloning, expression, and purification of His-MBP-SRSF1-3

The human SRSF1-3 gene was PCR amplified using primers PK776 and PK777 (Appendix B), digested with *Bam*HI and *Hind*III, and cloned in a vector at *Bam*HI/*Hind*III to give rise to His-MBP-SRSF1-3. Similarly, His-MBP-SRSF1-3 clones were processed as described for His-SRSF1-3. Purification of His-MBP-SRSF1-3 was performed on the Ni-NTA gravity column, and washed with lysis buffer with 40 mM and 80 mM imidazole. Subsequently, elution was performed with lysis buffer containing 100 mM, 200 mM and 300 mM imidazole. Lastly, His-MBP-SRSF1-3 was digested with TEV protease at 21°C for 4 hrs, and samples were run on SDS-PAGE.

2.2.30 Topoisomerase 1 inhibition assay

Plasmid DNA relaxation assays were performed by incubating TOP1 in a buffer containing 10 mM Tris-Cl pH 7.9, 1 mM EDTA, 0.15 M NaCl, 0.1% BSA 0.1 mM spermidine and 5% glycerol in the presence and absence of His-MBP-SRSF1-3 and kept at 37°C for 1hr. Subsequently, the reaction was stopped by adding stop buffer (0.125% bromophenol blue, 25% glycerol and 5% sarkosyl) and treated with proteinase K to improve the band's quality. The DNA was isolated via treatment with chloroform/isoamyl alcohol (24:1) followed by loading of upper blue-coloured aqueous phase onto 1% agarose in TAE buffer (40 mM Tris-acetate, pH 8.0 and 2 mM EDTA) and ran till the dye front reached $\frac{3}{4}$ the of the gel. Further, the gel was stained with ethidium bromide and imaged using the GE gel doc system.

2.2.31 Solubilization, purification, refolding, and dialysis of His-SRSF1-3

The Cell pellets of His-SRSF1-3 were washed twice with lysis buffer supplemented with 2M urea. Finally, the insoluble fraction was resuspended in solubilization buffer (lysis buffer supplemented with 6M GuHCl) and kept at room temperature for 1 hr. Further, the solubilized protein was obtained by centrifugation loaded onto the Ni-NTA gravity column, and purified as described for His-MBP-SRSF1-3. Purified His-SRSF1-3 was refolded by dilution methods in a refolding buffer (500 mM L-arginine, 10% glycerol, 100 mM HEPES, 300 mM NaCl, pH 8.0) at 4°C for 24 hrs. Finally, refolded His-SRSF1-3 was dialyzed against 10% glycerol, 100 mM HEPES, and 300 mM NaCl, pH 8.0.

Chapter 3

Chapter 3

Activation-induced cytidine deaminase an antibody diversification enzyme interacts with chromatin modifier UBN1 in B-cells

3.1 Introduction

B-cells diversify their antibody repository from a handful of immunoglobulins (Ig) genes via VDJ recombination, somatic hypermutation (SHM) and class switch recombination (CSR). Activation-induced cytidine deaminase (AID) is the initiator of SHM and CSR in the activated B-cells. An AID-introduced point mutation in the variable region of Ig genes leads to affinity maturation, whereas switch regions result in Ig isotype switching [94]. AID being a deleterious enzyme its activity is predominantly confined to the Ig locus. However, AID can be off-targeted to non-Ig genes in transcriptionally active B-cells, especially to proto-oncogenes, resulting in dysregulation of proto-oncogenes and chromosomal translocation [30].

AID plays a crucial role in the immune system, and; therefore, extensive investigations have been conducted to identify its mechanism of targeting and its interaction partners. Remarkably, even before the discovery of AID, it was well documented that transcription is required for SHM [48]. Furthermore, transcription initiation, transcription elongation, RNA PolII pausing, and termination are also required for AID action [48,50]. However, transcription through the genes is often encountered by nucleosomes (its composition and architecture) limiting the accessibility of AID to single-stranded DNA. Interestingly, immune cells utilize the facilitated chromatin transcription complex (FACT) and HIRA chaperone complexes to efficiently modify nucleosomes for AID targeting [56,57]. Indeed, FACT and HIRA deposit histone H3.3 (H3.3) at Ig genes that in turn provide AID-accessible single-stranded DNA [95]. Additionally, AID was found to directly interact with CTNNB1, SRSF1-3, and PTBP2 which dictates the AID-mediated antibody diversification [64–66].

HIRA chaperon complex is composed of ubinuclein 1 (UBN1), anti-silencing function 1A histone chaperone (ASF1a), HIRA, and calcineurin binding protein 1 (CABIN1). Interestingly, Zheng *et al.*, reported that replication protein A (RPA) interaction with HIRA protein, a member of the HIRA chaperon complex, and H3.3 is crucial for HIRA-mediated, H3.3 deposition at gene regulatory elements [76]. Incidentally, an earlier study showed that RPA was also involved in interaction with AID in activated B-cells [63]. In brief, AID interaction with RPA is well established, and RPA is also reported to interact with the HIRA protein of the HIRA chaperone complex.

In this study, we try to explore whether there is any interaction of AID with UBN1 a member of the HIRA chaperon complex. In our study, we employed *in-silico* studies like molecular docking and molecular dynamics (MD) simulation, co-immunoprecipitation, *in-vitro* pull-down assay as well as double immunofluorescence assay along with proximity ligation assay to demonstrate the interaction between AID and UBN1. This study unfolds yet another AID interacting partner, UBN1. However, the AID-UBN1 interaction and its significance in the context of SHM or CSR still need to be explored.

3.2 Results

3.2.1 *In-silico* estimation of UBN1-AID interaction

AID is a key mediator of the antibody diversity in the activated B-cells. AID actions and targeting is an orchestrated event that requires the crosstalk between numerous proteins and cis-acting elements. Interestingly, UBN1 is also a crucial member of the chromatin modifier complex that cross-talks with one of the AID interacting partners [76]. In this study, we wanted to explore whether AID and UBN1 interact with each other or not, which may have a role in antibody diversity. Furthermore, to shed light on the molecular interaction between UBN1-AID, homology modelling and molecular docking studies were used. Due to very little sequence similarity, we modelled only the N-terminal domain (1-175) of UBN1 using the i-TASSER web server. On the other hand, the modelled structure of AID was already available in the alpha-fold webserver. Ramachandran plot was calculated for modelled UBN1 to validate the structure. A molecular docking study was conducted for the modelled structure using Rosetta v.3.2 using the online Rosie web server [96]. Among the top ten models, the lowest energy-minimized structure was used for the input of the MD study. Initial inspection of UBN1-AID showed that interaction was mainly driven through the initial N-terminal residues of UBN1. The three-dimensional conformation of the docked complex and its conformational change after the MD run are shown (**Fig. 3.1 A and B**). Conformational change upon MD indicates the stability of the binding.

To estimate the molecular mechanism of the UBN1 and AID, an MD simulation was conducted. A 100 ns long equilibrium stage was conducted to stabilize the complex as well as UBN1 and AID. The reason behind the long equilibrium step is to stabilize the model complex before the binding free energy calculation. The final production run of 50 ns shows the stable conformation of for complex as well as in both the protein as shown from the time evolution of root mean square deviation (RMSD) (**Fig. 3.2 A**).

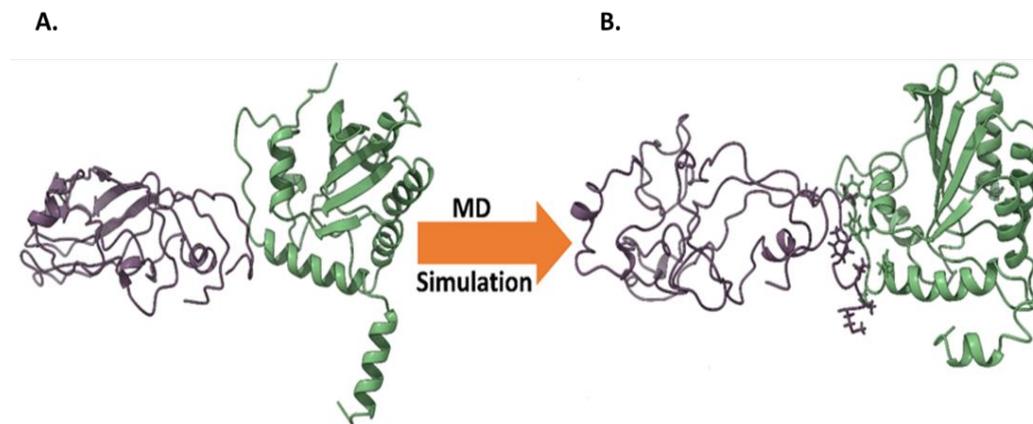


Fig. 3.1. *In-silico* study on the interaction of UBN1 and AID. (A) UBN1 and AID complex after docking, (B) UBN1 and AID complex after MD simulation. UBN1: Violet, AID: Green.

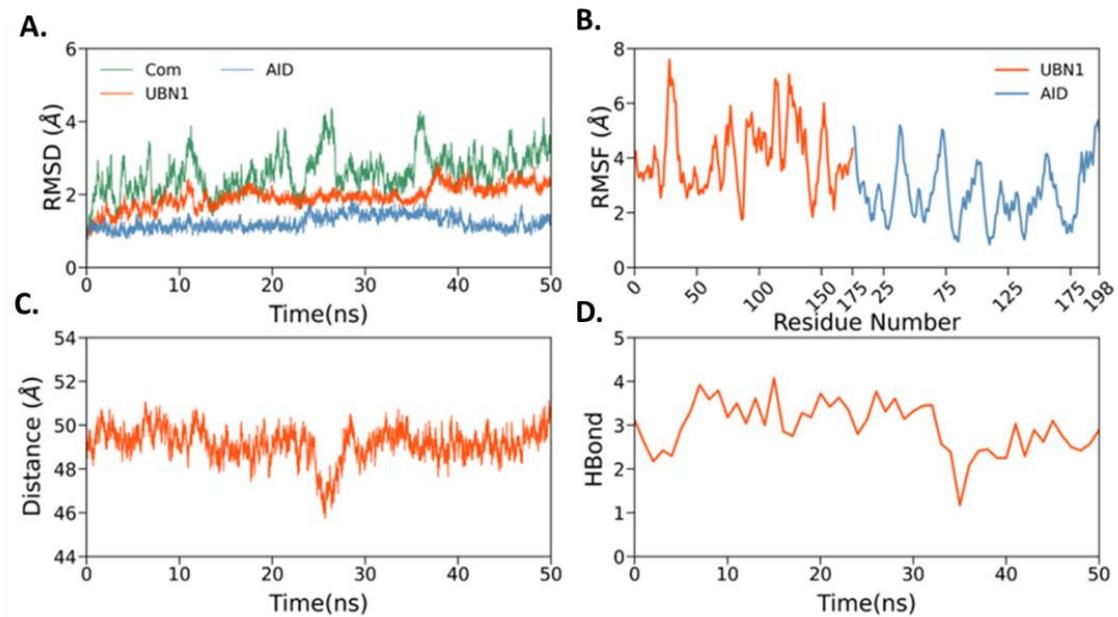


Fig. 3.2. RMSD value of AID & UBN1 atoms (A) Time evolution of root mean squared deviation (RMSD) of complex and its components, (B) Root mean squared fluctuation (RMSF) of UBN1 and AID, (C) Time evolution of protein-protein center of mass distance, (D) Time evolution of hydrogen bonds between UBN1 and AID.

Further, to estimate the residual fluctuation, we estimate the root mean squared fluctuation (RMSF) (**Fig. 3.2 B**), which shows higher flexibility of UBN1 compared to AID. However, the interface residues of UBN1 (initial N-terminal residues) show lesser fluctuation indicating a stable binding. Also, to estimate the stability of the binding between UBN1 and AID, the centre of mass distance was calculated, which remains stable in due course of our simulation (**Fig. 3.2 C**). This was also supported by the stable 2-3 hydrogen bonds between the UBN1 and AID through the simulation length (**Fig. 3.2 D**).

To estimate the binding in more detail, molecular mechanics Poisson Boltzmann surface area (MM/PBSA) was used throughout the MD trajectory (**Table 3.1**). Different components of the binding free energy are also shown graphically (**Fig. 3.3 A**). It is evident from (**Table 3.1**), that intermolecular van der Waals interaction (E_{vdW}), electrostatic interaction (E_{elec}), and non-polar solvation (G_{np}) favored the interaction, while polar solvation (G_{pol}) disfavoured the binding. The binding free energy between the proteins is -31.18 kcal/mol. The interaction pattern shows a higher contribution from electrostatic components (-167.05 kcal/mol),); however, this completely cancels due to an increase in the polar solvation contribution (169.88 kcal/mol). So, the main interaction was governed due to the van der Waals contribution (-29.90 kcal/mol). Furthermore, this is supported by our Ligplot⁺ analysis, which shows a greater number of hydrophobic contacts (**Fig. 3.4**). Also, to evaluate the key residues in the binding process, the decomposition of binding free energy was calculated (**Fig. 3.3 B**). The key residues in the binding process are S2, P4, and F9 for UBN1 and F141, E163, and H166 for AID. This indicates the binding residues mainly reside around the N-terminal and C-terminal of UBN1 and AID, respectively. Due to the lack of sequence similarities, we restrict our model to the N-terminal part of UBN1. So, the findings of the computational investigation can shed light on the molecular mechanism of the interactions. This will be one of the first efforts to quantify the interaction pattern between UBN1 and AID.

	ΔE_{vdW}	ΔE_{elec}	ΔG_{pol}	ΔG_{np}	$\Delta E_{\text{MM}}^{\text{a}}$	$\Delta G_{\text{solv}}^{\text{b}}$	$\Delta G_{\text{bind}}^{\text{c}}$
UBN1-Aid complex	-29.90 (0.07)	-167.05 (0.59)	169.88 (0.54)	-4.13 (0.01)	-196.94 (0.58)	165.76 (0.54)	-31.18 (0.11)

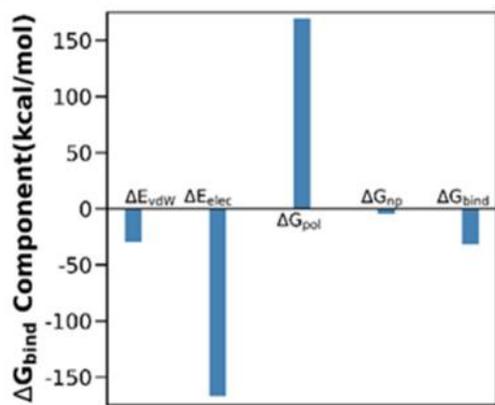
$$a = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}}$$

$$b = \Delta G_{\text{pol}} + \Delta G_{\text{np}}$$

$$c = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}} + \Delta G_{\text{pol}} + \Delta G_{\text{np}}$$

Table 3.1. Binding free energy components (kcal/mol) for the binding of the UBN1-AID complex. ΔE_{vdW} : van der Waals energy; ΔE_{elec} : electrostatics energy in the gas phase; ΔG_{pol} : polar solvation energy; ΔG_{np} : nonpolar solvation energy; ΔG_{bind} : total binding free energy.

A.



B.

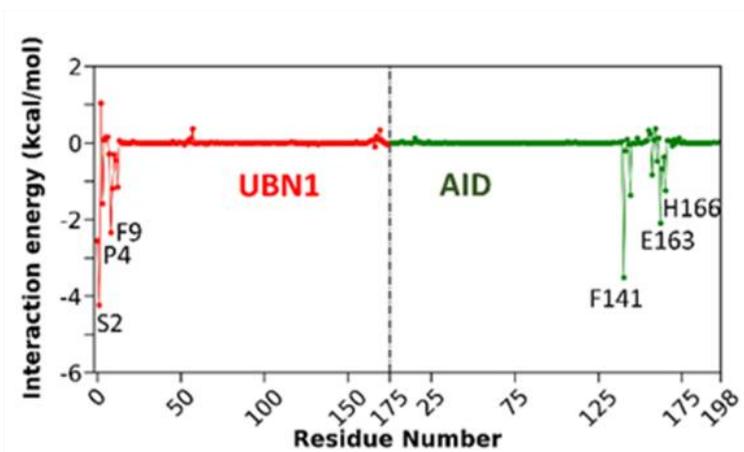


Fig. 3.3 Energy components for binding of AID to UBN1 (A) Binding free energy and its component for UBN1 and AID complex, **(B)** Residue wise contribution of binding free energy. Key residues are mentioned in the plot.

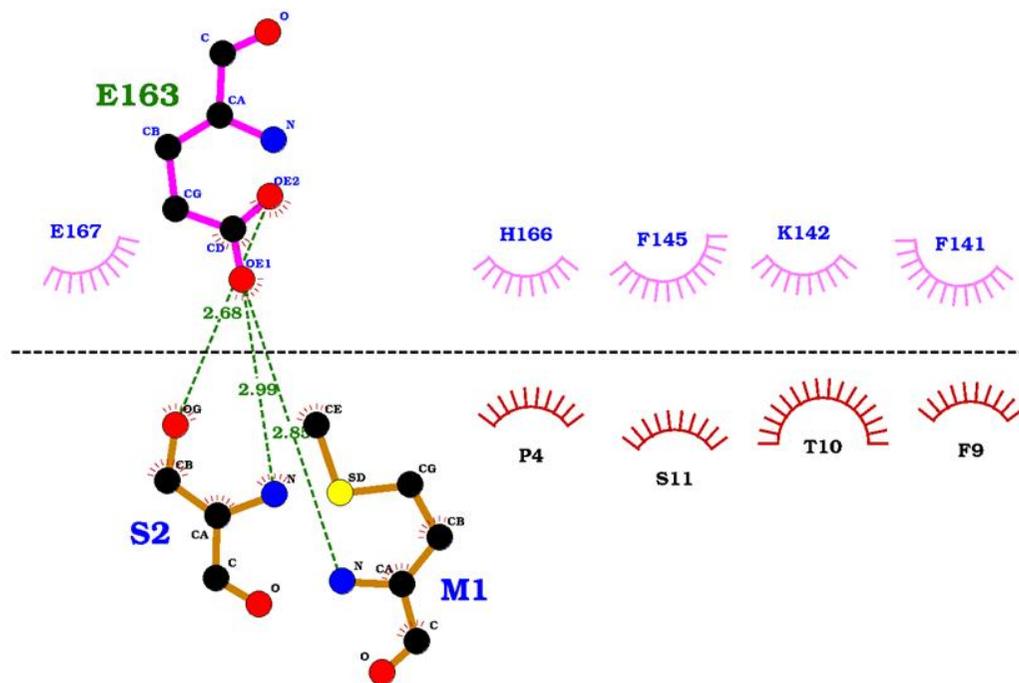


Fig. 3.4. The protein-protein interaction profile between UBN1 and AID constructed using Ligplot⁺. Hydrogen bonds are shown in green dotted line. Semicircles show the residues involving hydrophobic contact.

3.2.2 Interaction of AID with UBN1

In-silico studies predicted that UBN1 interacts with AID. To further, validate whether UBN1 AID interaction happens in B-cells, we performed co-immunoprecipitation experiments. In the co-immunoprecipitation assay, we pulled down AID using an anti-AID antibody from human Raji cells and chicken DT40 ψ VKO cells and detected using an anti-UBN1 antibody. As anticipated, AID interacting with partner UBN1 was detected in both cell lines (**Fig. 3.5 A & Fig. 3.6. A**). Similarly, we pulled down UBN1 using an anti-UBN1 antibody in Raji and DT40 ψ V KO and probed with an anti-AID antibody. Likewise, UBN1 interacting partner AID was detected (**Fig. 3.5 B & Fig. 3.6. B**). So, based on our co-immunoprecipitation we report that AID interacts with a member of HIRA chaperon complex UBN1 in B-cells. Interestingly, as AID interacts with UBN1 how this interaction influences the SHM, or CSR at the Ig locus needs to be explored further. Moreover, the co-immunoprecipitation assay was also performed from DT40 AID KO cell lysate and pulled down using anti-UBN1 and WB anti-AID and vice versa to eliminate any non-specific pull-down. As anticipated no signals (**Fig. 3.7**) were detected in the Co-IP sample which suggests pull down in DT40 ψ V KO cells were specific. Further, a pull-down experiment using an anti-ASF1a antibody as well as an anti-HIRA antibody to identify any association of AID with other members of the HIRA chaperon complex. Interestingly, AID was not found to be associated with either ASF1a or HIRA (**Fig 3.8, 3.9 & 3.10**) that illustrate AID is at least not interacting with UBN1 via HIRA or ASF1a.

Further, to validate that AID interaction with UBN1 is specific, we performed an *in-vitro* pull-down assay where either purified AID or UBN11-175 was incubated with cell lysate. We incubated purified His-MBP-AID with Raji and DT40 ψ V KO cell lysate and pulled down using Ni-NTA beads, and UBN1 was detected in the pulldown sample (**Fig. 3.11 A and C**). Similarly, His-MBP-UBN11-175 protein was incubated with Raji and DT40 ψ V KO cell lysate and pulled down using Ni-NTA and AID was detected (**Fig. 3.11 B and D**).

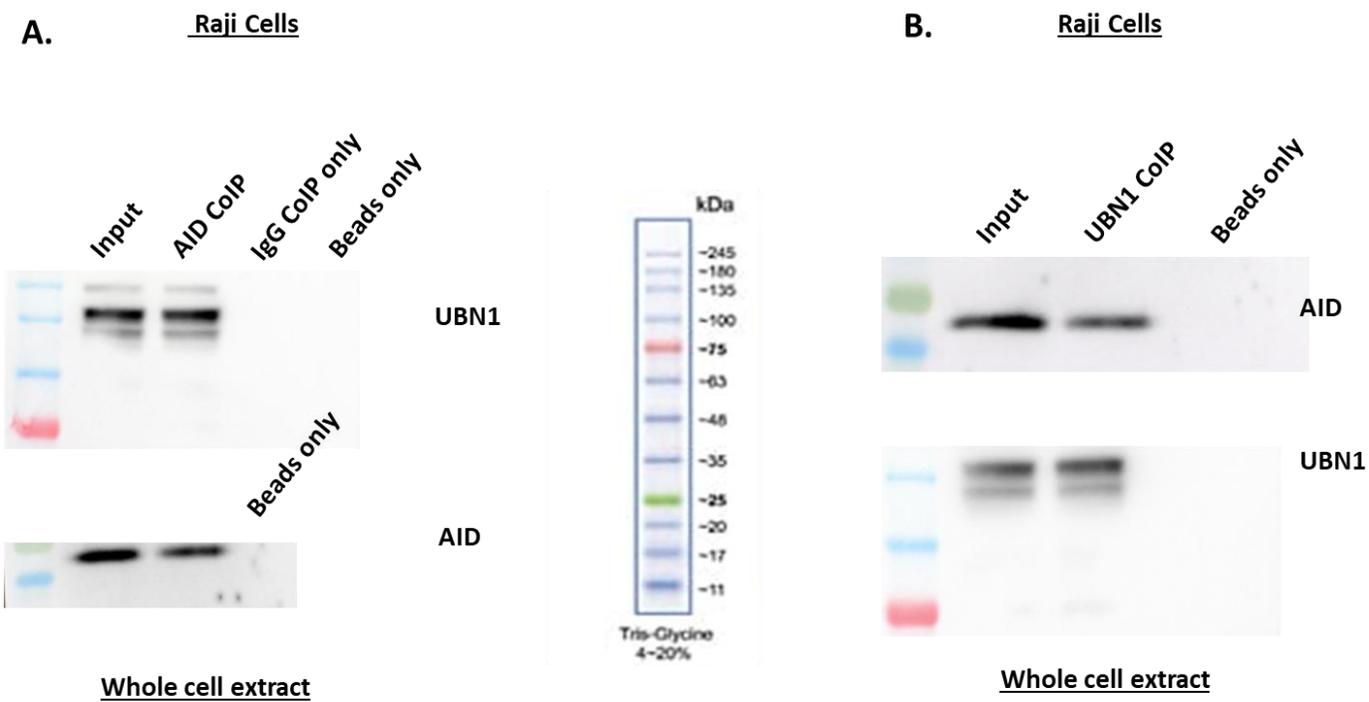
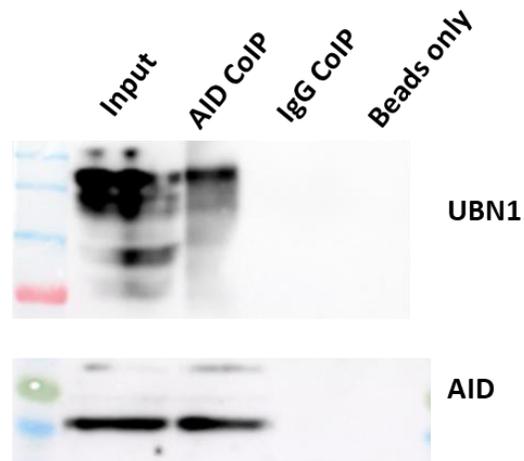


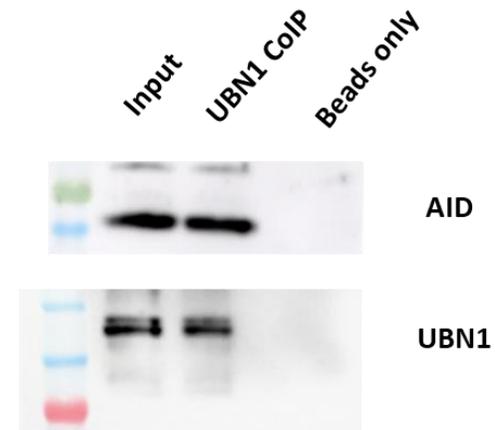
Fig. 3.5. AID interaction with UBN1 in Raji cells. (A) Co-immunoprecipitation of AID from whole cell lysate of Raji cells using anti-AID mAb and protein A/G beads, detection via WB using anti-UBN1 mAb. (B) Co-immunoprecipitation of UBN1 from whole cell lysate was performed on Raji cells using anti-UBN1 pAb and proteinA/G beads, and detection via WB using anti-AID mAb.

A. DT40 ψ V KO cells



Whole cell extract

B. DT40 ψ V KO cells



Whole cell extract

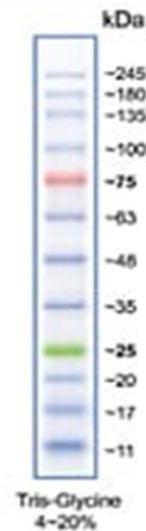


Fig. 3.6. AID interaction with UBN1 DT40 ψ V KO cells. (A) Co-immunoprecipitation of AID from whole cell lysate DT40 ψ V KO cells using anti-AID mAb and protein A/G beads, detection via WB using anti-UBN1 mAb. (B) Co-immunoprecipitation of UBN1 from whole cell lysate was performed on DT40 ψ V KO cells using anti-UBN1 pAb and proteinA/G beads, and detection via WB using anti-AID mAb.

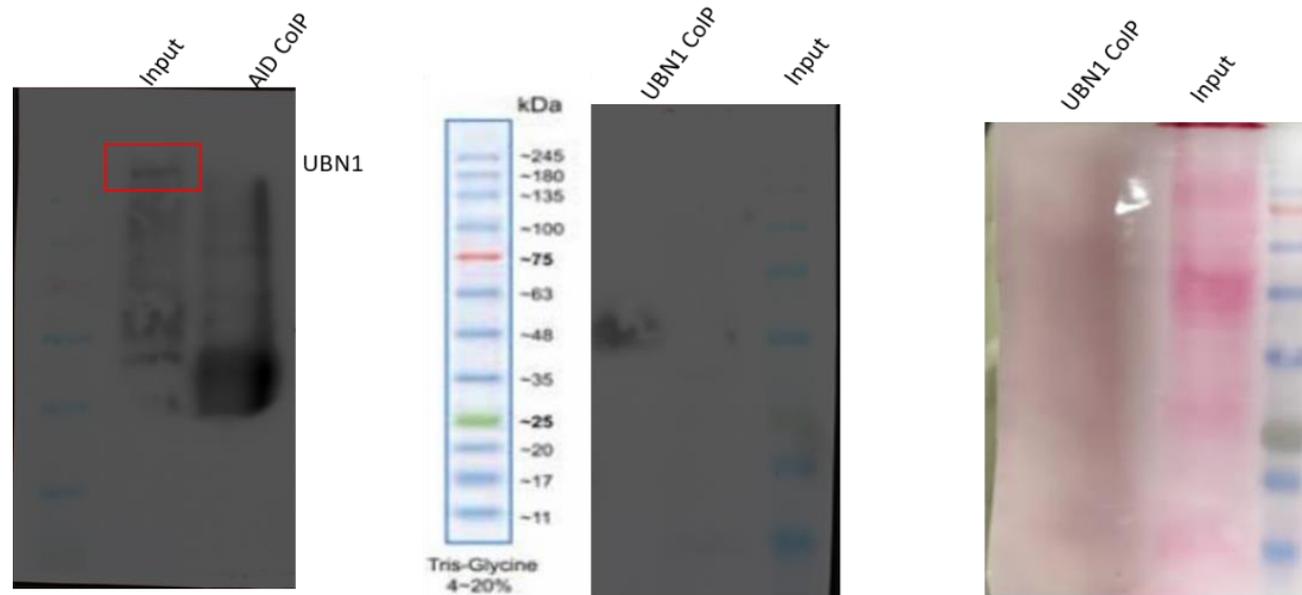


Fig. 3.7. Co-immunoprecipitation of UBN1 and AID in DT40 AID KO cells. A. Cell lysate was used to pull down with an anti-AID antibody and probed with an anti-UBN1 antibody. B. Cell lysate was used to pull down using an anti-UBN1 antibody and probed with an anti-AID antibody. As these cells are AID KO, no AID signal was detected in the input. C. Ponceau staining of western blot (Fig. B).

A. Raji Cells



Whole cell extract

B. Raji Cells



Whole cell extract

Fig. 3.8. AID interaction with ASF1a in Raji cells. (A) Co-immunoprecipitation of ASF1a from whole cell lysate was performed on Raji cells using anti-ASF1a pAb and protein A/G beads, detection via WB using anti-AID mAb. (B) Co-immunoprecipitation of AID from whole cell lysate of Raji cells using anti-AID mAb and protein A/G beads, detection via WB using anti-ASF1a mAb.

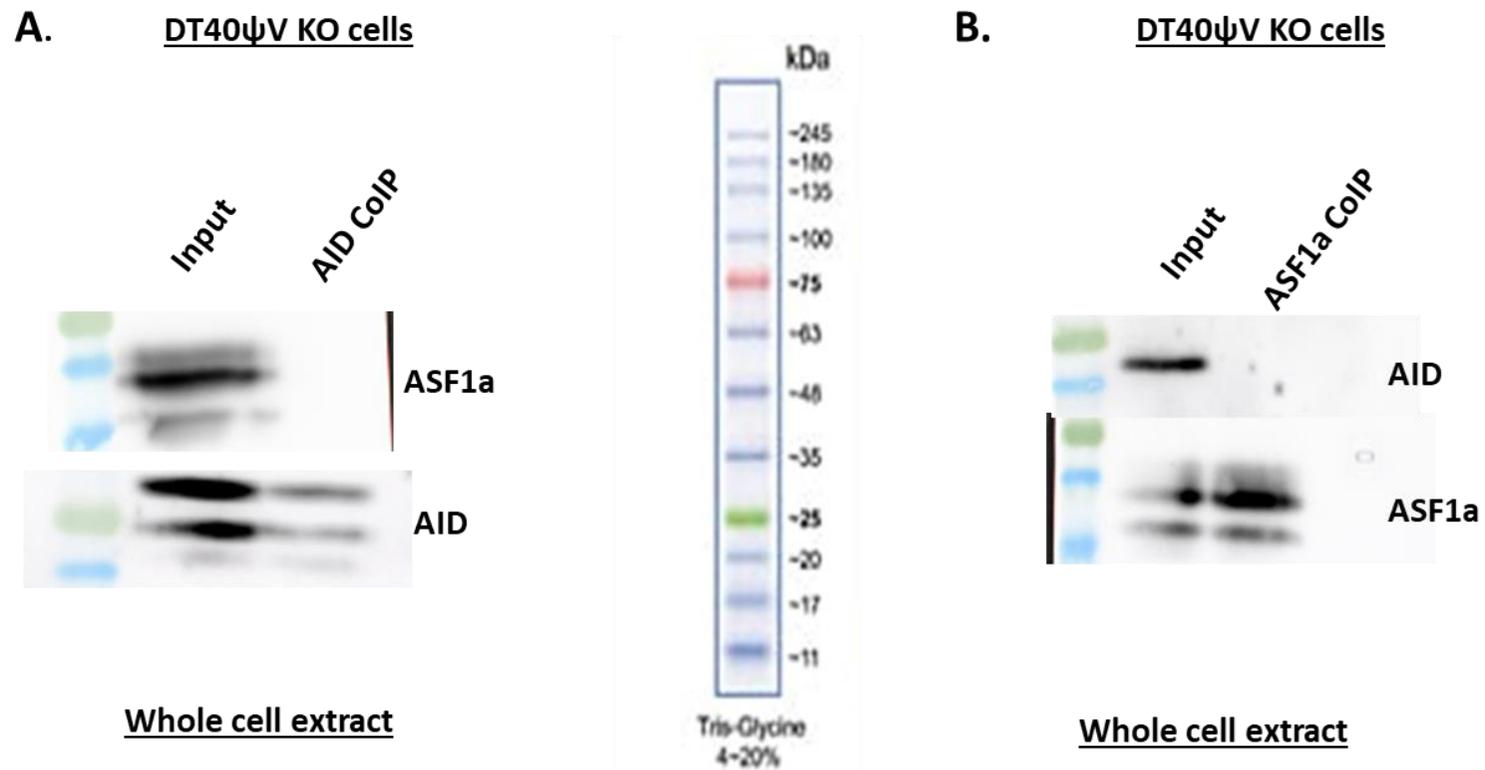


Fig. 3.9. AID interaction with ASF1a in DT40 ψ V KO cells. (A) Co-immunoprecipitation of ASF1a from whole cell lysate was performed on DT40 ψ V KO cells using anti-ASF1a pAb and protein A/G beads, detection via WB using anti-AID mAb. (B) Co-immunoprecipitation of AID from whole cell lysate of DT40 ψ V KO cells using anti-AID mAb and protein A/G beads, detection via WB using anti-ASF1a mAb.

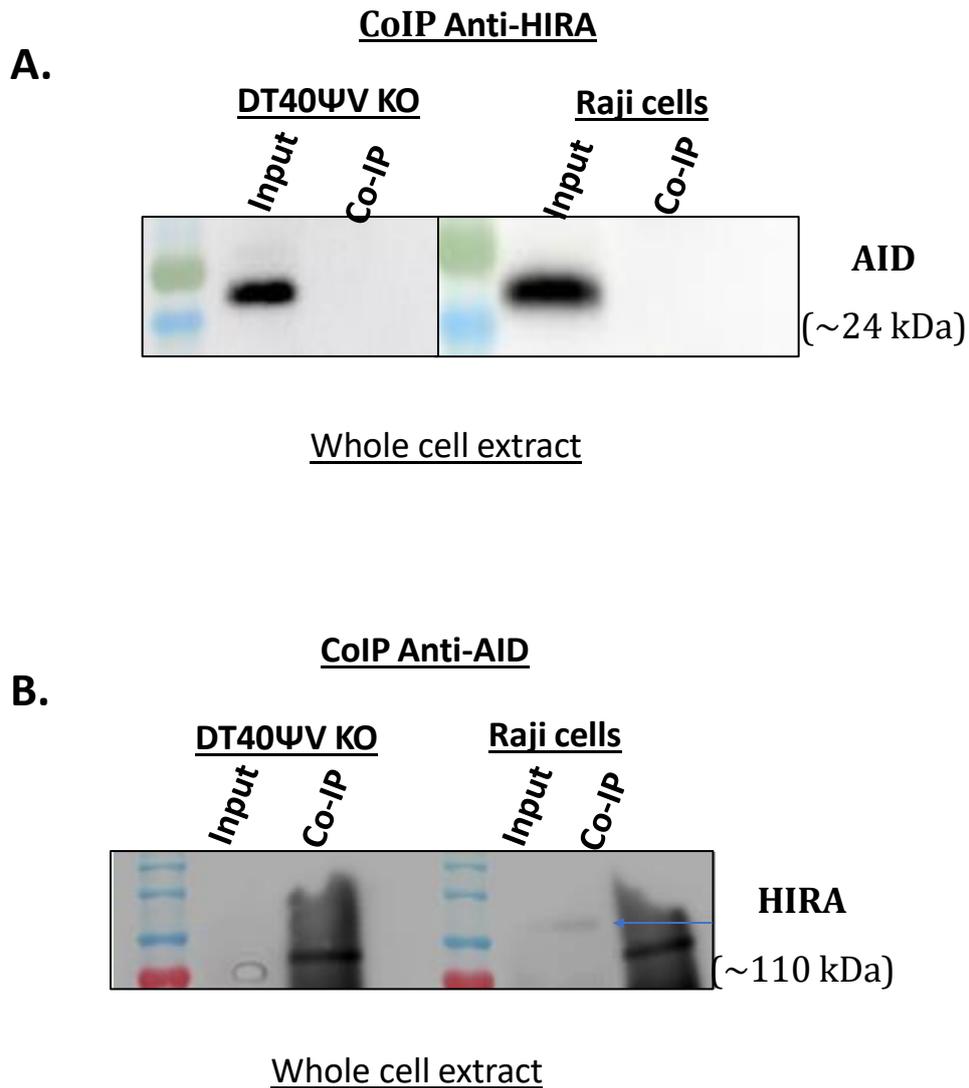


Fig. 3.10. AID interaction with HIRA in Raji and DT40 ψ V KO cells. (A) Co-immunoprecipitation of HIRA from whole cell lysate was performed on Raji and DT40 ψ V KO cells using anti-HIRA pAb and protein A/G beads, detection via WB using anti-AID mAb. (B) Co-immunoprecipitation of AID from whole cell lysate of Raji and DT40 ψ V KO cells using anti-AID mAb and protein A/G beads, detection via WB using anti-HIRA mAb.

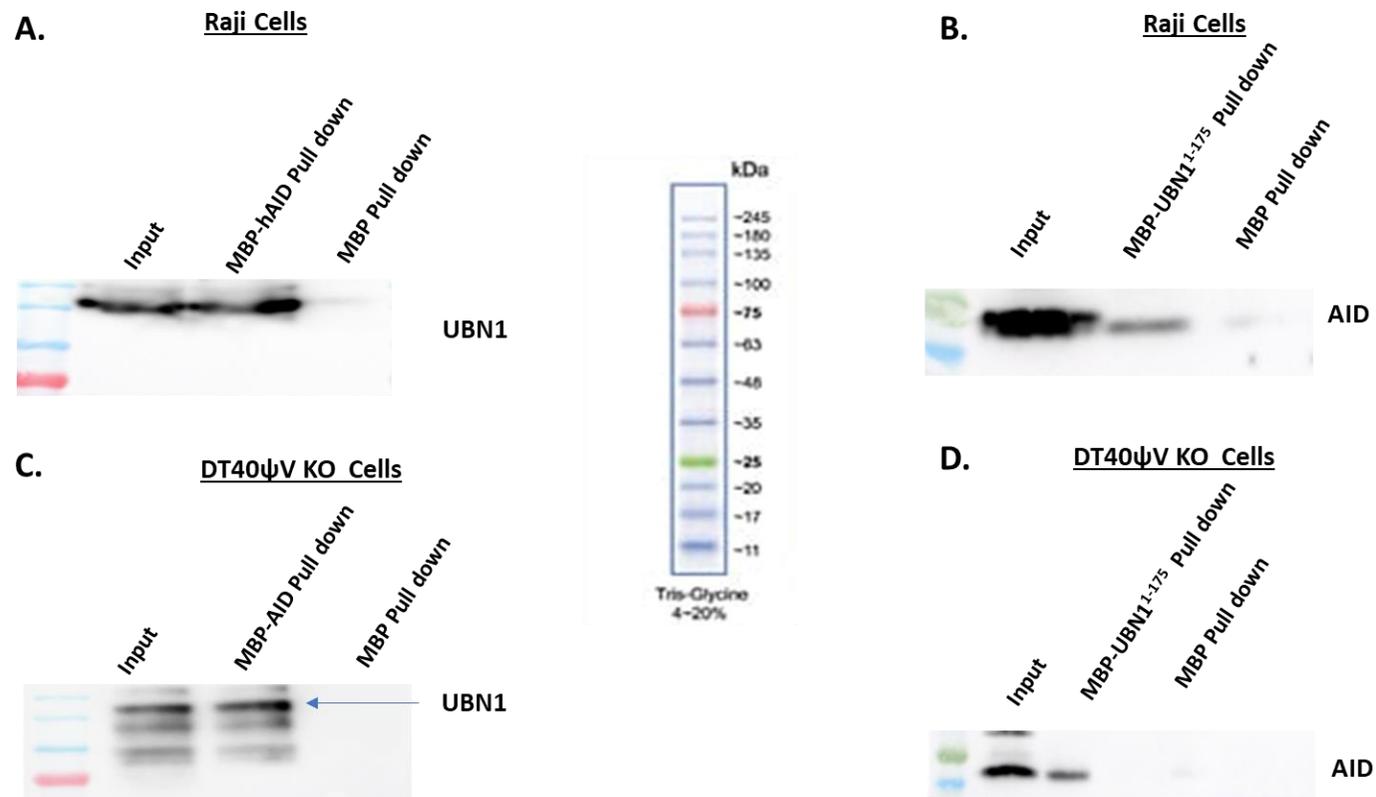


Fig. 3.11. His-tagged pull-down assay for AID interaction with UBN1 in Raji cells and DT40ψV KO cells. (A) His tagged MBP-AID or MBP were immobilized on Ni-NTA beads followed by incubation with Raji cell lysate and analysed by western blotting using anti-UBN1 pAb. (B) His tagged MBP-UBN1¹⁻¹⁷⁵ or MBP were immobilized on Ni-NTA beads followed by incubation with Raji cell lysate and analysed by western blotting using anti-AID mAb. (C) His tagged MBP-AID or MBP were immobilized on Ni-NTA beads followed by incubation with DT40ψV KO cell lysate and analysed by western blotting using anti-UBN1 pAb. (D) His tagged MBP-UBN1¹⁻¹⁷⁵ or MBP were immobilized on Ni-NTA beads followed by incubation with DT40ψV KO cell lysate and analysed by western blotting using anti-AID mAb.

As anticipated, His-MBP-hAID was found to interact with UBN1 and vice-versa. Moreover, His-MBP was used as a control to eliminate any possibility of purification and solubilization tag interacting with UBN1 or AID. Additionally, no band was observed in the control sample, which suggests the specific interaction between AID and UBN1.

Moreover, to confirm AID interaction with UBN1, we performed a pull-down assay using purified His-MBP-hAID, His-MBP, His-MBP-Flag-UBN1¹⁻¹⁷⁵, MBP-hAID (without His-tag) and MBP-Flag-UBN1¹⁻¹⁷⁵ (without His-tag). We incubated His-MBP-hAID with MBP-Flag-UBN1¹⁻¹⁷⁵ and pulled it down using Ni-NTA beads, and MBP-Flag-UBN1¹⁻¹⁷⁵ was detected using anti-flag Ab (**Fig. 3.12 A and C**). Similarly, His-MBP-Flag-UBN1¹⁻¹⁷⁵ was incubated with MBP-hAID and pulled down using Ni-NTA beads, and MBP-hAID was detected using an anti-AID antibody. Furthermore, no signal was detected in His-MBP control neither in His-MBP-hAID nor His-MBP-Flag-UBN1¹⁻¹⁷⁵ pull down (**Fig. 3.12 B and D**). As anticipated, our *in-vitro* pull-down assay showed AID interacts with UBN1 and vice versa. Thus, similar to our co-immunoprecipitation assay, the pulldown assay using cell lysate, as well as the pull-down assay with purified proteins, strongly suggest that AID is interacting with UBN1.

3.2.3 Co-localization studies for AID with UBN1

As our *in-silico* and co-immunoprecipitation data indicate that AID interacts with UBN1, we wanted to confirm whether UBN1 and AID are co-localized inside the B-cells. Thus, we performed a co-localization assay of both proteins using double immunofluorescence staining with anti-AID monoclonal antibody and anti-UBN1 polyclonal antibody in Raji as well as in DT40 DT40 ψ V KO cells (**Fig. 3.13, 3.14, 3.15 and 3.16**). As expected, AID is predominantly localized to the cytoplasm as it consists of a weak nuclear localization signal and a strong nuclear export signal, whereas a small fraction of AID is localized in the nucleus.

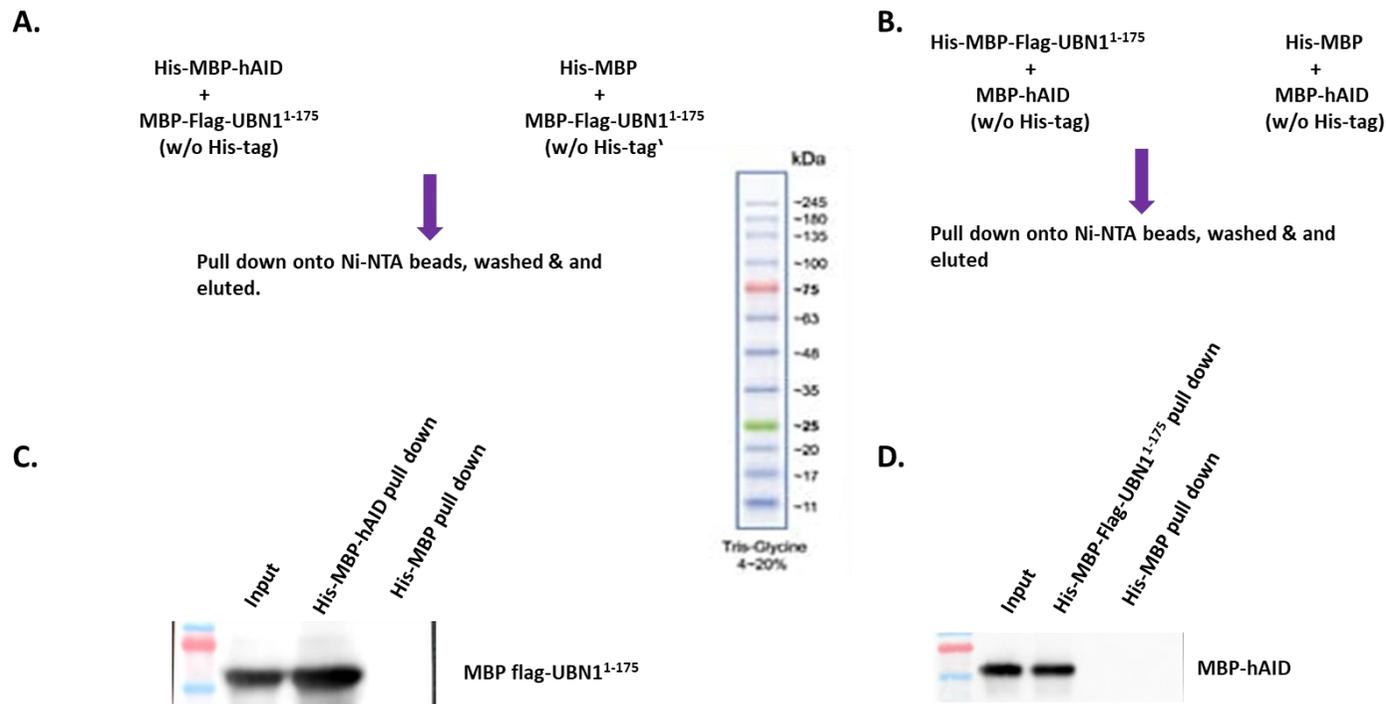


Fig. 3.12. *In-vitro* His-tagged pull-down assay using purified protein to detect the interaction between UBN1 and hAID. Fig. A & B represent the strategy used for pull-down in Fig. C & D, respectively. (C) MBP-Flag-UBN1¹⁻¹⁷⁵ (w/o His tag) was added to His-MBP-hAID and His-MBP, followed by pull-down using Ni-NTA beads and detection by western blotting using anti-flag-mAb. (D) MBP-hAID (w/o His tag) was added to His-MBP-Flag-UBN1¹⁻¹⁷⁵ & His-MBP followed by pull-down using Ni-NTA beads and detected by western blotting using anti-AID mAb.

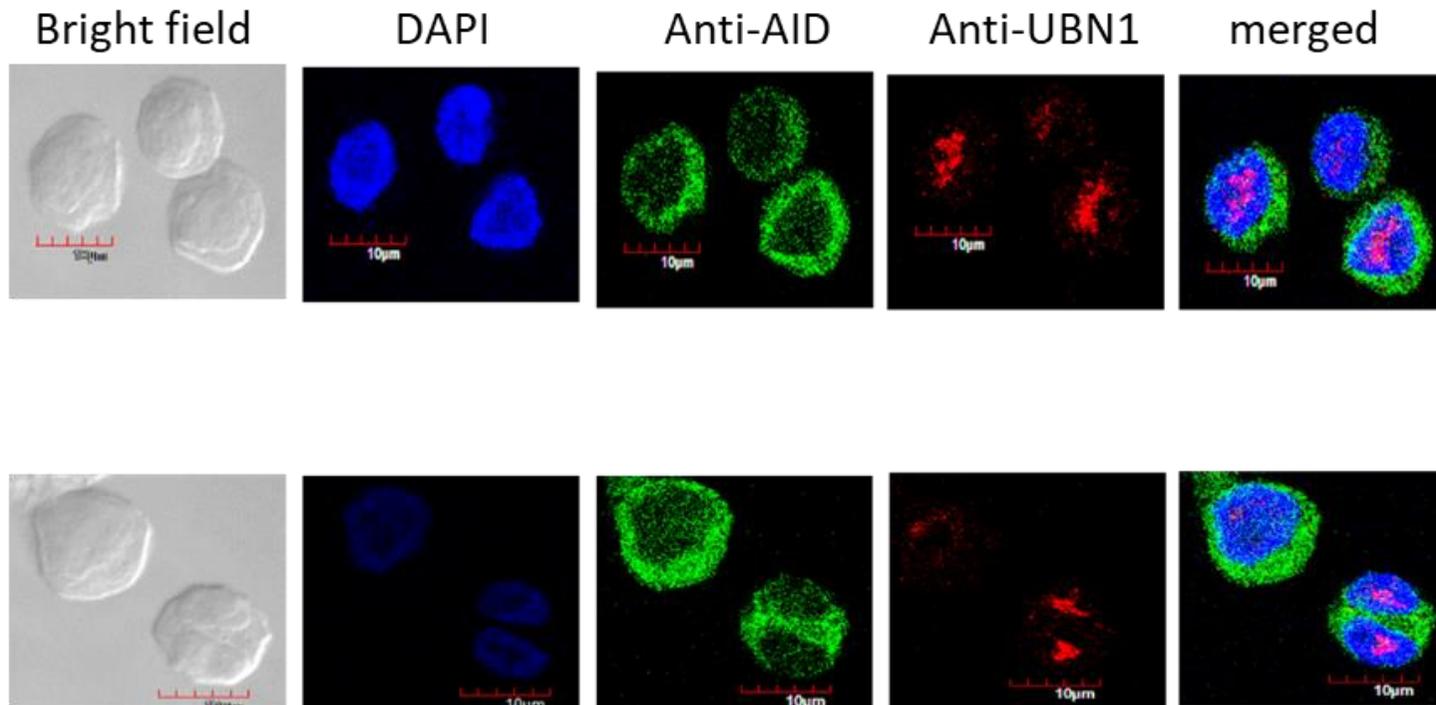


Fig. 3.13. Co-localization of AID with UBN1 in DT40ΨV KO cells. Cells were fixed and double stained with monoclonal anti-AID (green) and polyclonal anti-UBN1 (red) antibodies. AID is predominantly located in the cytoplasm whereas UBN1 is mostly inside the nucleus. Yellow and orange spots are produced due to the co-localization of AID with UBN1.

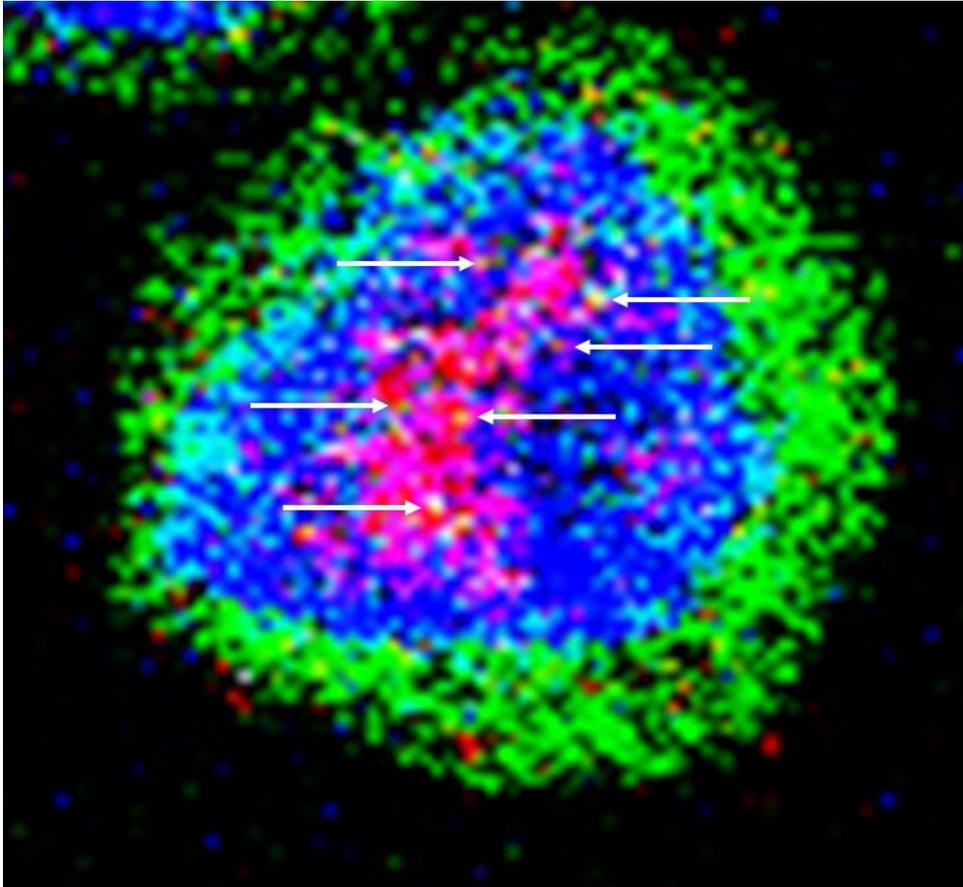


Fig. 3.14. Merged zoomed image of AID and UBN1 co-localized in DT40 ψ V KO cells. The yellow/orange spots represent the co-localization of AID and UBN1.

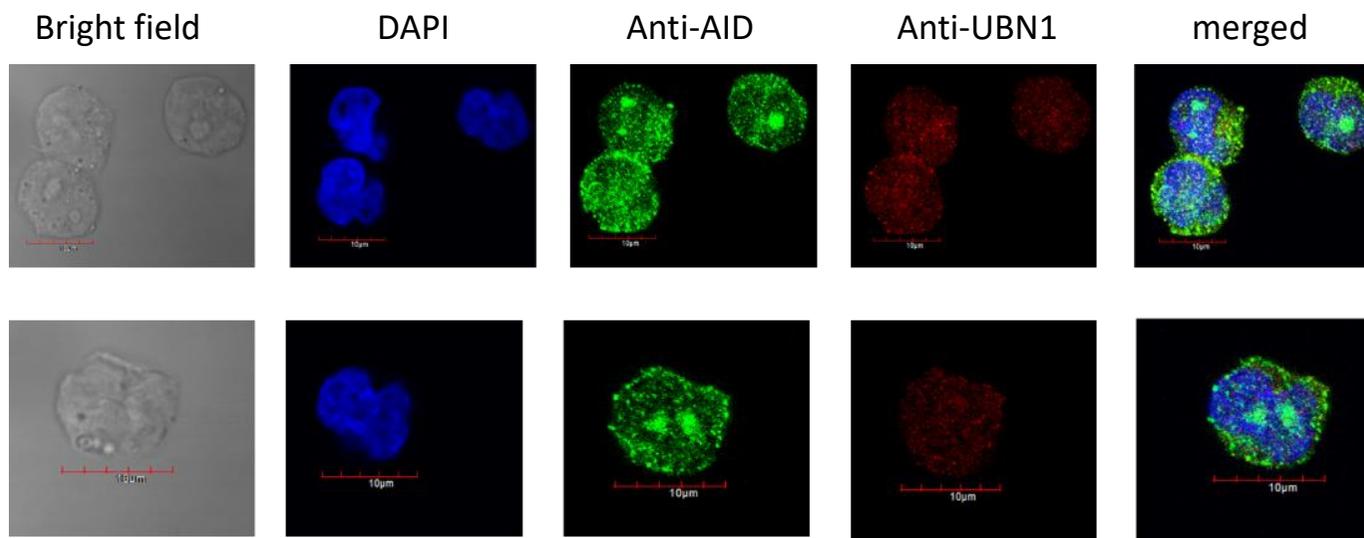


Fig. 3.15. Co-localization of AID with UBN1 in Raji cells. Cells were fixed and double stained with monoclonal anti-AID (green) and polyclonal anti-UBN1 (red) antibodies. AID is predominantly located in the cytoplasm whereas UBN1 is mostly inside the nucleus. Yellow and orange spots are produced due to the co-localization of AID with UBN1.

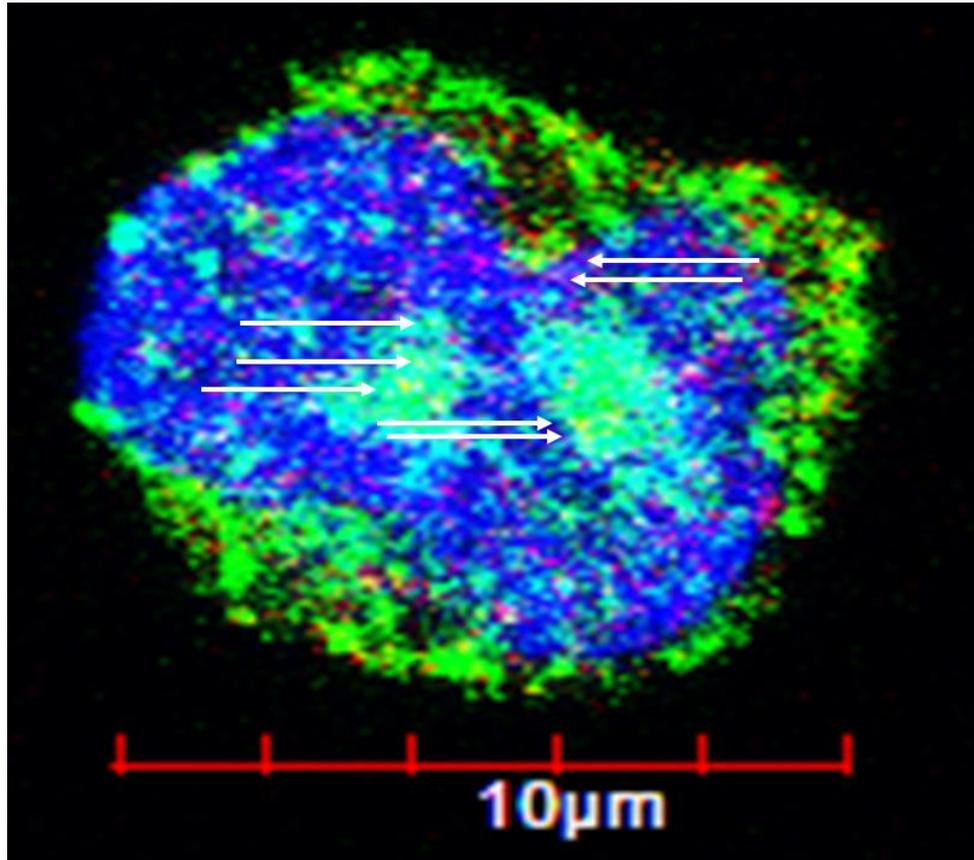


Fig. 3.16. Merged zoomed image of AID and UBN1 co-localized in Raji cells. The yellow/orange spots represent the co-localization of AID and UBN1.

On the contrary, UBN1 is primarily localized in the nucleus of DT40 Ψ V KO cells, whereas it is distributed throughout the cytosol and nucleus in Raji cells. As UBN1 is a part of the HIRA chaperon complex that acts on chromatin, its predominant localization in the nucleus is indispensable for regulating the dynamics of chromatin. However, several yellow and orange spots were observed inside the nucleus of Raji as well as DT40 Ψ V KO cells which validated our previous observations in the co-immunoprecipitation assay that UBN1 interacts with AID inside B-cells. Further, the co-localization assay was also performed on DT40 AID KO cells to eliminate any non-specific co-localization of UBN1 and AID. As anticipated, the co-localization assay on DT40 AID KO cells (**Fig. 3.17**) reveals there was no significant discrete green signal due to the absence of AID protein. So, AID and UBN1 co-localization in DT40 Ψ V KO cells is due to the presence of each other at the same location.

3.2.4 Proximity ligation assay for direct interaction between AID and UBN1

Additionally, to understand the type of molecular interaction taking place between UBN1 and AID we performed a proximity ligation assay. Raji cells were incubated with anti-UBN1 and anti-AID antibodies, subsequently, secondary antibodies (as mentioned by the manufacturer) were added. Strikingly, we detected PLA spots (red) in Raji cells, but the intensity and number of spots were quite low (**Fig. 3.18**). Furthermore, all the PLA spots were found inside the nucleus. As optimum PLA spots can only be formed if the two proteins are within the range of 10Å, it suggested that the interaction between UBN1 and AID could be either weak or transient. Similarly, the PLA assay was also performed on DT40 Ψ V KO cells (**Fig. 3.19**), and we detected PLA spots that suggest UBN1 and AID interaction inside the cell. Thus, our PLA data demonstrates AID and UBN1 interaction is direct in both DT40 Ψ V KO and Raji cells.

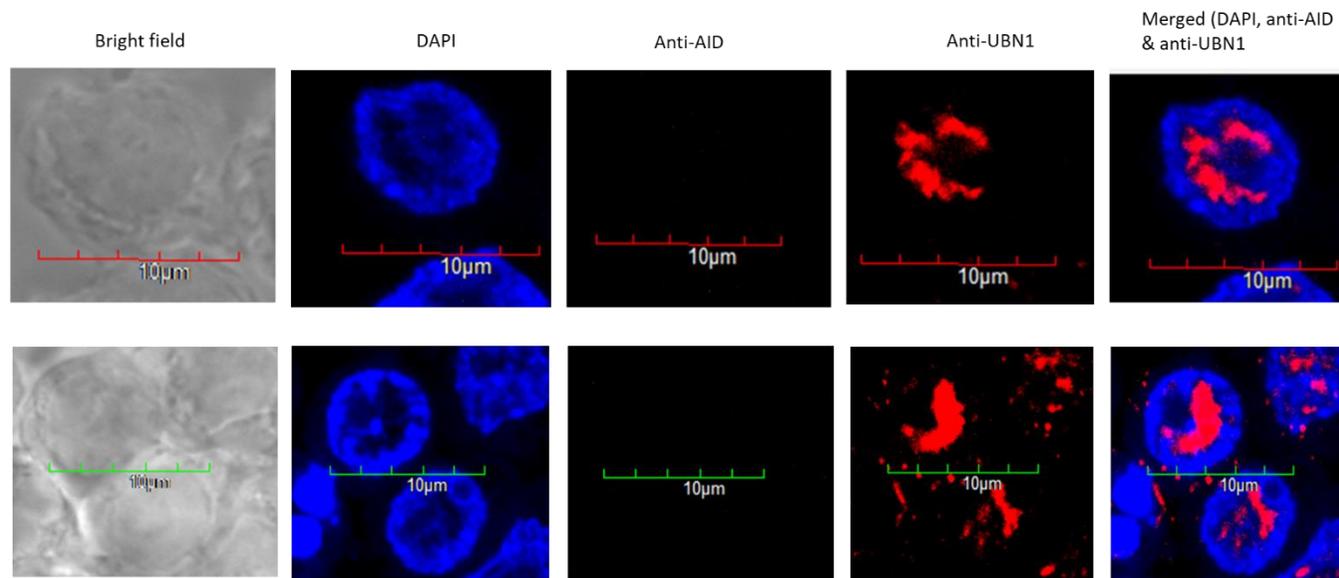


Fig. 3.17. Co-localization of AID with UBN1 in DT40 AID KO cells. Cells were fixed and double stained with monoclonal anti-AID (green) and polyclonal anti-UBN1 (red) antibodies. UBN1 is mostly inside the nucleus.

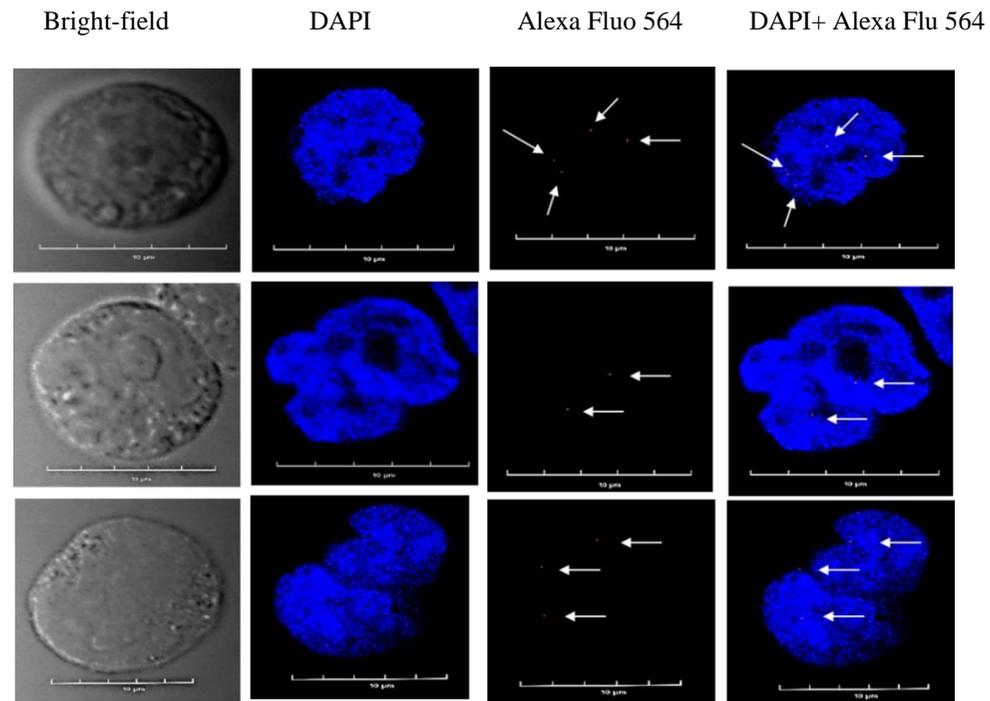


Fig. 3.18. Proximity ligation assay for AID and UBN1 in Raji cells. Images demonstrate the PLA signals in Raji cells and the proximity between two proteins, AID and UBN1. Nuclei were stained with DAPI. Anti-AID and anti-UBN1 primary antibodies were incubated, followed by incubation with secondary antibodies conjugated to oligonucleotides (PLA Probes anti-mouse MINUS and PLA Probes anti-rabbit PLUS).

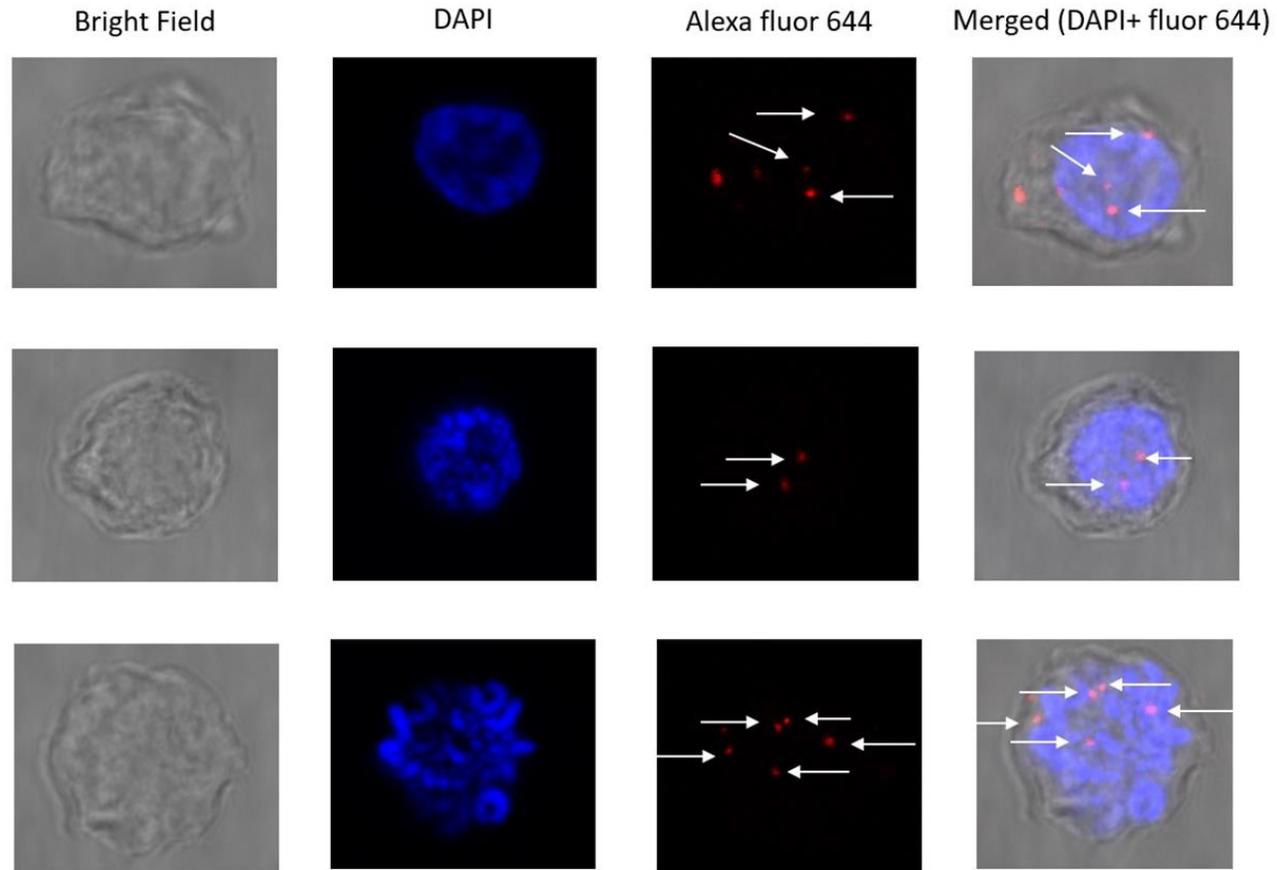


Fig. 3.19. Proximity ligation assay for AID and UBN1 in DT40 Ψ V KO cells. Images demonstrate the PLA signals in DT40 Ψ V KO cells and the proximity between two proteins, AID and UBN1. Nuclei were stained with DAPI. Anti-AID and anti-UBN1 primary antibodies were incubated, followed by incubation with secondary antibodies conjugated to oligonucleotides (PLA Probes anti-mouse MINUS and PLA Probes anti-rabbit PLUS).

3.3 Summary

AID is reported to interact with numerous proteins that influence its functions [64,65,97]. Moreover, AID interaction with RPA and RPA interaction with HIRA protein is well established [63,76]. We explored the possibility of AID interaction with members of the HIRA chaperon complex, such as UBN1, HIRA as well and ASF1a. Interestingly, our *in silico* studies predicted that UBN1 interacts with UBN1. Subsequently, using co-immunoprecipitation we found that AID interacts with UBN1 in Raji as well as DT40 ψ V KO cell lines. Additionally, *in-vitro* pull-down assay also suggests that UBN1 interacts with AID. Moreover, AID and UBN1 are found to be co-localized in the nucleus as detected by double immunofluorescence assay. Proximity ligation assay also validated our finding that AID interacts with UBN1. Moreover, AID interaction with UBN1 may be responsible for recruiting AID chromatin-accessible regions that can influence SHM. This study finds yet another AID interacting partner UBN1, that may influence the AID-mediated antibody diversity.

Chapter 4

Chapter 4

Heterologous expression of hAID in *E. coli* leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored.

4.1 Introduction

Activated B-cells produce affinity matured (by the process of SHM) and switched antibodies (by the process of CSR) to combat the pathogens. Notably, activation-induced cytidine deaminase (AID) is the initiator of SHM as well as CSR. AID gene was discovered by Honjo's group in 1999 [4]. Subsequent years showed hundreds of research groups involved in finding the structure, function, regulation, and mechanism of AID action. AID mRNA is composed of 5 exons (E1 = 1-3, E2 = 4-52, E3 = 53-142, E4 = 143-181, E5 = 182-198), and codes for a 198 amino acid protein [25,98]. Numerous AID splice isoforms are reported, such as AID δ E4a (the first 10 amino acids are missing), AID δ E4 (exon 4 is eliminated), AID δ E3E4AID (exons 3 and 4 are missing), AID-ivs3 (inclusion of intron 3) [25]. The main parts of AID are the N-terminal nuclear localization signal (NLS) sequence, middle catalytic domain, APOBEC-like domain, and C-terminal nuclear export signal sequence. Functional Studies on domains of AID have revealed that the amino-terminal is indispensable for SHM whereas the carboxy-terminal is required for CSR [22]. AID targeting Ig genes required the orchestrated set of events and crosstalk between numerous proteins (the details of processes and interacting partners are already discussed in the review literature as well as the introduction of chapter 3).

AID is a crucial enzyme whose aberrant expression and mistargeting lead to diverse types of pathological conditions [100, 101]. AID overexpression is reported in follicular lymphoma, DLBCL [102], and chronic lymphocytic leukemia [99]. Moreover, with overexpressed AID, cells accumulate somatic mutation in proto-oncogenes like cMYC, Pax5, etc. resulting in the dysregulation of genes, and thereby promoting oncogenesis [103]. Additionally, besides

the full-length AID, distinct types of splice isoforms of AID are reported in lymphomas such as AID δ E4a, AID δ E4, AID δ E3E4 and AID-ivs3. Moreover, the splice variants of AID exhibit a diverse range of activity from hyper SHM or inactivation of CSR, as well as affecting its nuclear localization [24]. Normal B-cell suppresses the generation of an alternative splice variant of AID indicating a crucial regulation that limits the AID mutagenic activity. Jin et.al partly solved the mystery behind the appearance of various types of splice isoforms of AID. Interestingly, RBM5 an RNA binding protein is responsible for AID exon 4 skipping in cell lines [104]. An intriguing observation was that, in malignancies, the cells switch from AID-FL to an alternative splice variant that showed enhanced hyper-mutational activity [25].

AID protein extraction is essential for a better understanding of AID structure as well as function. Recombinant AID extraction from *E. coli* strains [105] as well as insect cell line *SF9* [106] is routinely done and is well documented in the literature. In the quest to perform a biochemical study on AID, we expressed recombinant human AID from *E. coli* cells. Surprisingly, overexpressed had produced two induced bands, one corresponds to the approximate molecular weight of hAID whereas another's band showed a slight difference in migration. Subsequently, this chapter describes the various techniques used to answer the possible reasons behind the appearance of two bands of hAID. We used peptide mapping, western blotting, mutational analysis, changing the solubilization tag, and protein purification to address the appearance of two bands of hAID. We found that despite mutating the putative splice site, switching the tag from Nus A to MBP had produced full-length hAID as well as splice isoform. Additionally, peptide mapping reveals both bands are of hAID, the top band belongs to full-length hAID, whereas the lower band is hAID δ C (has missing residue). Nevertheless, the molecular cues behind the appearance of splice isoform hAID δ C need to be explored whether it is due to RNA splicing or protein splicing or other biological phenomena.

4.2 Results

4.2.1 Expression of WT hAID-His

AID gene from Raji cells (human B-cell line) was cloned into pET43a to give rise to pHAD. *E. coli* cells were induced with 0.1 mM IPTG for 4hrs at 37°C to produce hAID. Incidentally, the expression of hAID was predominantly detected in the pellet (**Fig. 4. 1A**) as an inclusion body may be formed due to the absence of post-translational modification or improper folding of proteins in bacteria. Surprisingly, in addition to the expected band of hAID at ~ 25 kDa, we observed another band at a slightly lower molecular weight. To confirm the presence of His-tag on the two induced proteins, we performed western blotting using an anti-His antibody. Western blot against His-tag using anti-His antibody showed both the induced bands are of overexpressed hAID (**Fig. 4. 1B**). Subsequently, western blot of hAID-His against anti-AID antibody showed only a single band (**Fig. 4. 1C**) corresponding to the size of full-length hAID (~ 25 kDa). There were two distinct bands for the expressed protein, an upper band at the expected size of hAID protein (~ 25 kDa) and a distinct lower band (at ~ 23 kDa) showing a slightly faster migration.

4.2.2 Mass spectral analysis of hAID-His

To further confirm that both induced bands are of AID protein and to get more insight into the identity of two induced bands of protein, LC-MS was performed. The mass spectral analysis of both proteins suggests that they are of hAID with almost 50 % sequence coverage (**Fig. 4.2A** Appendix D). The trypsin digestion map of hAID protein (**Fig. 4.2B**) showed a trypsin site at the c-terminal, so the possibility of retrieving complete coverage is unlikely. This also eliminates any possibility of host cell protein showing an induced band. After comparison with the reported splice isoforms of AID and considering the expression patterns in *E. coli*, the top band is the FL-hAID, whereas the lower band is most likely of hAID with a deletion of 10 amino acids.

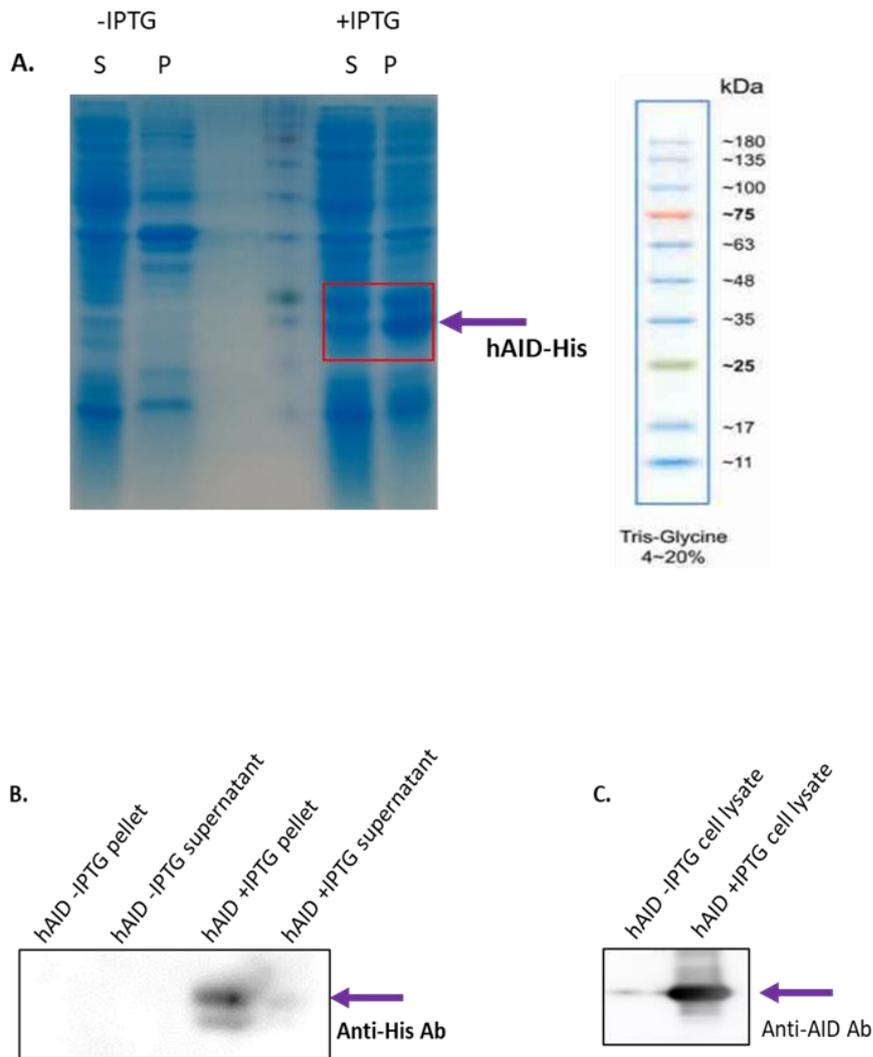


Fig. 4.1. Expression and western blotting of hAID-His. (A) Expression of the hAID-His in *E. coli* cells. lane 1: uninduced supernatant; lane 2: uninduced pellet; lane 3: induced supernatant; lane 4: induced pellet. (B) Western blot of uninduced and induced hAID-His with anti-His antibody. lane 1: uninduced pellet; lane 2: uninduced supernatant; lane 3: induced pellet; lane 4: induced supernatant. (C) Western blot of uninduced and induced hAID-His with anti-AID antibody. lane 1: uninduced cell lysate; lane 2: induced cell lysate.

A.

	10	20	30	40	50	60
hAID	MDSLLMNRK	FLYQFKNVR	AKGRRETYLC	YVVKRRDSAT	SFSLDFGYLR	NKNGCHVELL
hAID_UB	MDSLLMNRK	FLYQFK NVR	AKGRRETYLC	YVVK RRDSAT	SFSLDFGYLR	NKNGCHVELL
hAID_LB	MDSLLMNRK	FLYQFK NVR	AKGRRETYLC	YVVK RRDSAT	SFSLDFGYLR	NKNGCHVELL
	70	80	90	100	110	120
hAID	FLRYISDWDL	DPGRCYRVTW	FTSWSPCYDC	ARHVADFLRG	NPNLSLRIFT	ARLYFCEDRK
hAID_UB	FLRYISDWDL	DPGRCYRVTW	FTSWSPCYDC	ARHVADFLRG	NPNLSLRIFT	ARLYFCEDRK
hAID_LB	FLRYISDWDL	DPGRCYRVTW	FTSWSPCYDC	ARHVADFLRG	NPNLSLRIFT	ARLYFCEDR K
	130	140	150	160	170	180
hAID	AEPEGLRRLH	RAGVQIAIMT	FKDYFYCWNT	FVENHERTEFK	AWEGLHENS	RLSRQLRRIL
hAID_UB	AEPEGLRRLH	RAGVQIAIMT	FKDYFYCWNT	FVENHERTEFK	AWEGLHENS	RLSRQLRRIL
hAID_LB	AEPEGLRRLH	RAGVQIAIMT	FKDYFYCWNT	FVENHERTEFK	AWEGLHENS	RLSRQLRRIL
	190	198				
hAID	LPLYEVDDL	DAFRTLGL				
hAID_UB	LPLYEVDDL	DAFRTLGL				
hAID_LB	LPLYEVDDL	DAFRTLGL				

B.

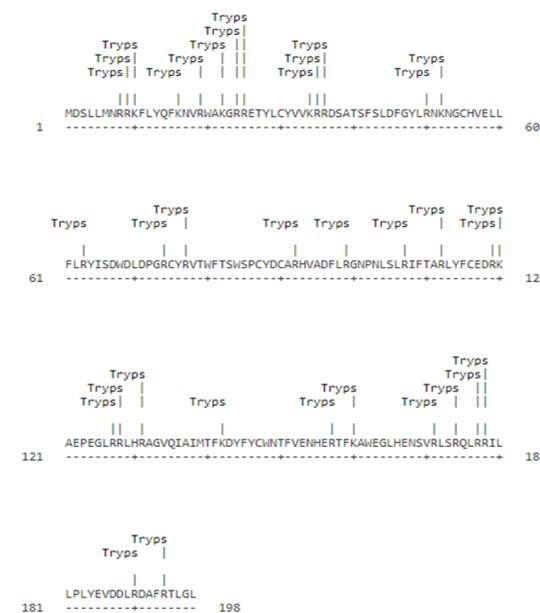


Fig. 4.2. Peptide mapping and trypsin digestion map of hAID. (A) The peptide sequences of the upper and lower bands were aligned with hAID. The red highlighted sequence represents the peptide coverage found in mass spectrometry of the upper and lower band. (B) Trypsin digestion map of the hAID protein sequence.

Moreover, hAID with a deletion of 10 aa from the start of exon 4 is a splice isoform of hAID known as hAID δ E4a and is reported to be expressed in normal B-cell as well as during various types of lymphoma. Surprisingly, *E. coli* cells expressed the hAID splice variant beside FL-hAID from a transformed plasmid containing the cDNA of FL-hAID. *Bacteria* do contain self-catalyzing introns with conserved dinucleotide sequences at 5' and 3' that result in its removal via transesterification reaction [107].

4.2.3 Expression of Nus-hAID-His and His-MBP-hAID

In an attempt to get soluble AID protein, the AID gene was cloned into pET43a to give rise to pNus-hAID, so that AID can be expressed as a fusion protein with Nus solubilisation tag. *E. coli* cells were induced with 0.1 mM IPTG for 4 hrs at 37°C to produce Nus-hAID. As anticipated, adding the N-terminal Nus-tag (a solubilization tag) yielded the Nus-hAID in the soluble as an insoluble fraction (**Fig. 4.3 A**). Very similar to WT hAID, Nus-hAID showed two induced bands at similar expression conditions. One band correspond to the size of Nus-hAID (~ 79 kDa), whereas another band is a slightly smaller size protein. As the NusA protein is prone to proteolysis we asked whether replacing the Nus tag with other reported solublization tag still leads to the expression of two induced bands or not. MBP is well documented in the literature to have strong solubilization effects on many insoluble proteins, as a result, we substituted the Nus tag with the MBP tag to create His-MBP-hAID. As anticipated, His-MBP-hAID overexpression produced two bands similar to overexpression of WThAID and Nus-hAID-His, one band corresponding to the full-length His-MBP-hAID whereas another band is a slightly smaller size protein (**Fig. 4.4 A**). To confirm the affinity of two isoforms of His-tagged Nus-hAID proteins and MBP-hAID, we performed affinity chromatography experiments using Ni-NTA resin. The purification of soluble Nus-hAID-His as well as His-MBP-hAID using Ni-NTA resin again showed two distinct bands of in Nus-hAID-His (**Fig. 4.3 B**) as well as His-MBP-hAID (**Fig. 4.4 B**),

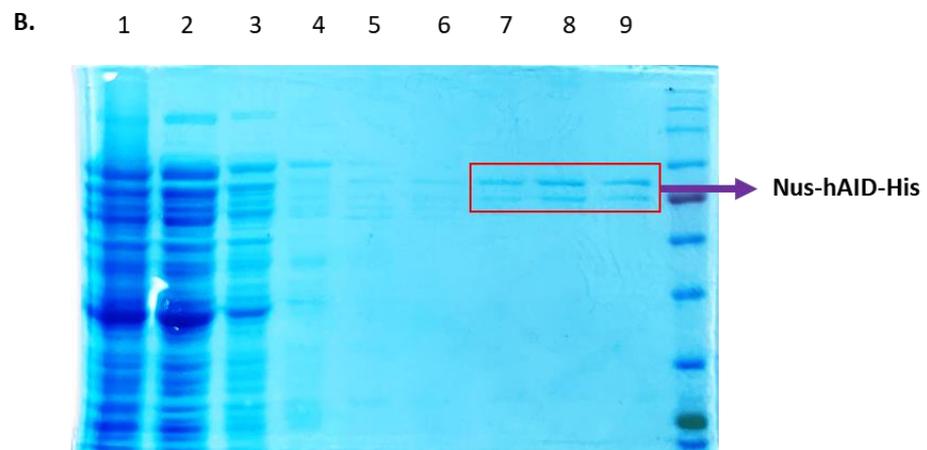
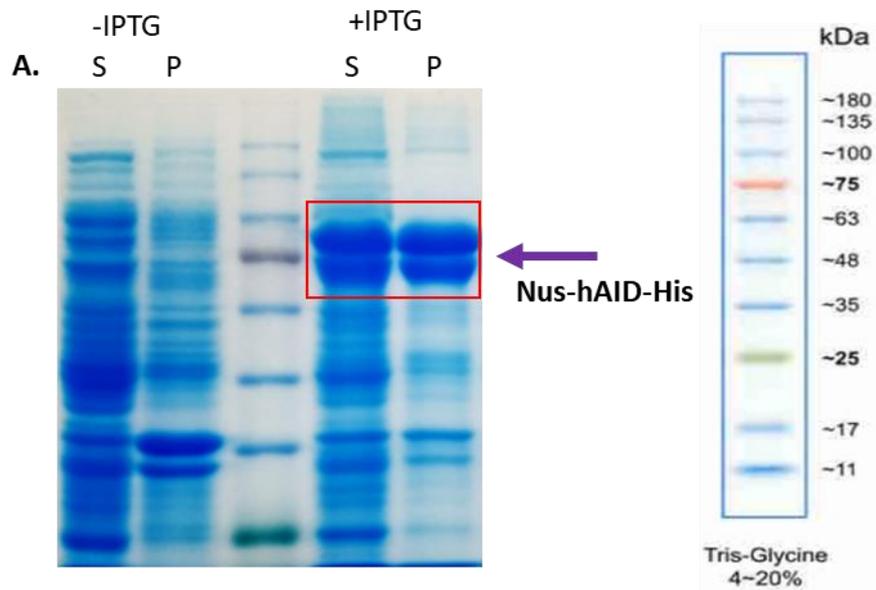


Fig. 4.3. Expression and purification of Nus-hAID-His from *E. coli* cells. (A) Induction of Nus-hAID-His in *E. coli* cells lane 1: uninduced supernatant; lane 2: uninduced pellet; lane 3: induced supernatant; lane 4: induced pellet. (B) Purification of Nus-hAID-His on Ni-NTA gravity column. Lane 1: input sample; lane 2: flow-through; lane 3-6: wash with 20 mM, 40 mM, 60 mM and 80 mM imidazole; lane 7-9: elution with 100 mM, 200 mM and 400 mM imidazole respectively.

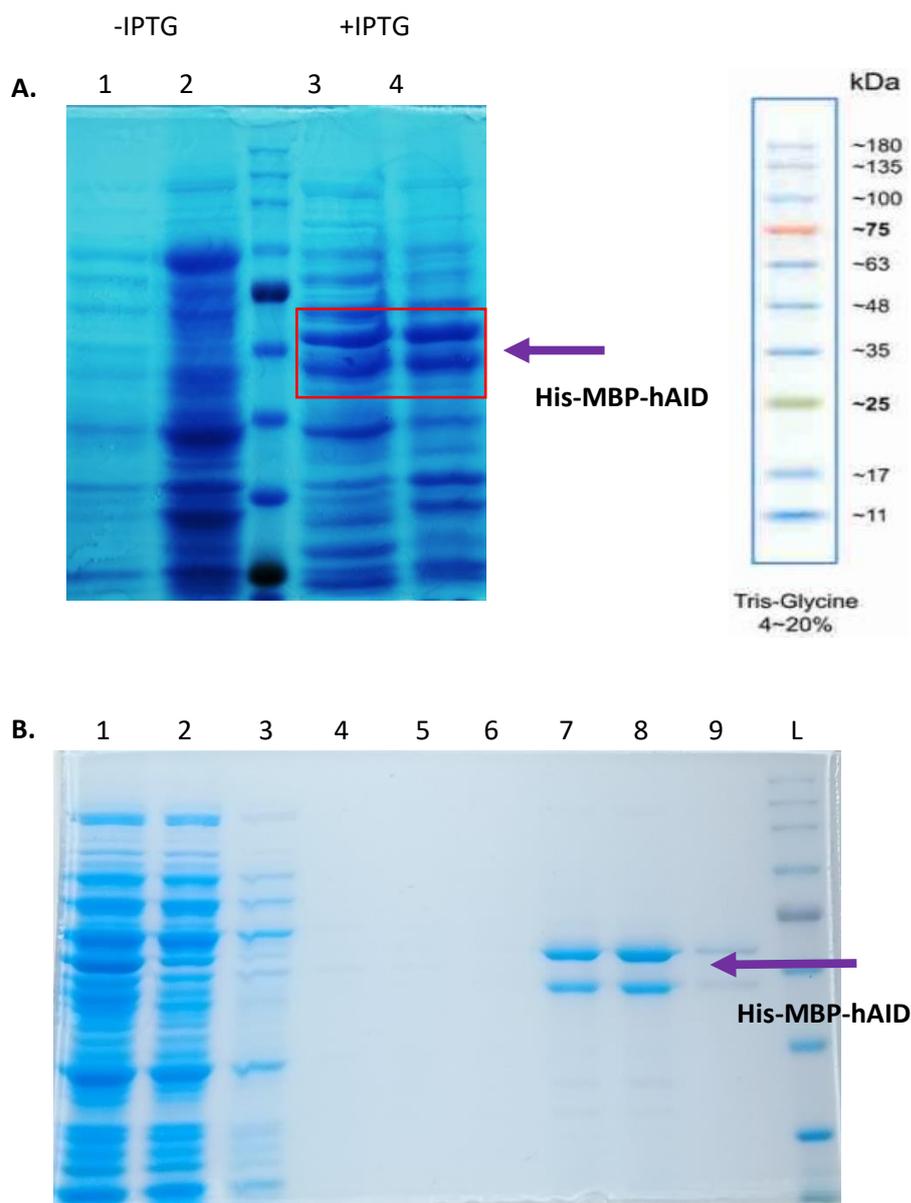


Fig.4.4. Expression and purification of His-MBP-hAID. (A) Expression of His-MBP-hAID in *E. coli* cells. Lane 1: uninduced supernatant; lane 2: uninduced pellet; lane 3: induced supernatant; lane 4: induced pellet. (B) Purification of His-MBP-hAID on Ni-NTA gravity column. Lane 1: input sample; lane 2: flow-through; lane 3-6: wash with 40 mM, 60 mM, 80 mM and 100 mM of imidazole; lane 7-9 elution with 200 mM, 400 mM and 500 mM imidazole respectively.

suggesting that both the proteins are His-tagged and show affinity to Ni-NTA resin. To further validate that the two isoforms of each His-tagged Nus-hAID and MBP-hAID indeed contain His-tag, we performed western blotting using an anti-His antibody. Interestingly, like WT-hAID, both Nus-hAID-His (**Fig. 4.5 A**) as well as His-MBP-hAID (**Fig. 4.6 A**) showed two distinct bands as observed in induction as well as purification SDS-PAGE gel. Likewise, western blot of His-tagged Nus-hAID and MBP-hAID were performed using an anti-AID antibody that produced a single band corresponding to full-length Nus-hAID (**Fig. 4.5 B**) and MBP-hAID (**Fig. 4.6 B**) similar to only hAID. As Western blotting experiments with the anti-AID antibody are only able to give a single band despite both the bands belonging to overexpressed AID, we looked at the epitope region of AID against which the anti-AID antibody was raised. As mentioned by the ThermoFisher manufacturer anti-AID antibody was raised against the c-terminal domain of AID, anti-AID antibodies may be able to recognize only full-length hAID, whereas the lower band may have some missing amino acids at the c-terminal region that in turn abolish the binding site of the antibody. So, the western data for anti-His antibody and anti-AID antibody for both Nus-hAID-His as well as His-MBP-hAID predicts that lower bands may have a few missing amino acids at the c-terminal domain.

4.2.4 hAID and hAID δ E4a

The DNA sequence alignment of hAID and its isoform *hAID δ E4a* showed that 30 bp are missing from 427 to 456 in the latter compared to FL-hAID. The missing region of hAID from 427 to 456 bp is a part of the exon 4 and we assume this region as a putative intron. The start of the putative intron in hAID is marked as a 5' splice site, whereas the end of the intron is named as a 3' splice site (**Fig. 4.7**).

4.2.5 Expression of splice mutant of Nus-hAID-His

As suggested by mass spectral data (**Fig. 4.2 A**) that probably 10 amino acids are missing, we asked whether the mutation at the putative 5' splice site (SS), 3'SS, and 5'3'SS of hAID intron has any effect on

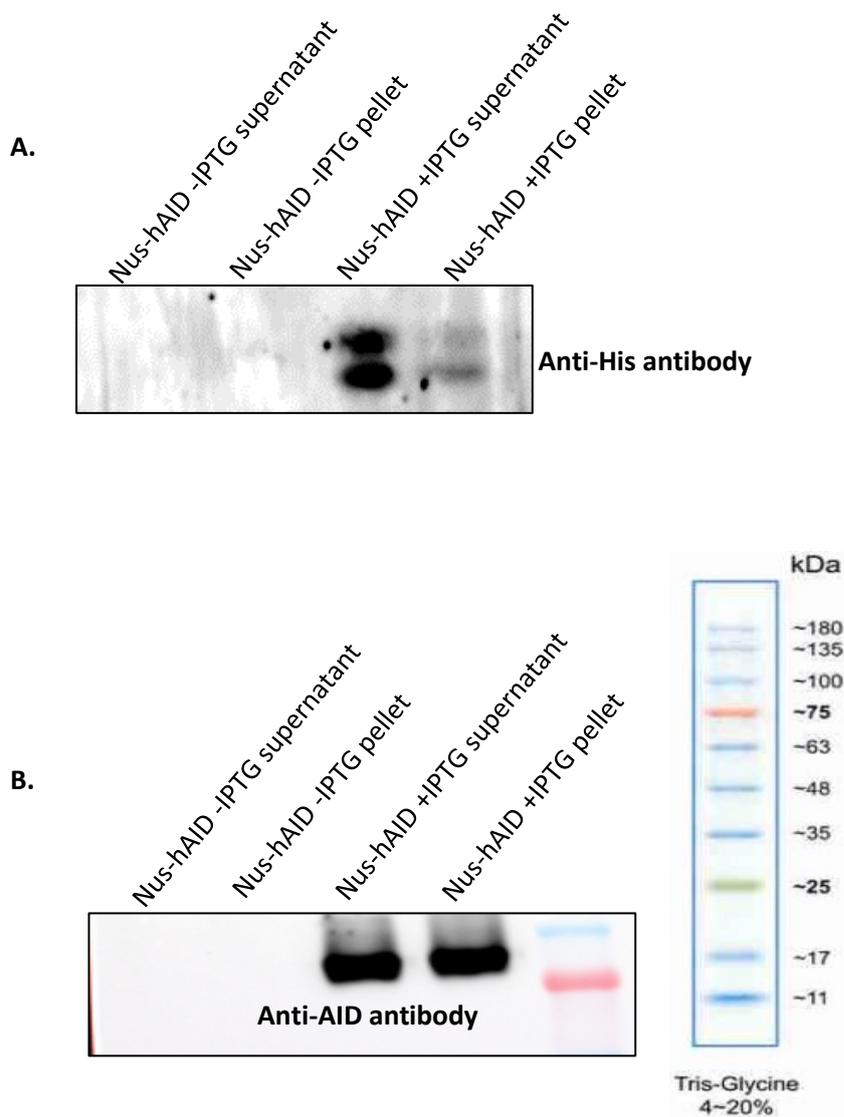


Fig. 4.5. Western blot of Nus-hAID-His using antibody. (A) Western blot of uninduced and induced hAID with anti-His antibody. lane 1: uninduced supernatant; lane 2: uninduced pellet; lane 3: induced supernatant; lane 4: induced pellet. (B) Western blot of uninduced and induced hAID with anti-AID antibody. lane 1: uninduced supernatant, lane 2: uninduced pellet, lane 3: induced supernatant, lane 4: induced pellet.

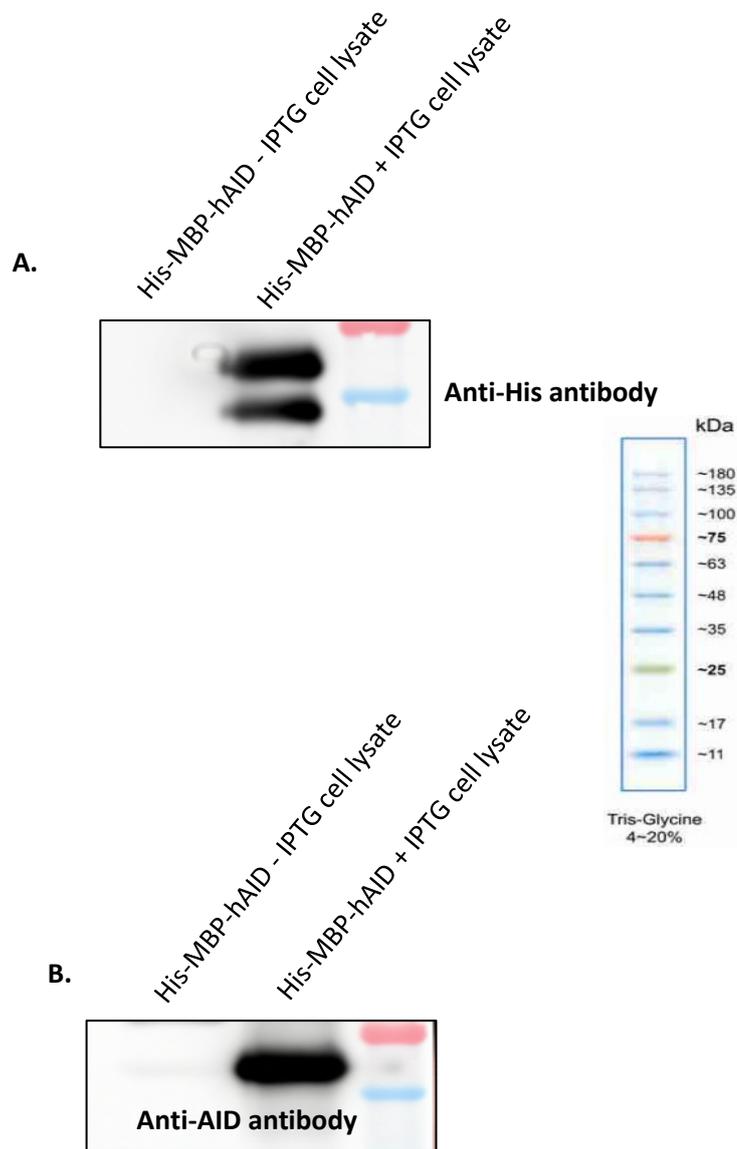


Fig. 4.6. Western blot of His-MBP-hAID. (A) and (B) Western blot of uninduced and induced cell lysate His-MBP-hAID with anti-His antibody and anti-AID antibody, respectively. lane 1: uninduced cell lysate, lane 2: induced cell lysate.

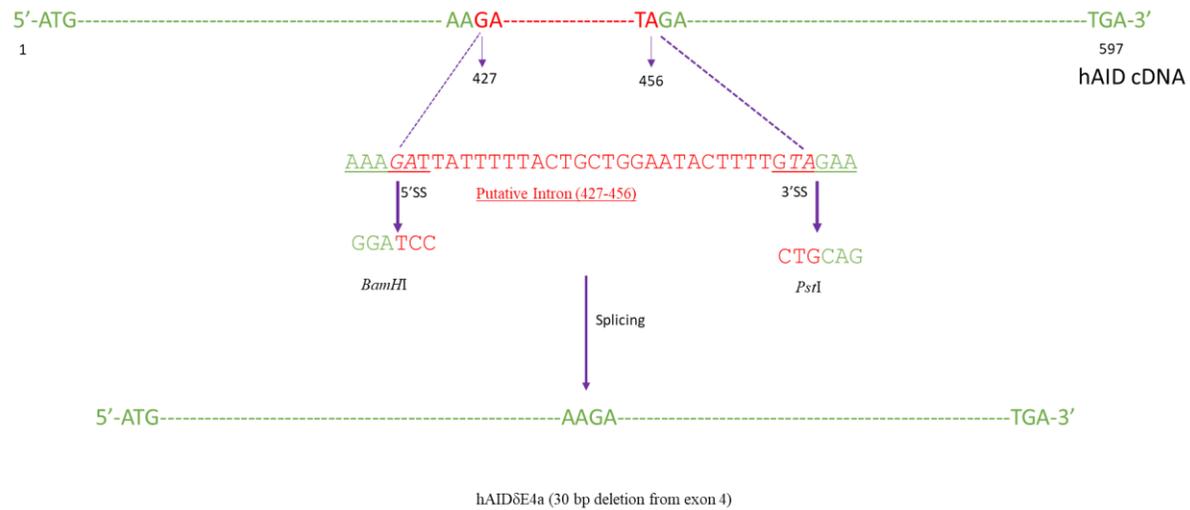


Fig. 4.7. Identification of putative hAID splice site. hAID gene 30 bp (427-456bp called putative intron) and flanking sequence are indicated above. The 5' splice site is mutated to *Bam*HI whereas the 3' splice site was mutated to *Pst*I. Splicing of hAID 30 bp from the start of exon 4 leads to the formation of hAIDΔE4a.

the expression pattern of hAID splice variant. Subsequently, we created three splice junction mutants as explained in the method section and explained in **Fig. S1**. As both hAID and Nus-hAID showed similar expression patterns as two bands appear in each case, the subsequent splice junction mutants study was performed on Nus-hAID. All three independent mutants of recombinant Nus-hAID-His were expressed in *E. coli* cells. The expression profile of 5'SSM, 3' SSM, and 5'3' SSM of Nus-hAID-His showed the bands of FL-hAID as well as hAID splice variants like Nus-hAID (**Fig. 4.8 A**). The mutation in the putative splice site of hAID did not influence the expression of the AID splice isoform. Most likely, only mutation at the putative splice site may not be sufficient to ablate the splicing of hAID. Thus, despite mutation at the splice site, the expression profile of the splice isoform of hAID, *hAID δ E4a* remains unaffected.

4.2.6 PCR from cDNA of Induced and Uninduced Nus-hAID

As peptide mapping data of hAID-His suggested, heterologous expression of hAID in *E. coli* produces FL-hAID as well as its isoform hAID δ E4a. To explore the possibility of any modifications happening at the mRNA levels, leading to the expression of hAID δ E4a in *E. coli*, we isolated mRNA from induced and uninduced Nus-hAID-His bacterial cells and prepared cDNA. Further, using these cDNA, we performed PCR with AID primers PK719 and PK720 (Appendix B) and ran 1.5% agarose gel. PCR data for all the splice mutants suggested the presence of only a band of 145 bp corresponding to the FL-AID whereas we could not detect the band of around 115 bp of hAID δ E4a (**Fig. 4.8 B**).

4.2.7 Mass spectral analysis of Nus-hAID-His and His-MBP-hAID

To further get insight into the two induced bands in the case of Nus-hAID-His and His-MBP-hAID peptide mapping was performed. Interestingly, mass spectrometry peptide mapping data of Nus-hAID-His (**Fig. 4.9**) and His-MBP-hAID (**Fig. 7B**) reveals that the upper band belong to full-length hAID whereas the lower band is also AID but with

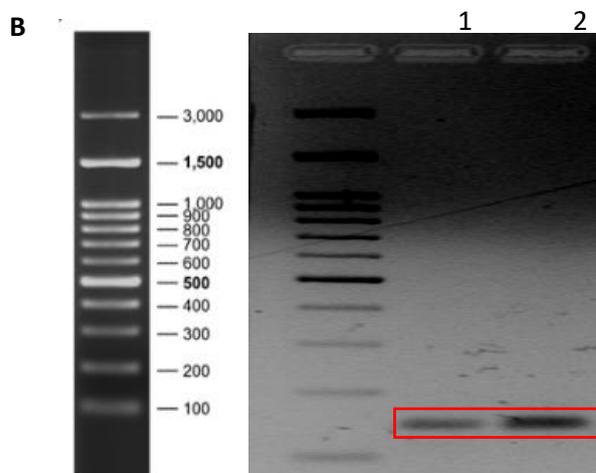
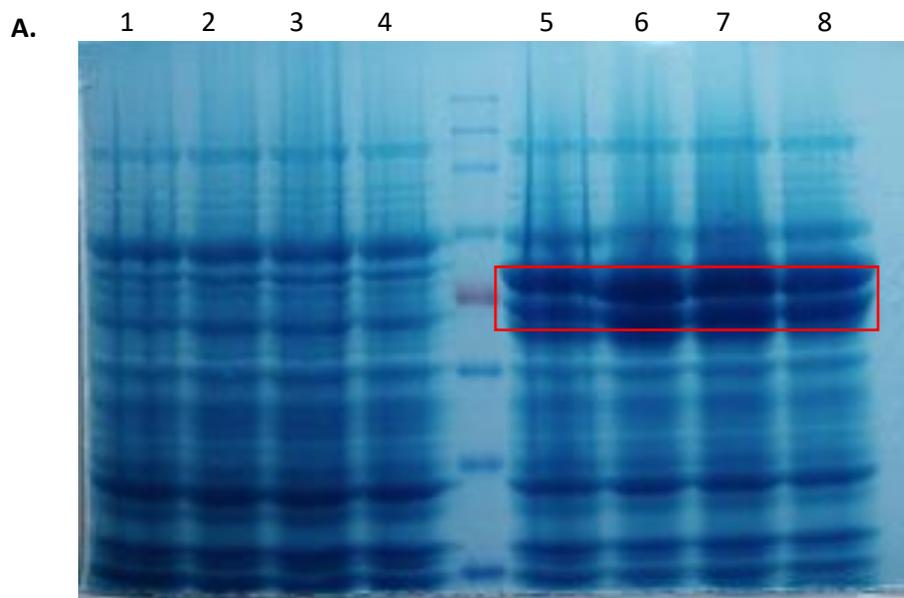


Fig. 4.8. Expression of splice mutant and PCR from cDNA (A) Nus-hAID-His expression with all other splice mutant. Lanes 1, 2, 3, 4 and Lane 5, 6,7,8 showed the uninduced and induced Nus-hAID-His with 5' SM, 3' SM and 5'3' SM respectively. **(B).** PCR from cDNA of uninduced and induced Nus- hAID-His. Lane 1: uninduced cDNA, Lane 2: Induced cDNA.

but have a few missing amino acids. As samples used for mass spectrometry partially purified protein, the chances of the upper band contaminating the lower band are very high. Thus, we neglected peptides of any lower band for which the abundance value is lower than 10% of the upper band. For calculating the number of missing amino acids, we consider only those amino acids which are present in the upper band but missing in the lower band.

Strikingly, Nus-hAID-His lower band showed missing amino acids from the c-terminal domain (**Fig. 4.9**). Thus, the overexpressed Nus-hAID-His showed upper bands (~79 kda) and lower band of (~74 kda). Similarly, His-MBP-hAID peptide sequencing data reveals (**Fig. 4.10**) there is a lot of amino acids, around 60, are missing from the c-terminal of the lower band. So, the overexpressed His-MBP-hAID showed two induced bands. Our mass spectrometry data for Nus-hAID-His and His-MBP-hAID showed a significant number of missing amino acids from the c-terminal domain, which is in collaboration with our western blot data of anti-AID antibody that showed only a single band as c-terminal domain amino acids are indeed absent.

4.2.8 RNA structure prediction of AID mRNA

The RNA structure of hAID was predicted using RNA fold software. RNA folding structure obtained from the software is similar with a minimum free energy state of folding (**Fig. 4.11**). The hAID structure contains a lot of stem and loop base pairing among themselves, which resembles the structure of group II intron [108]. These predictions suggest that the mRNA of hAID has the potential to fold into a structure like self-splicing intron.

	10	20	30	40	50	60								
hAID	MDSL	LLMN	RNRK	FLYQ	FKNVRW	AKGR	RETYLC	YVVK	RRRDSAT	SFSL	DFGYLR	NKNG	CHVELL	
Nus-hAID-His UB	MDSL	LLMN	RNRK	FLYQ	FKNVRW	AKGR	RETYLC	YVVK	RRRDSAT	SFSL	DFGYLR	NKNG	CHVELL	
Nus-hAID-His LB	MDSL	LLMN	RNRK	FLYQ	FKNVRW	AKGR	RETYLC	YVVK	RRRDSAT	SFSL	DFGYLR	NKNG	CHVELL	
	70	80	90	100	110	120								
hAID	FLRY	ISDW	DL	DPGR	CYRV	TW	FTSW	SPCYDC	ARHV	ADFLRG	NPNL	SLRIFT	ARLY	FCEDRK
Nus-hAID-His UB	FLRY	ISDW	DL	DPGR	CYRV	VTW	FTSW	SPCYDC	ARHV	ADFLRG	NPNL	SLRIFT	ARLY	FCEDRK
Nus-hAID-His LB	FLRY	ISDW	DL	DPGR	CYRV	VTW	FTSW	SPCYDC	ARHV	ADFLRG	NPNL	SLRIFT	ARLY	FCEDRK
	130	140	150	160	170	180								
hAID	AEPE	GLRRLH	RAGV	QIAIMT	FKDY	FYCWNT	FVEN	HERTFK	AWEG	LHENS	SV	RLSR	QLRR	IL
Nus-hAID-His UB	AEPE	GLRRLH	RAGV	QIAIMT	FKDY	FYCWNT	FVEN	HERTFK	AWEG	LHENS	SV	RLSR	QLRR	IL
Nus-hAID-His LB	AEPE	GLRRLH	RAGV	QIAIMT	FKDY	FYCWNT	FVEN	HERTFK	AWEG	LHENS	SV	RLSR	QLRR	IL
	190	198												
hAID	LPLY	EVDDL	R	DAFR	TLGL									
Nus-hAID-His UB	LPLY	EVDDL	R	DAFR	TLGL									
Nus-hAID-His LB	LPLY	EVDDL	R	DAFR	TLGL									

Fig. 4.9. Peptide sequencing of Nus-hAID-His in which upper and lower bands are aligned with hAID. The red letters denote the sequence covered in peptide sequencing and dotted boxes show the missing sequencing between the upper and lower band.

	10	20	30	40	50	60
hAID	MDSL	LMNR	RK	FLYQ	FKNV	RW
His-MBP-hAID UB	MDSL	LMNR	RK	FLYQ	FKNV	RW
His-MBP-hAID LB	MDSL	LMNR	RK	FLYQ	FKNV	RW
	70	80	90	100	110	120
hAID	FLRY	ISDW	DL	DPGR	CYRV	TW
His-MBP-hAID UB	FLRY	ISDW	DL	DPGR	CYRV	TW
His-MBP-hAID LB	FLRY	ISDW	DL	DPGR	CYRV	TW
	130	140	150	160	170	180
hAID	AEPE	GLRR	LH	RAGV	QIAIM	T
His-MBP-hAID UB	AEPE	GLRR	LH	RAGV	QIAIM	T
His-MBP-hAID LB	AEPE	GLRR	LH	RAGV	QIAIM	T
	190	198				
hAID	LPLY	EVDDL	R	DAFR	TLGL	
His-MBP-hAID UB	LPLY	EVDDL	R	DAFR	TLGL	
His-MBP-hAID LB	LPLY	EVDDL	R	DAFR	TLGL	

Fig. 4.10. Peptide sequencing of His-MBP-hAID in which upper and lower bands are aligned with hAID. The red letters denote the sequence covered in peptide sequencing and dotted boxes show the missing sequencing between the upper and lower band.

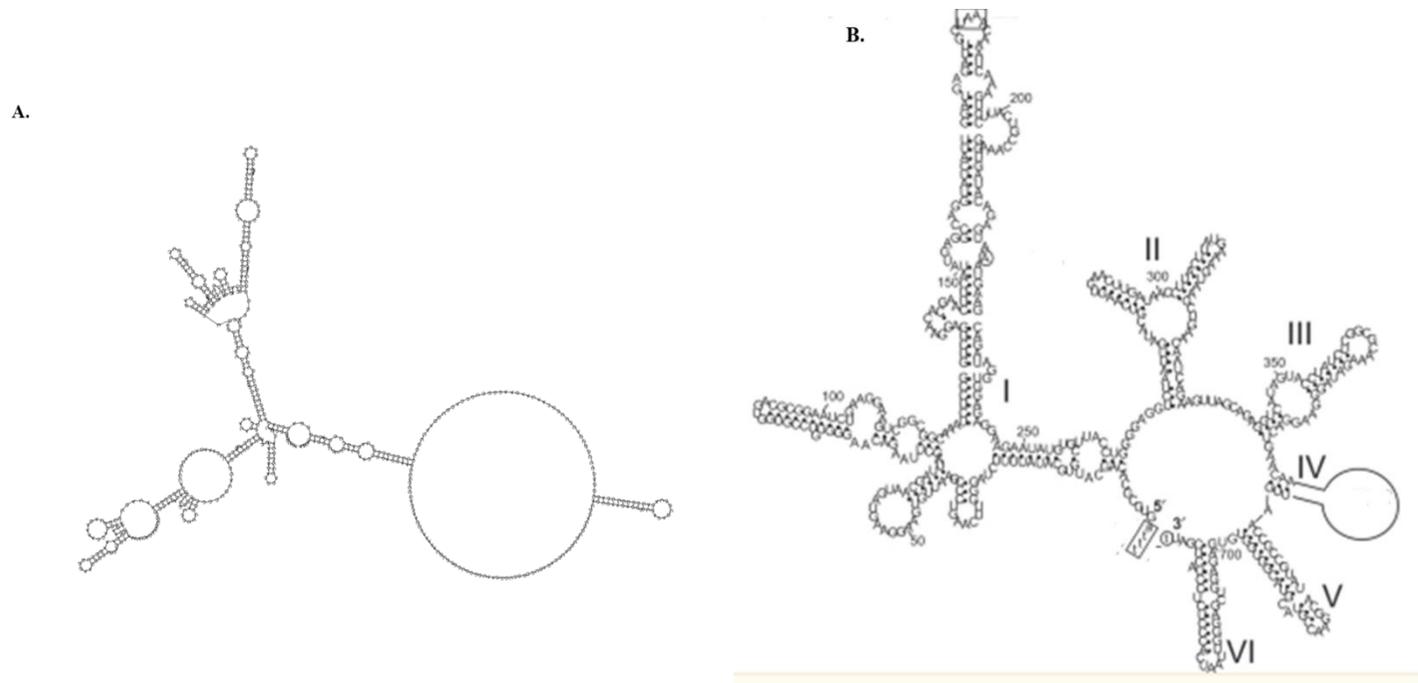


Fig. 4.11. mRNA structure of hAID as predicted by RNA fold server (A) mRNA structure of AID as predicted by RNA fold server. (B) A group II introns from *Bacillus cereus*. The AID structure is similar to group II introns.

4.3 Summary

AID occupies the central place in mediating SHM as well as CSR. Remarkably, the heterologous expression of hAID in *E. coli* cells yielded two bands. Furthermore, peptide mapping showed that the upper band is full-length hAID whereas the lower band is a splice isoform of hAID (contains a few missing residues). Astonishingly, AID splice isoform AID δ E4a is well documented in the literature to be expressed in B-cell malignancy [25]. However, the mutation of the putative splice site of the intron still yielded two hAID bands that suggest these mutations do not influence the appearance of the splice isoform of hAID. Western blotting of overexpressed hAID also showed two bands with an anti-His antibody which confirms that both bands have intact C-terminus. Similarly, the addition of a solubilization tag either Nus tagged or MBP yielded two bands in protein expression and purification as well as western blotting. Finally, peptide mapping of Nus-hAID-His and His-MBP-hAID showed that the lower bands are hAID δ C. The present study is the first report that suggests the hAID overexpression in *E. coli* produces a splice isoform hAID δ C.

Chapter 5

Chapter 5

SRSF1-3 interaction with TOP1 likely inhibits its activity

5.1 Introduction

AID mutagenic activity for SHM and CSR at the Ig locus is directly linked to transcription [48]. Additionally, RNA PolII pausing and termination also promote AID activity [50]. Normally, transcription causes local denaturation of double-stranded DNA, which in turn promotes the formation of single-stranded DNA. Interestingly, this transcription-induced single-stranded DNA acts as a substrate for AID. RNA polII progression creates negative and positive supercoiling of DNA downstream and upstream of its movements. Finally, RNA polII-mediated negative supercoiling results in the creation of single-stranded patches of about 6-7 nucleotides that directly coincide with AID activity [49].

Transcription induces the supercoiling of DNA and causes torsional strain that must be resolved for the smooth movement of RNA PolII. Incidentally, the supercoiling of DNA is relieved by topoisomerase 1 (TOP1), a globally present enzyme in all vertebrates. Astonishingly, Kobayashi et.al, for the first time demonstrated that AID activity is linked with TOP1 [74]. Furthermore, AID downregulates the expression of TOP1, which in turn assists the CSR by enriching the cleavage of the switch region. Nevertheless, the complete shutdown of TOP1 activity by CPT inhibits the S region's cleavage. Interestingly, the role of TOP1 is not only restricted to CSR but its involvement in SHM is also well documented. Likewise, AID was also reported to reduce TOP1 levels in SHM-undergoing cells. The reduction of TOP1 levels enhances the frequency of somatic mutation in B-cells [75].

Notably, AID is also reported to interact with numerous factors that influence antigen-dependent antibody diversity one of which is SRSF1-3 [66,71]. SRSF1 is a splice isoform of a protein involved in alternative splicing and RNA metabolism, SRSF1. SRSF1-3 importance

is not only restricted to its interaction with AID but being a splicing regulator, plays a diverse role that affects antibody diversity. Remarkably, the chicken cell lines devoid of SRSF1-3 failed to undergo SHM which highlights the significance of SRSF1-3 in SHM [71]. Recently, SRSF1-3 knock-in clones upregulated the expression of chromatin modifiers UBN1 as well as SATB2 which may have a role in antibody diversity. Additionally, SRSF1-3 was also involved in changing the splicing pattern of many genes that are directly or indirectly associated with AID [72].

SRSF1 is found to interact with TOP1, which is reported to inhibit the activity of the latter [77]. Likewise, SRSF1-3 was also found to interact with TOP1 [66]. Nevertheless, the influence of SRSF1-3 on the TOP1 activity is not documented. Our current study using computational methods highlights that SRSF1-3 indeed stabilizes the TOP1 structure. Moreover, the SRSF1-3 is most likely to influence the TOP1 activity, which may have a direct effect on SHM/CSR.

5.2 Results

5.2.1 *In-silico* studies of SRSF1-3 and TOP1 interaction

The conformational stability of a protein-protein complex depends on the intermolecular interactions and solvent conditions around the protein. In the absence of a complex structure of TOP1 with SRSF1-3, and to shed insights into their binding mechanism and conformational changes in TOP1 upon its interaction with the SRSF1-3, therefore, we investigated the overall conformational dynamics and stability of the complex using the MD simulations. Moreover, the MMPBSA scheme was utilized to further evaluate their protein-protein binding pose, solvent effect, and binding free energy of the complex. The human TOP1 is composed of 765 amino acids (**Fig. 5.1 A**), whereas SRSF1-3 is 244 amino acids long (**Fig. 5.1 B**). To estimate the molecular mechanism of the TOP1 and SRSF1-3, an MD simulation was conducted for 600 ns. We monitored the time evolution of the root-mean-square deviations (RMSD) of C α atoms of both proteins as well as compared with the apo TOP1, suggesting that in the final production runs of 600 ns, the stable conformation of protein was achieved. Furthermore, we investigated the overall structural stability and calculated the probability density of each protein as well as the apo TOP1 throughout the MD simulation as shown (**Fig. 5.2 A**). It reveals that compared to the apo TOP1, the interaction with the SRSF1-3 stabilizes the topoisomerase, as seen by the sharp peak of the complex. In addition, it shows that TOP1 is quite stable compared to the SRSF1-3 in the complex, and a clear picture of the stabilization of TOP1 can be observed by the sharp peak in contrast to the flattened peak of apo TOP1. Also, the different regions of the TOP1 in its complex system were compared with the apo TOP1 (**Fig. 5.2 B**). It suggests that overall, the structure got stabilized upon its interaction with the SRSF1-3 as can be seen from the shift in the peak of different regions of TOP1 in the complex compared with the apo system. Especially, the CT, cap, and subdomain of the topoisomerase showed a significant difference suggesting that these regions mainly interact with SRSF1-3.

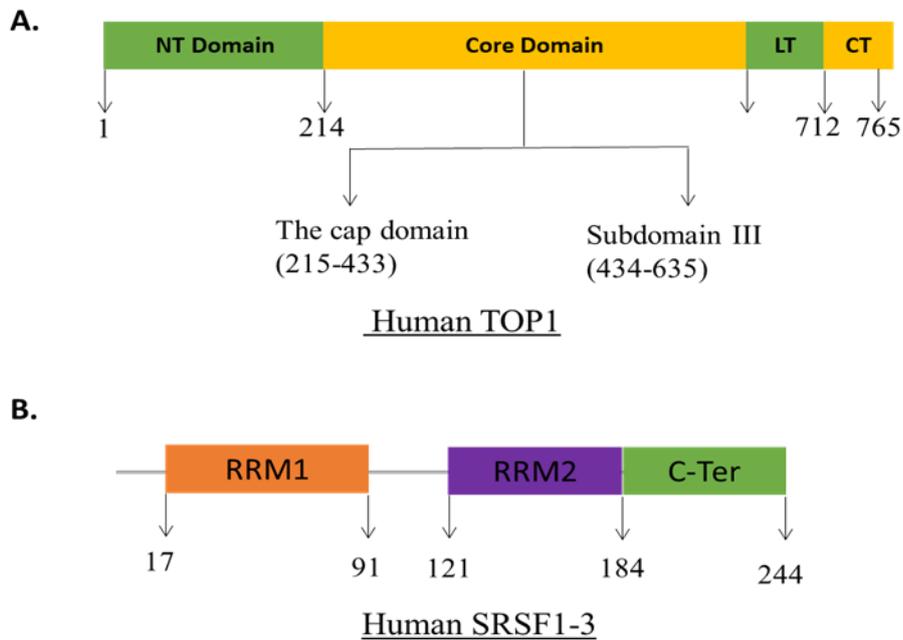


Fig. 5.1. Domain organization of TOP1 and SRSF1-3. (A) The sequence of human TOP1 domain organization is N-terminal domain (NT domain), core domain, linker domain (LT) and C-terminal domain. (B) The sequence of human SRSF1-3 domain organization is RNA recognition motif 1 (RRM1), spacer region, RRM2, and C-terminal domain (C-ter).

In addition, we also evaluated the different regions and stability of SRSF1-3, such as RRM1, RRM2, spacer, and C-terminal domain (CTD) throughout the MD simulation (**Fig. 5.2 C**). It reveals that RRM1 and RRM2 of the SRSF1-3 are more stable than the spacer and CTD, favouring their interaction with the TOP1. Next, we investigated each residual fluctuation of both proteins and calculated the root-mean-square fluctuations (RMSF) throughout the MD (**Fig. 5.2 D**). The fluctuations of TOP1 significantly reduce upon binding with the SRSF1-3, especially in regions, such as cap (residue numbers), subdomain (residue numbers), and LT (residue numbers). It further supports our finding that these regions are involved in the interaction with the SRSF1-3.

Further, the structural compactness of systems can be accessed by the radius of gyration (R_g) (**Fig. 5.3 A**). As compared to the flat peak seen for the apo TOP1, which showed variability, in the case of the complex, the compactness of the TOP1 increases, suggesting conformational rearrangements of protein upon its interaction with the SRSF1-3. Additionally, the solvent-accessible surface area (SASA) was computed (**Fig. 5.3 B**). The probability density graph showed a reduction of $\sim 2000 \text{ \AA}^2$ after complexation. It reveals that the accessibility of solvent is reduced, which enhances the interaction between proteins. To elucidate the binding mechanism in more detail, molecular mechanics Poisson Boltzmann surface area (MM/PBSA) was used throughout the MD trajectory. Different components of binding free energy are also shown graphically in **Fig. 3A**. The binding free energy between proteins is -125.18 kcal/mol . It is evident (Table 5.1) that the interaction between TOP1 and SRSF1-3 is mainly governed by the hydrophobic interaction (-145.98 kcal/mol). The polar and non-polar solvation-free energy also favours the binding however, the electrostatic interaction energy (130.37 kcal/mol) between the two proteins binding disfavours their interactions. Also, to evaluate the key residues in the process, the per-residual decomposition of binding free energy was calculated (**Fig. 5.4**). The key residues involved in binding include T270

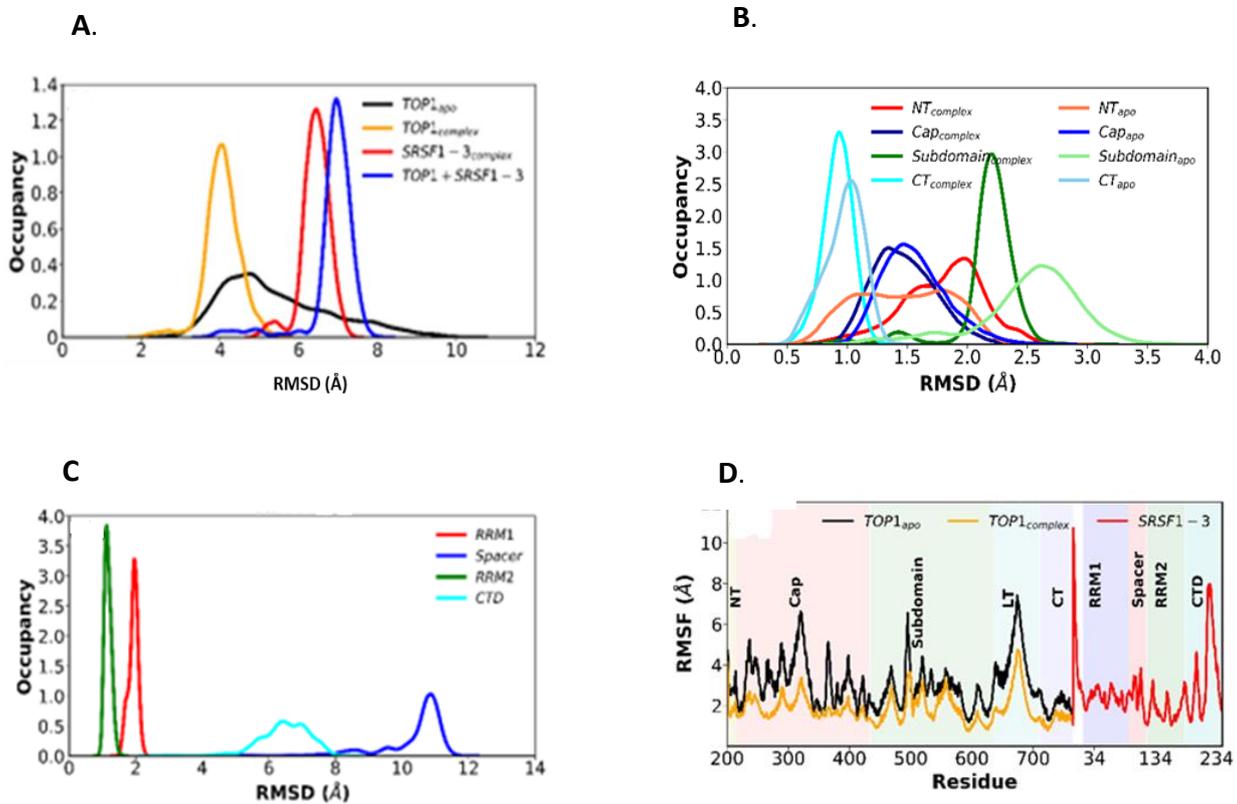


Fig. 5.2. Structural dynamics conformation of TOP1 bound and unbound to SRSF1-3. (A) The root-mean-square deviations (RMSD) of the backbone of TOP1 complex and apo, (B) The different regions of the TOP1 RMSD of both apo and complex, (C) The different regions of RMSD of SRSF1-3 in the complex, (D) the root-mean-square fluctuations (RMSF) of apo and complex.

K638, E641, E702 from TOP1 and R118, E161, and R210 from the SRSF1-3. This indicates that the main interacting residues come from the cap, subdomain, and LT domains of TOP1 and RRM1, RRM2, and CTD of SRSF1-3. In this way, firstly using the computational approach, we would be able to shed light on the interaction mechanisms between the TOPI and SRSF1-3 that may aid in future studies related to them.

5.2.2 Expression of His-SRSF1-3 and His-SRSF1-3ΔC

The human SRSF1-3 gene was cloned into a vector downstream to the N-terminal His-tag to give rise to His-SRSF1-3. Subsequently, the expression of His-SRSF1-3 at 37°C/4hrs as well as 16° C /18 hrs showed the expression in the pellet but not in the supernatant fraction (**Fig. 5.5**). The reason for SRSF1-3 expressed in the pellet is most likely due to the improper folding of proteins or missing a typical post-translational modification in *E. coli* cells which in turn promote the formation of an inclusion body. Furthermore, to express the SRSF1-3 in the soluble fraction, we removed the c-terminal portion and cloned it as mentioned above to create His-SRSF1-3ΔC. Removing the c-terminal portion of some proteins enhances the solubility of the protein and is well-reported in the literature. Unexpectedly, His-SRSF1-3ΔCRS expression was also seen in the pellet but not in supernatant like His-SRSF1-3 (**Fig. 5.6**). Probably, in this case, mere removal for the c-terminal region is not adequate to drag SRSF1-3 in the soluble fraction.

5.2.3 Expression, and purification of His-MBP-SRSF1-3

The SRSF1-3 gene was cloned into a vector downstream of the His-MBP-TEV protease site to give rise to His-MBP-SRSF1-3. As anticipated, the expression of His-MBP-SRSF1-3 at 16°C/ 18 hrs was seen in the soluble fraction. MBP is a well-reported solubilization tag that enhances solubility as well as the yield of the fusion partner. Subsequently, as His-MBP-SRSF1-3 contains an N-terminal His tag, the affinity purification of His-MBP-SRSF1-3 was done on Ni-NTA resin. Further, His-MBP-SRSF1-3 binds to Ni-NTA beads and is eluted at 500 mM imidazole (**Fig. 5.7**). The purity of eluted His-MBP-SRSF1-3 was

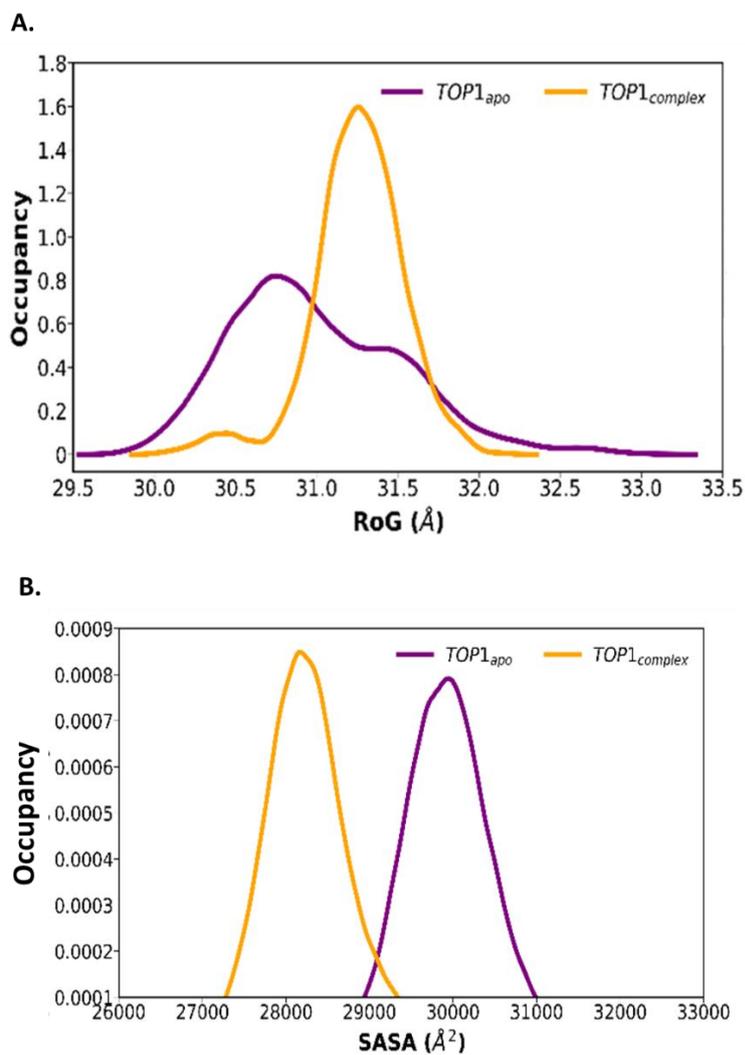


Fig. 5.3. Compactness and solvent accessibility analysis of TOP1 bound and unbound to SRSF1-3. (A) The probability density of radius of gyration (RoG) of both apo and complex, (B) The solvent-accessible-surface-area (SASA) of both apo and complex in MD simulations.

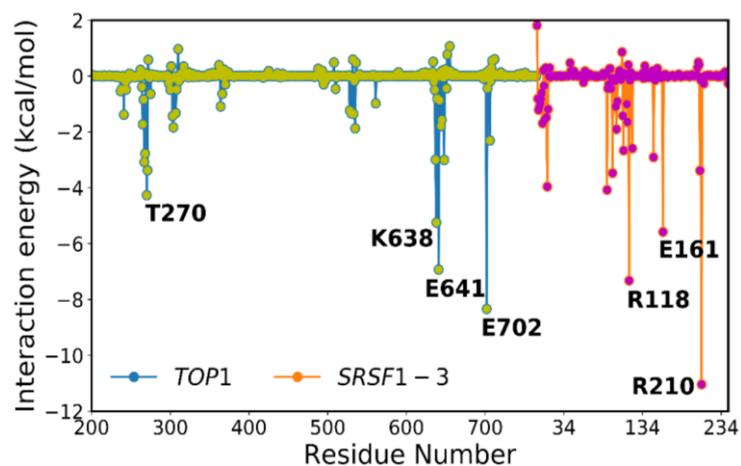


Fig. 5.4 The per-residual decomposition of binding free energy of the TOP1 with SRSF1-3 complex obtained from MMGBSA.

	ΔE_{elec}	ΔE_{vdW}	ΔG_{np}	ΔG_{pol}	ΔE^a MM	ΔG^b_{solv}	ΔG^c_{bind}
TOPI-	130.37	-	-	-	-	-	-
SRSF1-		145.98	18.45	91.10	15.61	109.56	125.18
3							
complex							
$^a\Delta E_{\text{MM}} = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}}, ^bG_{\text{solv}} = \Delta G_{\text{np}} + \Delta G_{\text{pol}}, ^c\Delta G_{\text{bind}} = \Delta E_{\text{vdW}} + \Delta E_{\text{elec}} + \Delta G_{\text{np}} + \Delta G_{\text{pol}}$							

Table 5.1 The different components of binding free energy obtained from MMPBSA for the TOP1 with SRSF1-3 complex.

analysed on SDS-PAGE which is approximately 90%. Finally, the His-MBP-SRSF1-3 was treated with TEV protease to remove the His-MBP from His-MBP-SRSF1-3 but as soon as the tag was removed the SRSF1-3 was precipitated (**Fig. 5.8**). Thus, we used purified and buffer exchanged His-MBP-SRSF1-3 for TOP1 activity assay.

5.2.4 TOP1 inhibition assay

To study the effect of His-MBP-SRSF1-3 on human TOP1, the plasmid relaxation assay was performed. Subsequently, supercoiled plasmid DNA, human TOP1 and the varying molar ratio of His-MBP-SRSF1-3 were mixed. Further, the TOP1 activity in the presence or absence of His-MBP-SRSF1-3 was similar. Incidentally, the plasmid relaxation assay suggested that His-MBP-SRSF1-3 do not influence the TOP1 activity (**Fig. 5.9**). It may happen that the presence of N-terminal His-MBP may have affected the interaction of SRSF1-3 with TOP1 which results in no change in the TOP1 activity profile even in the presence of SRSF1-3.

5.2.5 Solubilization, purification, refolding and dialysis of His-SRSF1-3

As His-MBP-SRSF1-3 doesn't influence the activity of TOP1 activity we performed solubilization, purification, refolding, and dialysis of His-SRSF1-3. His-SRSF1-3 was solubilized by treatment with 6M GuHCl and percentage solubility was measured by SDS-PAGE. GuHCl solubilized His-SRSF1-3 was applied to Ni-NTA resin and eluted at 80 mM, 100 mM, and 250 mM imidazole (**Fig. 5.10**). Subsequently, purified His-SRSF1-3 was refolded by dilution methods that yielded approximately 50% refolded His-SRSF1-3. Further, His-SRSF1-3 was dialyzed and concentrated but the SDS-PAGE reveals that the yield was too low for performing a plasmid relaxation assay. We are optimizing the buffer condition that can help in recovering a higher concentration of His-SRSF1-3.

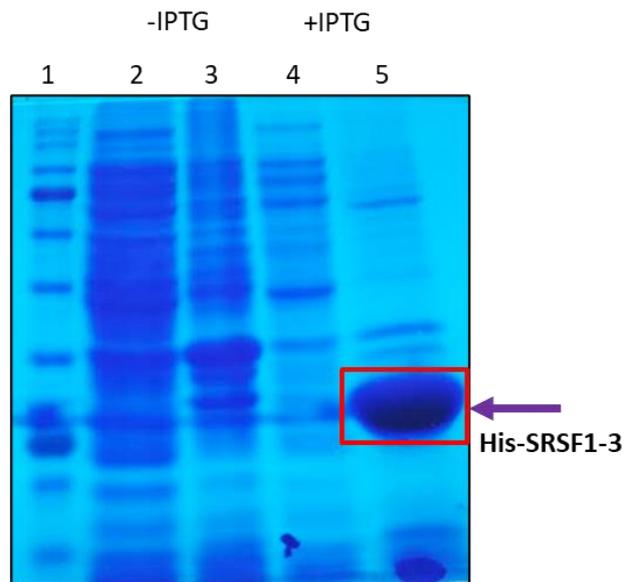


Fig. 5.5. Expression of His-SRSF1-3. Lane 1: pre-stained ladder; lane 2: uninduced supernatant; lane 3: uninduced pellet; lane 4: induced supernatant; lane 5: induced pellet.

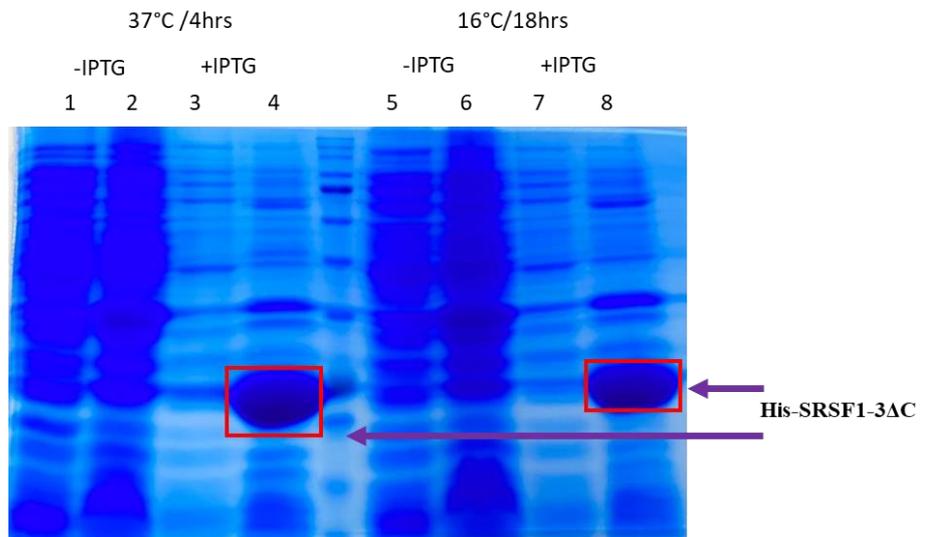


Fig. 5.6. Expression of His-SRSF1-3ΔC. The induction of His-SRSF1-3ΔC was performed at 37°C for 4 hrs as well as 16°C for 4 hrs. lane 1: uninduced supernatant at 37°C; lane 2: uninduced pellet at 37°C; lane 3: induced supernatant at 37°C; lane 4: induced pellet at 37°C; lane 5: uninduced supernatant at 16°C; lane 6: uninduced pellet at 16°C; lane 7: induced supernatant at 16°C; lane 8: induced pellet at 16°C.

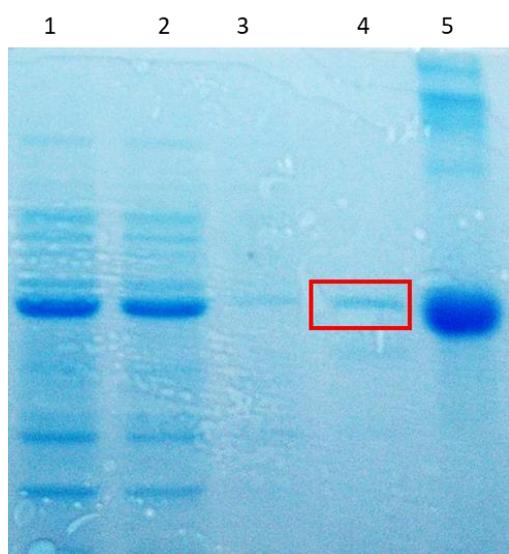


Fig. 5.7. Purification of His-MBP-SRSF1-3 by using Ni-NTA beads. Lane 1: Input sample; lane 2: flow-through; lane 3: wash using 40 mM imidazole; lane 4: elution with 500 mM imidazole; lane 5: BSA.

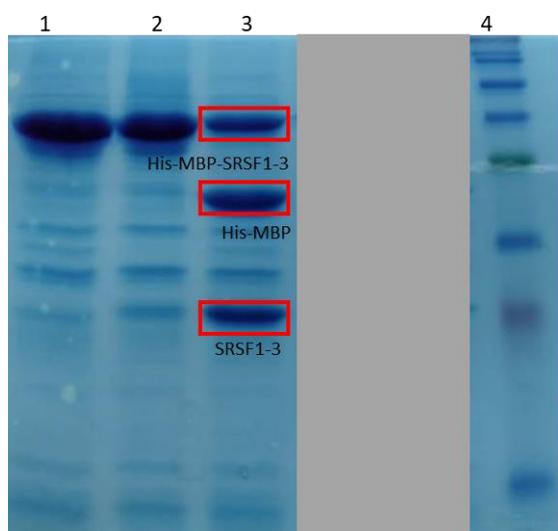


Fig. 5.8. Removal of solubilization tag by treatment with TEV protease. Lane 1: Partially purified and buffer exchanged His-MBP-SRSF1-3; lane 2: His-MBP-SRSF1-3 with TEV protease input; lane 3: His-MBP-SRSF1-3 treated with TEV protease; lane 4: pre-stained protein ladder.

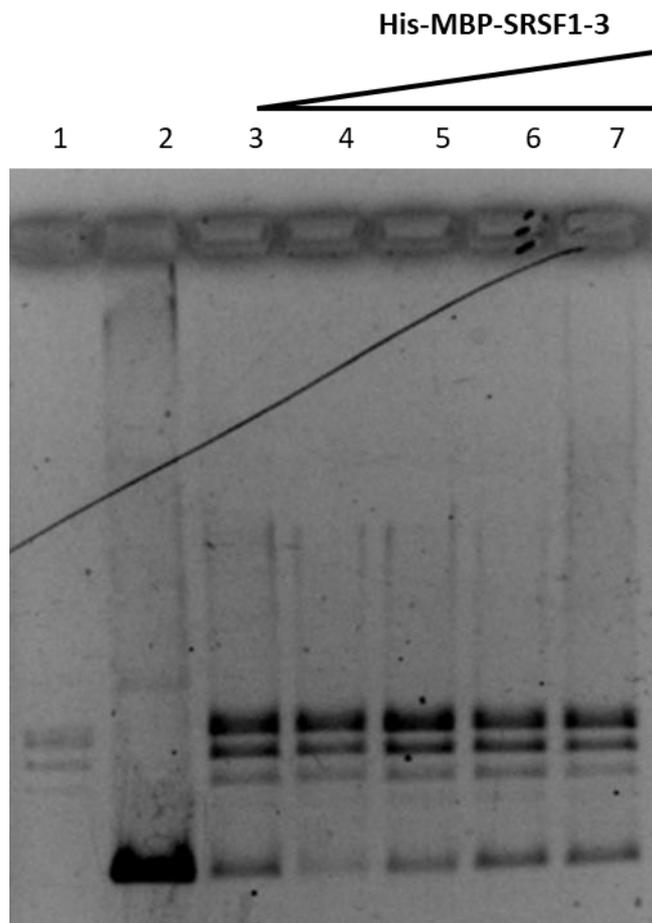


Fig. 5.9. Investigation of consequences of SRSF1-3 on TOP1 plasmid relaxation activity. Lane 1: marker DNA; lane 2: supercoiled DNA (SS); lane 3: SS DNA mixed with TOP1; lane 4-7: SS DNA, TOP1 and increasing concentration of His-MBP-SRSF1-3.

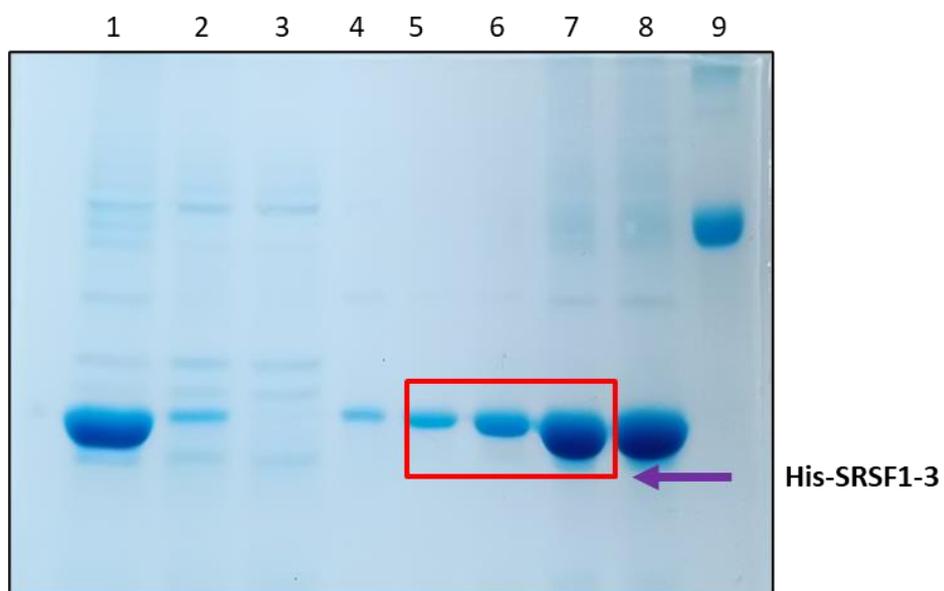


Fig5.10 Purification of solubilized His-SRSF1-3. Lane 1: Solubilized His-SRSF1-3 as input; lane 2: flowthrough; lane 3 and 4: wash with 20 mM and 40 mM imidazole; lane 5 to 7: elution with 80 mM, 100 mM, 250 mM imidazole; lane 8: stripping; lane 9: BSA.

5.3 Summary

SRSF1-3 protein is a splice isoform of the very first member of the serine-arginine-rich protein family SRSF1, which is pivotal for SHM [70,71]. Interestingly, the literature suggests SRSF1-3 was found to interact with yet another crucial factor of SHM/CSR, TOP1 [66]. Our current study uses *in-silico* methods of molecular dynamics simulations showing that SRSF1-3 in complex with TOP1, stabilizes the structural fluctuations of TOP1. Additionally, the binding free energy of binding TOP1 and SRSF1-3 is -125.18 kcal/mol, where hydrophobic interaction mainly contributes to the binding. The RRM1 and RRM2 domains of SRSF1-3, and the cap, subdomain and LT domain of TOP1 were involved in the interaction. Moreover, per residue fluctuations of TOP1 are significantly decreased in the presence of SRSF1-3, which suggests SRSF1-3 increases the stability of TOP1. Furthermore, we explored the effect of SRSF1-3 on the TOP1 DNA relaxation study. As His-SRSF1-3 was expressed in inclusion bodies we expressed SRSF1-3 as His-MBP-SRSF1-3 (MBP solubilization tag), used for TOP1 activity assay. Unexpectedly, His-MBP-SRSF1-3 did not affect TOP1 activity, it can be partly explained as SRSF1-3 domains in His-MBP-SRSF1-3 may not be accessible for interaction with TOP1 for mediating its effect. Incidentally, we expressed, purified, refolded and buffer exchanged His-SRSF1-3 and will show its effect on TOP1 activity. Moreover, SRSF1 is well-documented to inhibit TOP1 activity as SRSF1-3 (being higher sequence similarity with SRSF1); likely, SRSF1-3 will also influence the TOP1 activity. In summary, SHM will be impacted by whatever influence that SRSF1-3 has on TOP1 activity.

Chapter 6

Chapter 6

Conclusion and scope for future work

6.1 Introduction

Germinal centre B-cells undergo antibody diversification by SHM and CSR. Interestingly, SHM and CSR are initiated by an enzyme produced in B-cells known as AID. AID-induced deamination at the Ig locus (at variable and switch regions leads to SHM and CSR, respectively) results in the recruitment of error-prone DNA repair enzymes that finally lead to mutations. However, these mutations at the Ig locus are indispensable for the expression of high-affinity antibodies. Nevertheless, AID expression levels are directly correlated to immunity (optimal level), immunosuppression (minimal level), and lymphoma (higher level) [5,30]. Additionally, AID is predominantly confined to Ig genes but its off-targeting to numerous proto-oncogenes such as BCL6, and cMYC in lymphoma and leukemia are well documented in the literature.

AID is a hotspot of extensive research in the field of immunology after its discovery two decades ago, but still, AID targeting to Ig locus is not thoroughly understood. Remarkably, transcription of Ig locus had been found to attract AID as it creates temporary single-stranded DNA where AID can act [109]. Further, augmentation in transcription levels of Ig locus enriched AID-induced mutations. Moreover, transcribing RNA PolII encounters nucleosomes that have been shown to restrict the AID action. Nevertheless, actively transcribing B-cells recruits chromatin modifiers such as the FACT and HIRA chaperone complex, resulting in AID-accessible open chromatin structure [56,57]. HIRA complex is a large multimeric protein complex comprising HIRA, CABIN1, UBN1, and ASF1a. Lately, it was reported that the UBN1 protein of the complex specifically interacts with H3.3, which results in the deposition of H3.3 at the Ig locus, which is also a hallmark of SHM [56,110]. Additionally, UBN1 is also required for the deposition of H3.3

at gene regulatory elements in ES cells of mice [62].

The diverse functional role of AID in initiating SHM and CSR is hidden in its functional domain. AID mRNA is composed of five exons which encode four reported functional domains distributed over 198 amino acids. AID N-terminal domain participates in SHM, whereas the C-terminal domain engages in CSR. Besides full-length AID numerous splice isoforms of AID such as AID δ E4a (a 30bp from the start of exon 4 is missing), and AID δ E4 (exon 4 is missing) are reported in malignancies [23,24]. Strikingly, the AID splice variant showed a range of activities such as hyper somatic mutations, a marginal increase in nuclear localization to diminished CSR activity. Furthermore, cell-based studies have shown that RBM5 (a splicing factor) is responsible for exon 4 skipping leading to the production of AID δ E4 but still, it failed to conclude how other AID splice variants were produced [104].

Interestingly, AID was found to interact with splicing factor SRSF1-3. SRSF1-3 belongs to the serine/arginine-rich splicing regulator that typically consists of an RNA binding domain (RRM) and is an isoform of the founding member of the SR family SRSF1 [70]. SRSF1 has a diverse role in RNA splicing, mRNA metabolism, etc. Interestingly, SRSF1 interaction with TOP1 leads to the inhibition of TOP1 activity is well documented in the literature [77]. TOP1 is a widely distributed enzyme in all eukaryotes which assists in unwinding the DNA and removing the torsional stress. Additionally, in the context of activated B-cells TOP1 DNA relaxation activity is contrarily related to SHM and CSR. However, the complete inhibition of TOP1 activity results in the abolishment of SHM and CSR. Unexpectedly, SRSF1-3 which is indispensable for SHM is also found to interact with TOP1 [66]. Furthermore, SRSF1-3 is reported to upregulate chromatin modifiers such as UBN1 and SATB2 [72].

6.2 Activation-induced cytidine deaminase, an antibody diversification enzyme, interacts with chromatin modifier UBN1 in B-cells

UBN1 is a part of the HIRA multimeric protein complex, which also contains HIRA protein, CABIN1, and ASF1a. Moreover, the HIRA complex deposits H3.3 at the Ig locus, which in turn creates AID-accessible single-stranded DNA and assists in Ig diversification. Recently, it was reported that HIRA knockout remarkably affected the SHM as well as influenced the mutation pattern of Ig genes. The earlier report suggests that HIRA protein interacts with RPA. Interestingly, RPA interaction with AID was well documented. So, in our current study, we want to explore whether AID interacts with UBN1.

Our study using molecular docking and molecular simulation predicted that AID is interacting with UBN1 (**Fig. 3.1 and 3.2**). Furthermore, co-immunoprecipitation data suggest that AID interacts with UBN1 in chicken cell line DT40 ψ V KO and human cell line Raji cells (**Fig. 3.5 and 3.6**). Similarly, our *in-vitro* pull-down assay with purified protein and cell lysate demonstrates the interaction between UBN1 and AID (**Fig. 3.11 and 3.12**). Additionally, immunofluorescence microscopy using fluorescently labeled antibodies indicates that AID and UBN1 are co-localized in the cells, as represented by yellow/orange spots (**Fig. 3.10 and 3.12**). Further, PLA assay strongly suggests AID and UBN1 are direct in Raji and DT40 ψ V KO cells (**Fig. 3.18 and 3.19**). Interestingly, UBN1 is also found to directly interact with H3.3. Thus, it may happen that UBN1, AID and H3.3 may be a part of the same complex that gets deposited at the Ig locus and may influence SHM. Remarkably, RNA seq data in the literature highlights that HIRA, UBN1, and H3.3 are upregulated in germinal centre B-cells compared to naïve [57]. Further, HIRA, UBN1 and H3.3 are crucial for antibody diversification. So, the fate of UBN1 and AID interaction in the light of SHM or CSR needs to be explored.

6.3 Heterologous expression of hAID in *E. coli* leads to the production of a splice isoform of AID: hAID δ C, a mystery to be explored

AID enzyme is the standout performer in activated B-cells as its genome mutator activity produces high-affinity effector antibodies [7,12]. Surprisingly, our current study found that recombinant hAID-His expression in *E. coli* cells produced two induced bands (**Fig. 4.1 A**). Further, western blotting using anti-His antibody and anti-AID antibody showed two bands and a single band respectively (**Fig. 4.1 B and 4C**). Anti-His western blotting reveals the integrity of His-tag at the c-terminus in both bands whereas anti-AID western reveals that one band correspond to full-length AID and other bands were not picked. Interestingly, the peptide mapping data and MASCOT search for both bands of hAID-His showed that one band belong to full-length AID (FL-hAID) whereas the low molecular weight band have some missing amino acids from the start of exon 4 of hAID (**Fig. 4.2A**).

To get further insight into the appearance of lower molecular weight hAID isoform, we dig deep into the existing literature. Notably, we found the lower band of hAID is known as hAID δ E4a where 10 aa of exon 4 were missing. Further, the expression of the hAID spliced variant along with FL-hAID was reported in the patients suffering from BCR-ABL-positive acute lymphoblastic leukemia, chronic lymphocytic leukemia as well as B-cell tumours [99,103]. Astonishingly, AID-positive BCR-ABL-ALL patients showed a higher degree of abnormalities in the genes such as Pax 5, BTG1, etc in contrast to AID-negative BCR-ABL-ALL. As anticipated, the higher degree of abnormalities in AID-positive BCR-ABL-ALL may be directly proportional to the activity of splice variant of AID as a few of them have hyper somatic mutational activity, impaired CSR activity, or altered nuclear localization.

Once we have identified that hAID splice isoform hAID δ E4a is expressed in *E. coli* cells as well as reported in numerous B-cell related

lymphoma. Subsequently, we wanted to explore why human cDNA of AID is spliced in *E. coli*, we performed a PCR-based detection from the cDNA of the induced cells (**Fig. 4.8 B**). Unexpectedly, we failed to detect the band corresponding to hAID δ E4a. Classical nuclear spliceosome introns are absent in prokaryotes. However, Group I and II introns are generally reported in bacterial, phage, and rRNA genes of fungi, plants, and bacteria. Moreover, Group I and II introns can adopt a three-dimensional structure which promotes the self-splicing of introns [107]. As the hAID δ E4a isoform expressed in *E. coli* cells lacks 30 bp known as putative intron. Interestingly, small introns ranging from 30 to 42 bp in the gene encoding for hypoxanthine-guanine phosphoribosyl transferase were reported in a parasitic trematode *Schistosoma mansoni* [111]. Additionally, Russel et.al, 1993 have reported very short self-splicing introns ranging from 20-33 nucleotides in *Paramecium tetraurelia*. These small introns are capable of self-splicing as they contain conserved 5' and 3' SS whereas branch point consensus was missing [112]. Notably, the RNA fold tool predicted that hAID mRNA can adopt the 3D structure like self-splicing Group II introns. Eventually, we mutated all the splice sites of a putative intron of hAID and expressed it in the *E. coli* cells. All the splice mutants of hAID expressed FL-hAID along with hAID δ E4a (**Fig. 4.8 A**). So, any mutation in the putative splice site of hAID didn't affect the expression pattern of hAID splice isoform. Most likely, mere mutation of the splice site of hAID in *E. coli* is not sufficient to suppress the expression of hAID δ E4a besides FL-hAID.

To further validate that indeed the expression of hAID in *E. coli* cells produced two induced bands, we introduced a solubilization tag NusA (NusA-hAID-His) as well as MBP (His-MBP-hAID) and expressed in *E. coli* cells. Like hAID-His, both NusA-hAID-His (**Fig. 4.3 A**) and His-MBP-hAID (**Fig. 4.4 A**) expressed two bands, the upper band belonging to full-length protein and smaller lower bands. Remarkably, both the upper band and lower band of NusA-hAID-His (**Fig. 4.3 B**) as well as His-MBP-hAID (**Fig. 4.4. B**) showed the ability

of binding to the Ni-NTA column. Since His-tag is present at the C-terminal in the case of NusA-hAID-His it tells that both bands of NusA-hAID-His have intact c-terminal whereas the same is not true for His-MBP-hAID as here His-tag is present at N-terminal.

Further, peptide sequencing data of NusA-hAID-His (**Fig. 4.9**) and His-MBP-hAID (**Fig. 4.10**) showed that indeed the lower bands were devoid of 36 and 59 amino acids respectively near the c-terminal. Thus, our earlier mass spectrometry data with only hAID-His that showed another splice variant of hAID is hAID δ E4a is not the case as many more amino acids are missing so we named it as splice variant hAID δ C. Interestingly, peptide mapping data agree with western blotting using an anti-AID antibody that the lower band have missing c-terminal residue. Additionally, we looked into the protein splicing mechanism in *E. coli*, similar to mRNA where exons and introns are present, protein contains exteins and inteins [113]. These inteins contain N and C terminals and are responsible for removing intervening inteins and the exteins are joined together to form spliced protein. Interestingly, inteins contain conserved amino acids at the start such as Ala/Gln/Pro and Asp/Gln at the end whereas hAID contains Asp at the start (143rd position) and Val at the end (152nd position) [114].

Our study is the first report that hAID cDNA expressed in *E. coli* leads to the expression of a splice isoform hAID δ C. Subsequently, we tried mutating putative intron, but it does not influence the splicing. Further, we failed to find the conserved residue of an intein in hAID again becomes impossible to know whether protein splicing is responsible for the appearance of splice isoform. Finally, exploring the mystery behind the expression of the splice isoform of hAID in *E. coli* cells will be interesting.

6.4 SRSF1-3 interaction with TOP1 likely inhibits its activity

Our study highlights that SRSF1-3 interaction with TOP1 is likely to inhibit its activity. The *in-silico* study was performed to know the interaction between SRSF1-3 and TOP1. SRSF1-3 structure was

modeled by alphafold2 whereas the TOP1 structure was downloaded from the protein data bank and uploaded to the clusopro server for molecular docking. Interestingly, the docking model predicted that SRSF1-3 interacts with TOP1, and binding free energy was as high as -815 Kcal/mol. Subsequently, molecular dynamics simulation was performed to get a deep insight into the binding mechanism and how the interaction of SRSF1-3 with TOP1 influences the stability and conformation of TOP1. The MD simulation of 600 ns was sufficient to achieve the stable conformation of SRSF1-3 and TOP1 as seen by the RMSD value. Further, TOP1 in complex with SRSF1-3 showed a sharp peak of probability density compared to TOP1 alone, which suggests SRSF1-3 stabilizes TOP1 structure. Additionally, different regions of TOP1 in the complex were compared to TOP1 alone, which showed that overall, SRSF1-3 has a stabilizing effect on TOP1 as a sharp peak was observed. Remarkably, the Cap, subdomain, and CT domain of TOP1 were involved in the interaction with RRM1 and RRM2 of SRSF1-3. Each residue fluctuation data of TOP1 in complex is less in contrast to TOP1 alone suggesting that indeed TOP1 structure was stabilized in the presence of SRSF1-3. Furthermore, the radius of gyration and SASA value predicted that indeed TOP1 interacts with SRSF1-3 as seen by increased compactness and reduced solvent accessibility respectively.

As *in-silico* method predicted that SRSF1-3 stabilized the structure of TOP1, we performed *in vitro* studies to know do SRSF1-3 affects the activity of TOP1. We expressed His-tagged SRSF1-3 in *E. coli* cells but it formed inclusion bodies, which may be due to improper folding that forced us to look for alternate ways of expressing SRSF1-3. Subsequently, we added a solubilization tag MBP and expression of His-MBP-SRSF1-3 was seen in the soluble fraction. Purified His-MBP-SRSF1-3 was used for the TOP1 activity assay. Surprisingly, the plasmid relaxation assay revealed that TOP1 activity was not affected by the presence of His-MBP-SRSF1-3. The plausible explanation behind the failure of His-MBP-SRSF1-3 to influence TOP1 activity is most likely due to the unavailability of the N-terminal of SRSF1-3,

which is due to the presence of MBP at the N-terminus. Furthermore, we removed the tag from SRSF1-3 using TEV protease, however, the SRSF1-3 ended up forming the insoluble precipitate. Recently, we have solubilized, purified, refolded, and dialyzed SRSF1-3 from inclusion bodies but somehow the concentration of SRSF1-3 was not optimum for the inhibition of TOP1 activity. Nevertheless, we are optimizing the condition suitable to achieve a higher concentration of SRSF1-3. Finally, as SRSF1-3 is very crucial for SHM, its interaction with TOP1 is most likely to influence the TOP1 activity.

6.5 Conclusion

Our study identifies a new AID interacting partner, UBN1. We employed *in-silico* techniques such as molecular docking, and molecular dynamics simulations as well as *in-vitro* assays like co-immunoprecipitation, and proximity ligation assay, which suggest that AID interacts with UBN1. Recently, the importance of HIRA protein in the SHM was unfolded [57], similar to UBN1 is also part of the HIRA chaperon complex. Furthermore, UBN1 is directly involved in the interaction with H3.3, which is required for SHM [56,110]. So, it may happen that the UBN1-AID complex is loaded at the Ig locus already enriched in H3.3, which in turn makes AID-accessible single-stranded DNA. Finally, this study established the AID-UBN1 interaction; nevertheless, the fate of this interaction needs to be explored.

Our study, for the first time, highlights that heterologous expression of hAID in *E. coli* cells produced a splice isoform of hAID, hAID δ C. We employed protein purification, western blotting, peptide sequencing, mutation of the putative intron, and the addition of a solubilization tag which suggests that besides full-length hAID, a splice isoform of hAID is produced. Recently, Oguz et al, 2021 reported that the expression of human blood clotting factor IX cDNA in moss produced a splice isoform. Additionally, changing the splice site of factor IX suppressed the splicing [115]. However, in our case, we failed to find if the splicing of hAID inside *E. coli* is happening at the RNA or

protein level. Interestingly, different splice isoforms of hAID are predominantly expressed in acute as well as chronic B-cell-related lymphoma [103]. Further, the hAID splice variant showed miscellaneous activity such as hyper somatic mutation activity, diminished CSR as well as enhanced nuclear localization in contrast to FL-hAID. As hAID is a deleterious enzyme, its enhanced hyper somatic mutation activity as well as enhanced nuclear localization, may promote aggressive lymphoma. Unexpectedly, there are almost no reports that clearly suggest why or how the FL-hAID expressed in normal B-cells is transformed to alternatively spliced isoforms of hAID in B-cell lymphoma. Our study suggests that the hAID isoform is produced in *E. coli* cells. So, hAID mature RNA may adopt different self-splicing structures in B-cell lymphoma that may lead to the production of various splice isoforms of hAID.

SRSF1-3 interaction with TOP1 was studied using *in-silico* method that suggests that RRM1 and RRM2 domains of SRSF1-3 are involved in interaction with CT, Cap and subdomain of TOP1. Moreover, SRSF1-3 stabilized the TOP structure. Further, the *in-vitro* study of SRSF1-3 on TOP1 activity may give more idea on the significance of SRSF1-3 in antibody diversification. SRSF1-3 is a crucial player necessary for SHM, and it may happen one of the ways SRSF1-3 is able to mediate SHM via influencing the TOP1 activity. Remarkably, as the level of TOP1 activity is directly related to SHM and CSR [74,75], even the marginal local influence of TOP1 activity by SRSF1-3 can be reflected in SHM.

6.6 Future directions

AID is required for antibody diversification and its interaction with UBN1 is reported in this thesis. Moreover, as UBN1 is a part of the chromatin modifier, the importance of its interaction with AID needs to be explored. Additionally, disrupting the AID and UBN1 interaction by site-directed mutagenesis, knock-down, or knock-in of UBN1 in B-cell

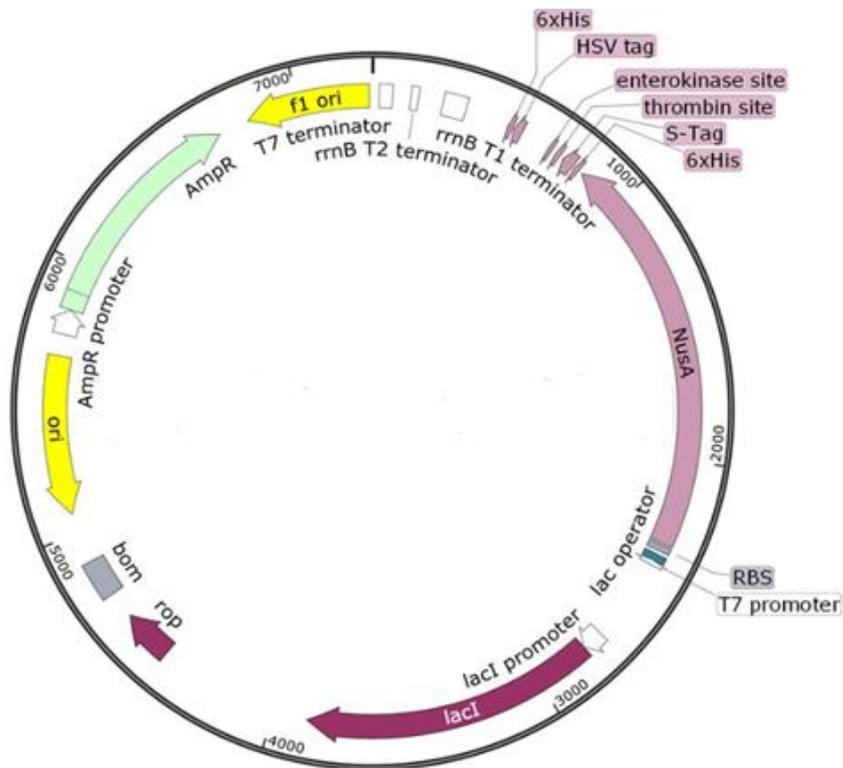
lines will give an idea of whether this interaction is required for SHM or CSR or both. Furthermore, UBN1 presence at a specific location at the Ig locus is directly correlated with the binding or mutagenic effect of AID can also be explored. The mystery of why AID expression in *E. coli* produced splice isoform is required to be explored. Interestingly, as cancer cells express various splice isoforms of AID, this identification of the reason for AID splicing inside *E. coli* cells will give an insight that AID mRNAs are self-capable of forming various splice isoforms or not in cancer. Furthermore, it will be exciting to study factors involved in the suppression of AID splicing inside normal cells, which can be used for either diagnosis or drug designing against AID-splice isoform-related lymphoma. Finally, the SRSF1-3 is most likely influencing the activity of TOP1 needs to be explored *in-vivo*. Moreover, the SRSF1-3 requirement in SHM is solely due to its influence on TOP1 activity or even the interaction of SRSF1-3 with TOP1 doesn't have any role in SHM.

Appendix-A

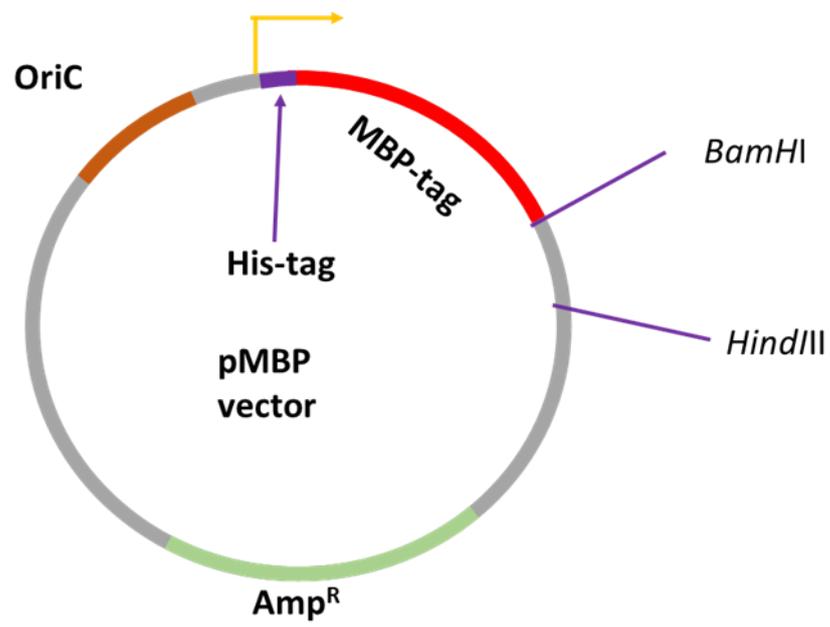
Vector maps

I. pET43a vector map.

II. MBP vector map.



I. pET43a vector map



II. MBP vector map

Appendix-B

List of primers used for PCR

Primers	Sequence 5' to 3'
PK509F	5'- GCATCTAGA ATGGACAGCCTCTTGATGAAGA -3'
PK510R	5'- ATATCTCGAGTCAAAGTCCCAGAGTTTTAAAGG -3'
PK515F	5'-ATTTCATATGGACAGCCTCTTGATGAACCGG-3'
PK516R	5'-ATATCTCGAGAAGTCCCAAAGTACGAAATGCG-3'
PK578F	5'AAAAC TAGTCTGGAAGTTCTGTTCCAGGGGCCCATGGACAGCCTCTTGATGAA -3'
PK719F	5'-TCTACTTCTGTGAGGACCGCAAG-3'
PK720R	5'-GCTTTGAAAGTTCTTTCGTGGTTTTTC-3'
PK721R	5'- GGATCCGAAGGTCATGATGGCTATTTGC-3'
PK722F	5'TTTGGATCCTATTTTTACTGCTGGAATACTTTTGTAG-3'
PK723R	5'AATCTGCAGAAAAGTATTCCAGCAGTAAAAATAGGATCC-3'
PK724F	5'-AAACTGCAG AACACGAAAGAACTTTCAAAGCC-3'
PK725R	5'AATCTGCAGAAAAGTATTCCAGCAGTAAAAATAATCTTTG -3'
PK774F	5'-GCTTGGATCCGACAGCCTCTTGATGAACCGGAG-3'
PK776F	5'-CGTGGATCCTCCGGAGGGGGCGTGATCCG-3'
PK777R	5'-CGTAAGCTTACAGCGCAGAAATAGACATCAAACCTTAAATAA-3'
PK879F	5'- TTT GCTAGC ATGTCCGGAG GGGGCGTG-3'
PK910R	5'-CGTAAGCTTTCAAAGTCCCAAAGTACGAAATG-3'
PK951R	5'-TTT AAGCTTCAGCGCAGAAATAGACATCAAACCTTAAATAA-3'
PK911R	5'-TTTAAGCTTCTCATGAGATCTAAACTTAGTG-3'
PK775R	5'-ATATCTCGAGAAGTCCCAAAGTACGAAATGCG-3'
PK934	5'GCTTGGATCCGACTACAAAGACGATGACGACAAGTCGGAGCCCCACAGGGTC-3'
PK935	5'-CGT AAGCTT TCAAGACTCTGATGCTTGTCTAAACTGCAGGG-3'
PK957	5'-AAAACATATG AAA ATA AAA ACA GGT GCA CGC ATC-3'
PK960	5'-AAAAGCTAGCCTTGGTGATACGAGTCTGCG-3'

Appendix-C

Antibodies used for co-immunoprecipitation, western blotting, and immunofluorescence microscopy

Primary Antibodies	Manufacturer
Mouse anti-AID mAB Cat #39-2500	Thermo Fisher Scientific
Mouse anti-His mAB Cat #MA1-21315	Thermo Fisher Scientific
Rabbit anti-UBN1 pAB HPA069045	Sigma Aldrich
Rabbit anti-HIRA Cat # SAB4503047	Millipore-sigma
Rabbit anti-ASF1a cat# ABE149	Sigma
Monoclonal anti-flag M2 antibody F3165	Sigma
Secondary Antibodies	
Goat anti-mouse IgH-HRP, HS201	Trans Bionovo
Goat anti-rabbit HRP	GeNeI
Goat anti-mouse Alexa Fluor 488	Abcam
Goat anti-rabbit Alexa Fluor 594	Santa Cruz Biotechnology
Duolink In situ PLA Probe Anti-Rabbit PLUS	Sigma Aldrich
Duolink In situ PLA Probe Anti-Mouse MINUS	Sigma Aldrich
Co-immunoprecipitation	
Protein A/G plus agarose beads, SC-2003	Santa Cruz Biotechnology

Appendix D

I. hAID-His upper band MASCOT data

II. hAID-His lower band MASCOT data

III. Peptide mapping of Nus-hAID-His

IV. Peptide mapping of His-MBP-hAID

hAID-His upper band MASCOT result

MASCOT Search Results

Protein View: sp|Q9GZX7|AICDA_HUMAN

Single-stranded DNA cytosine deaminase OS=Homo sapiens OX=9606 GN=AICDA PE=1 SV=1

Database: Human_Uniprot_2018
Score: 4769
Monoisotopic mass (M_r): 24337
Calculated pI: 9.50
Taxonomy: **Homo sapiens**

Sequence similarity is available as an [NCBI BLAST search of sp|Q9GZX7|AICDA_HUMAN against nr](#).

Search parameters

MS data file: D:\Chhaya-Data\Ccamp-Client_data\03092018_ReqID-4481_Monika\mgf\03092018_ReqID_4481_Lowerband.mgf
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: **Carbamidomethyl (C)**
Variable modifications: **Acetyl (Protein N-term), Oxidation (M)**

Protein sequence coverage: 46%

Matched peptides shown in **bold red**.

1 **MSLLMNRK** FLYQFKVVRM AKGRRETYLC YVVKRRDSAT **SFSLDFGYLR**
51 NKNQCHVELL FLRYISDWDL **DPRCYRVTW** FTSMSPCYDC ARHVADFLRG
101 **NPNLSLR**IFV ARLYFCEDRK AEPEGLRRLH **RAGVQIAIMT** **FKDYFYCWNT**
151 **FVENHERTFK** AMEGLHENSV RLSRQLRRIL **LPLYEVDDLK** **DAFR**TLGL

Unformatted sequence string: **198 residues** (for pasting into other applications).

hAID-His lower band MASCOT result

MASCOT Search Results

Protein View: [sp|Q9GZX7-2|AICDA_HUMAN](#)

Isoform 2 of Single-stranded DNA cytosine deaminase OS=Homo sapiens OX=9606 GN=AICDA

Database: Human_Uniprot_2018
Score: 5402
Monoisotopic mass (M_r): 22942
Calculated pI: 9.72
Taxonomy: [Homo sapiens](#)

Sequence similarity is available as [an NCBI BLAST search of sp|Q9GZX7-2|AICDA_HUMAN against nr.](#)

Search parameters

MS data file: D:\Chhaya-Data\Ccamp-Client_data\03092018_ReqID-4481_Monika\mgf\03092018_ReqID_4481_Upperband.mgf
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: [Carbamidomethyl \(C\)](#)
Variable modifications: [Acetyl \(Protein N-term\)](#), [Oxidation \(M\)](#)

Protein sequence coverage: 47%

Matched peptides shown in **bold red**.

```
1 MDSLIMNRK FLYQFKNVRW AKGRRETYLC YVVKRRDSAT SFSLDFGYLR  
51 NRNGCHVELL FLRYISDWDL DPGRCYRVW FTSWSPCYDC ARHVADFLRG  
101 NPNLSLRIFT ARLYFCEDRK AEPEGLRLRH RAGVQIAIMT FKENHERTFK  
151 AWEGLHENSV RLSRQLRRIL LPLYEVDDLR DAFRTLGL
```

Unformatted sequence string: [188 residues](#) (for pasting into other applications).

III. Peptide mapping of Nus-hAID-His

Checke	Protein	Master	Unique Sequence ID	Protein Accessi	Descri	Seque	FAST	Exp. c	Contamin	Sum P	#	Cover	#	Peptides	PSMs	#	Prot	#	Unique Peptide	#	AAs	MW [k	calc.	n	Score	Cover	#	PSM
FALSE	High	Master 1-1651104525291925452		1	seq1	seq1	MNKI	>seq1	0	FALSE	723	83	102	1492	102	102						701	79.8	5.26	2973	83	1492	
Checke	Confide	Annotated Sequence	Modif	Contam	#	Prot	#	PSM	Maste	Positions	#	Modifi	Theo.	lower band	Upper band	Ouan	Nus-hAID	lower	Nus-hAID	upper band	Confic	Percol	Percol	XCorr	Top A			
FALSE	High	[K]AAARDPQSR [A]		FALSE	1	1	2	seq1	seq1	1225-	900.5	110.2	89.8	High	High	High	Low	0.096	1	0.71	16.18							
FALSE	High	[K]AEPFGLR [R]		FALSE	1	1	4	seq1	seq1	1618-4	771.4	25.4	174.6	High	High	High	High	6E-04	2E-05	1.24	16.53							
FALSE	High	[K]AEPFGLR [L]		FALSE	1	1	3	seq1	seq1	1618-4	927.5	19.7	180.3	High	High	High	High	6E-04	0.004	1.56	16.02							
FALSE	High	[K]AGALMAAR [N]		FALSE	1	1	8	seq1	seq1	1478-	873.5	103.5	96.5	High	High	High	High	6E-04	1E-05	1.11	23.99							
FALSE	High	[K]AGALMAAR [N]	1xOxi	FALSE	1	1	3	seq1	seq1	1478-	889.5	77.4	122.6	High	High	High	High	6E-04	4E-04	2.75	20.37							
FALSE	High	[R]AGVQAIMTFK [D]		FALSE	1	1	2	seq1	seq1	1629-4	1179	150.6	49.4	High	High	High	High	6E-04	1E-04	2.54	34.71							
FALSE	High	[R]AGVQAIMTFK [D]	1xOxi	FALSE	1	1	2	seq1	seq1	1629-4	1195	135.1	64.9	High	High	High	High	6E-04	1E-05	2.18	30.84							
FALSE	High	[R]AKNALATIAQAQEEESLGDNKP [A]		FALSE	1	1	2	seq1	seq1	1410-	3480	190.8	9.2	High	High	Peak Found	High	6E-04	0.006	9.24	35.97							
FALSE	High	[R]AKNALATIAQAQEEESLGDNKP [A]		FALSE	1	1	4	seq1	seq1	1410-	4054	198.1	1.9	High	High	High	High	6E-04	3E-05	2.56	37.57							
FALSE	High	[R]AMVVDQFR [E]		FALSE	1	1	10	seq1	seq1	1124-	965.5	142.7	57.3	High	High	High	High	6E-04	2E-06	2.07	26.38							
FALSE	High	[R]AMVVDQFR [E]	1xOxi	FALSE	1	1	15	seq1	seq1	1124-	981.5	102.5	97.5	High	High	High	High	6E-04	4E-06	1.76	22.41							
FALSE	High	[R]AMVVDQFREHEGEIITGVVKK [K]		FALSE	1	1	13	seq1	seq1	1124-	2257	189.5	10.5	High	High	High	High	6E-04	2E-06	3.84	31.59							
FALSE	High	[R]AMVVDQFREHEGEIITGVVKK [K]	1xOxi	FALSE	1	1	8	seq1	seq1	1124-	2273	180.1	19.9	High	High	High	High	6E-04	2E-06	3.62	29.99							
FALSE	High	[R]AMVVDQFREHEGEIITGVVKK [K]		FALSE	1	1	5	seq1	seq1	1124-	2385	187.9	12.1	High	High	High	High	6E-04	3E-06	3.52	29.17							
FALSE	High	[R]AMVVDQFREHEGEIITGVVKK [K]	1xOxi	FALSE	1	1	3	seq1	seq1	1124-	2401	174	26	High	High	High	High	6E-04	8E-04	3.29	27.7							
FALSE	High	[K]AWEGLHENSVR [L]		IKLA'	IKLAW'	IKLA'	IKLA'	IKLA'	IKLA'	IKLAWEG'	IKLAV'	IKLA'	IKLAWEG'	IKLAWEG'	IKLA'	IKLAWEG'	IKLAWEG'	IKLAWEG'	IKLAWEG'	IKLAWEG'	IKLAWEG'	High	6E-04	2E-06	1.91	20.77		
FALSE	High	[R]DLAFKLAAR [G]		FALSE	1	1	2	seq1	seq1	1443-	1005	141	59	High	High	High	High	6E-04	1E-05	2.46	27.43							
FALSE	High	[R]DNLSDLGNNAAEAVILR [E]		FALSE	1	1	10	seq1	seq1	1148-	1827	197.3	2.7	High	High	High	High	6E-04	4E-05	3.71	39.57							
FALSE	High	[R]DNLSDLGNNAAEAVILR [E]	1xOxi	FALSE	1	1	2	seq1	seq1	1148-	2568	198.5	1.5	High	High	High	High	6E-04	0.006	3.4	41.01							
FALSE	High	[R]DNLSDLGNNAAEAVILR [E]		FALSE	1	1	2	seq1	seq1	1148-	2584	193.7	6.3	High	High	High	High	6E-04	0.001	3.71	38.61							
FALSE	High	[R]DSATSFDFGGLR [N]		FALSE	1	1	23	seq1	seq1	1534-	1579	197.1	2.9	High	High	High	High	6E-04	1E-07	3.47	41.57							
FALSE	High	[K]DYFYCVNTFVNHHER [T]		1xCar	FALSE	1	1	4	seq1	1640-4	2080	138.1	61.9	High	High	High	High	6E-04	9E-07	2.67	36.74							
FALSE	High	[R]EDMLPR [E]		FALSE	1	1	6	seq1	seq1	1165-	760.4	138.2	61.8	High	High	High	High	6E-04	1E-06	1.76	21.94							
FALSE	High	[R]EDMLPR [E]	1xOxi	FALSE	1	1	12	seq1	seq1	1165-	776.4	132.1	67.9	High	High	High	High	6E-04	6E-05	1.45	16.52							
FALSE	High	[R]EHEGEIITGVVKK [K]		FALSE	1	1	17	seq1	seq1	1132-	1311	140.7	59.3	High	High	High	High	6E-04	5E-05	3.06	23.35							
FALSE	High	[R]EHEGEIITGVVKK [V]		FALSE	1	1	5	seq1	seq1	1132-	1439	93.5	106.5	High	High	High	High	6E-04	2E-05	4.05	19.89							
FALSE	High	[R]EKIFELESALATATK [K]		FALSE	1	1	20	seq1	seq1	121-3	1722	194.4	5.6	High	High	High	High	6E-04	1E-06	4.11	40.53							
FALSE	High	[R]EKIFELESALATATK [K]		FALSE	1	1	6	seq1	seq1	121-3	1850	191.2	8.8	High	High	High	High	6E-04	7E-06	7.09	36.94							
FALSE	High	[K]EILAVVEAVSNEK [A]		FALSE	1	1	78	seq1	seq1	14-16	1401	170.8	29.2	High	High	High	High	6E-04	2E-07	3.52	34.83							
FALSE	High	[K]EILAVVEAVSNEK [A]		FALSE	1	1	4	seq1	seq1	14-20	1838	182.5	17.5	High	High	High	High	6E-04	6E-05	1.96	34.21							
FALSE	High	[K]EILAVVEAVSNEK [A]		FALSE	1	1	35	seq1	seq1	1391-	1926	189.5	10.5	High	High	High	High	6E-04	6E-09	3.87	39.86							
FALSE	High	[K]ELLEHGLDEPTVEALR [E]		FALSE	1	1	5	seq1	seq1	1391-	2211	166.5	33.5	High	High	High	High	6E-04	0.004	0.95	36.03							
FALSE	High	[K]ELLEHGLDEPTVEALR [E]		FALSE	1	1	50	seq1	seq1	174-8	962.5	92.6	107.4	High	High	High	High	6E-04	1E-06	1.45	23.16							
FALSE	High	[R]ENFRPGR [V]		FALSE	1	1	12	seq1	seq1	1171-	990.5	111.1	88.9	High	High	High	High	6E-04	6E-04	1.65	13.19							
FALSE	High	[R]ENFRPGR [V]		FALSE	1	1	5	seq1	seq1	1171-	1246	37.3	162.7	High	High	High	High	6E-04	0.014	2.18	16.2							
FALSE	High	[R]ENFRPGR [V]		FALSE	1	1	1	seq1	seq1	1171-	2473	200		High	High	Not Found	Mediu	0.013	0.321	0.92	25.22							
FALSE	High	[R]ETLYCVVKK [R]		1xCar	FALSE	1	1	30	seq1	1523-	1175	103.4	96.6	High	High	High	High	6E-04	3E-06	1.5	30.53							
FALSE	High	[R]ETLYCVVKK [R]		FALSE	1	1	2	seq1	seq1	1523-	1331	80.6	119.4	High	High	High	High	6E-04	6E-04	2.84	26.16							
FALSE	High	[R]ETLYCVVKK [R]		1xCar	FALSE	1	1	1	seq1	1523-	1487	121.7	78.3	High	High	Peak Found	High	6E-04	0.001	2.34	23.27							
FALSE	High	[K]FLYQPK [N]		FALSE	1	1	96	seq1	seq1	1508-	845.5	83.6	116.4	High	High	High	High	6E-04	8E-06	1.33	29.35							
FALSE	High	[K]FLYQPK [N]		FALSE	1	1	4	seq1	seq1	1508-	1215	136.4	63.6	High	High	High	High	6E-04	9E-05	2.38	25.49							
FALSE	High	[R]GAOLVTR [S]		FALSE	1	1	5	seq1	seq1	1192-	891.5	70.4	129.6	High	High	High	High	6E-04	5E-05	2.33	24.72							
FALSE	High	[R]GARVOAVSTLGGER [I]		FALSE	1	1	1	seq1	seq1	1256-	1530			No Oxi	Not Found	High	High	6E-04	9E-04	2.2								
FALSE	High	[R]GNPNLSLR [I]		FALSE	1	1	28	seq1	seq1	1597-4	870.5	68	132	High	High	High	High	6E-04	3E-06	1.1	21.05							
FALSE	High	[R]GNPNLSLR [I]		FALSE	1	1	2	seq1	seq1	1597-4	1459	183.3	167	High	High	High	High	6E-04	3E-04	2.17	30.09							
FALSE	High	[K]GRRETYLVVKK [R]		1xCar	FALSE	1	1	1	seq1	1520-	1544	200		High	High	Not Found	Mediu	0.013	0.374	1.3	22.65							
FALSE	High	[R]GVCTLEDLAEGQDLDADIEGL [A]		1xCar	FALSE	1	1	6	seq1	1452-	2819	200		High	High	Not Found	High	6E-04	3E-08	1.2	50.05							
FALSE	High	[R]GVLYSSVRFPEAR [G]		FALSE	1	1	34	seq1	seq1	1181-	1247	73.1	126.9	High	High	High	High	6E-04	2E-05	2.7	21.22							
FALSE	High	[R]GVLYSSVRFPEAR [G]		FALSE	1	1	2	seq1	seq1	1181-	2119	200		High	High	Not Found	High	6E-04	0.002	1.96	28.25							
FALSE	High	[K]HQAEAAHAIIDITFK [Y]		FALSE	1	1	116	seq1	seq1	1348-	1540	105.1	94.9	High	High	High	High	6E-04	3E-07	2.64	19.02							
FALSE	High	[K]HTMDIAVEAGNLAQAIGR [N]		FALSE	1	1	3	seq1	seq1	1303-	1867	189.7	10.3	High	High	High	High	6E-04	1E-06	4.14	34.79							
FALSE	High	[K]HTMDIAVEAGNLAQAIGR [N]	1xOxi	FALSE	1	1	2	seq1	seq1	1303-	1883	179.9	20.															

III. Peptide mapping of Nus-hAID-His

Checks	Protein	Master	Unique	Sequence ID	Protein Accession	Descri	Seque	FAST	Exp. c	Contaminants	Sum P	# Cover	# Peptides	# PSMs	# Prot	Unique Peptide	# AAs	MW	I calc.	p Score	Cover	# PSM	
FALSE	High	Master 1-	-1651104552391925452		1	seq1	seq1	MNKi	>seq1	0	FALSE	723	83	102	1492	102	102	701	79.8	5.26	2973	83	1492
Check	Confide	Annotated	Sequence	Modif	Contam	# Prot	# PSM	Maste	Positions	à Modif	# Theo.	lower band	Upper band	Quan	Nus-hAID	lower	Nus-hAID	upper band	Confic	Percol	Percol	XCorr	Top A
FALSE	High	[R]KFLYQFK[N]	FALSE	1	1	5	seq1	seq1	1507-		973.6	84.2	115.8	High	High	High	6E-04	7E-05	2.92	24.49			
FALSE	High	[R]KFLYQFKNVR[W]	FALSE	1	1	1	seq1	seq1	1507-		1343	179.4	20.6	High	Peak Found	High	6E-04	7E-04	3.17	22.45			
FALSE	High	[K]KKYEQEIDVR[V]	FALSE	1	1	17	seq1	seq1	137-4		1308	100.7	99.3	High	High	High	6E-04	2E-05	3.88	15.3			
FALSE	High	[R]KSGDFDTR[R]	FALSE	1	1	7	seq1	seq1	152-6		1073	101.3	98.7	High	High	High	6E-04	4E-06	2.59	20.29			
FALSE	High	[R]KSGDFDTRR[W]	FALSE	1	1	9	seq1	seq1	152-6		1229	73.5	126.5	High	High	High	6E-04	1E-04	1.98	16.13			
FALSE	High	[K]KVVNRDNISLDLGNNAEAVILR[L]	FALSE	1	1	2	seq1	seq1	1144-		2324	200		High	Not Found	High	6E-04	3E-04	2.28	31.44			
FALSE	High	[K]KYEQEIDVR[V]	FALSE	1	1	25	seq1	seq1	138-4		1180	100.9	99.1	High	High	High	6E-04	3E-07	2.86	16.75			
FALSE	High	[R]LASQLSGWELNVMVTVDLQAK	FALSE	1	1	11	seq1	seq1	1327-		2318	199	1	High	High	High	6E-04	4E-08	2.94	41.17			
FALSE	High	[R]LASQLSGWELNVMVTVDLQAK	1xOxi	FALSE	1	1	6	seq1	seq1	1327-	2334	198.1	1.9	High	High	High	6E-04	1E-06	5.47	38.33			
FALSE	High	[R]IDPVGACVGM[R][G]	1xCar	FALSE	1	1	11	seq1	seq1	1245-	1175	118.8	81.2	High	High	High	6E-04	1E-05	1.34	26.54			
FALSE	High	[R]IDPVGACVGM[R][G]	1xCar	FALSE	1	1	14	seq1	seq1	1245-	1191	74.1	125.9	High	High	High	6E-04	5E-05	2.4	23.19			
FALSE	High	[R]IDPVGACVGM[R][G]	1xCar	FALSE	1	1	1	seq1	seq1	1245-	1459	200		High	Not Found	Low	0.05	1	1.04	17.45			
FALSE	High	[R]EVPEEGEVIEK[A]	FALSE	1	1	59	seq1	seq1	1211-		1597	171.2	28.8	High	High	High	6E-04	9E-10	3.08	37.93			
FALSE	High	[R]EVPEEGEVIEKAAAR[D]	FALSE	1	1	1	seq1	seq1	1211-		1966	194.8	5.2	High	Peak Found	High	6E-04	1E-07	4.13	35.3			
FALSE	High	[K]FEALESALATATK[K]	FALSE	1	1	59	seq1	seq1	123-3		1465	186.9	13.1	High	High	High	6E-04	8E-09	2.92	40.8			
FALSE	High	[K]FEALESALATATKK[K]	FALSE	1	1	14	seq1	seq1	123-3		1593	176.5	23.5	High	High	High	6E-04	2E-06	3.19	36.17			
FALSE	High	[R]LLPLVEVDL[R][D]	FALSE	1	1	9	seq1	seq1	1676-		1459	129.9	70.1	High	High	High	6E-04	3E-07	3.22	40.47			
FALSE	High	[R]LLPLVEVDL[R][D]	FALSE	1	1	8	seq1	seq1	1676-		1948	167.9	32.1	High	High	High	6E-04	9E-07	2.14	41.72			
FALSE	High	[R]ITOTAKOVIVQK[V]	FALSE	1	1	6	seq1	seq1	1105-		1458	95.1	104.9	High	High	High	6E-04	8E-06	4.52	19.46			
FALSE	High	[R]ITOTAKOVIVQK[V][E]	FALSE	1	1	1	seq1	seq1	1105-		1713	139.1	60.9	High	Peak Found	High	6E-04	3E-04	2.27	20.25			
FALSE	High	[R]LYFCEDR[K]	1xCar	FALSE	1	1	38	seq1	seq1	1610-	1002	59.3	140.7	High	High	High	6E-04	6E-07	1.72	25.09			
FALSE	High	[R]LYFCEDR[K]	1xCar	FALSE	1	1	8	seq1	seq1	1610-	1131	57.9	142.1	High	High	High	6E-04	4E-05	1.43	20.05			
FALSE	High	[K]MKNKELAVVEAVSNEK[A]	FALSE	1	1	10	seq1	seq1	11-16		1774	173.2	26.8	High	High	High	6E-04	6E-05	5.16	35.04			
FALSE	High	[K]MKNKELAVVEAVSNEK[A]	1xOxi	FALSE	1	1	8	seq1	seq1	11-16	1790	141.4	58.6	High	High	High	6E-04	1E-06	2.44	32.52			
FALSE	High	[K]MKNKELAVVEAVSNEK[A]	1xAce	FALSE	1	1	2	seq1	seq1	11-16	1816	187.9	12.1	High	Peak Found	High	6E-04	0.005	1.45	41.73			
FALSE	High	[K]MKNKELAVVEAVSNEKALPR[IE]	FALSE	1	1	1	seq1	seq1	11-20		2211			No O	High	Not Found	High	6E-04	0.002	4.4			
FALSE	High	[K]MKNKELAVVEAVSNEKALPR[IE]	1xOxi	FALSE	1	1	1	seq1	seq1	11-20	2227			No O	High	Not Found	High	6E-04	4E-04	1.55			
FALSE	High	[K]NALATAQAQOESLGDNNKPADP	FALSE	1	1	12	seq1	seq1	1412-		2281	197.7	2.3	High	High	High	6E-04	1E-08	7.21	38.8			
FALSE	High	[K]NALATAQAQOESLGDNNKPADP	FALSE	1	1	7	seq1	seq1	1412-		3855	191.3	8.7	High	High	High	6E-04	8E-09	6.73	40			
FALSE	High	[K]NGCHVELLFLR[V]	1xCar	FALSE	1	1	11	seq1	seq1	1550-	1358	181.4	18.6	High	High	High	6E-04	4E-05	2.21	31.56			
FALSE	High	[R]NKNKGVVLLFLR[V]	1xCar	FALSE	1	1	8	seq1	seq1	1548-	1600	159.5	40.5	High	High	High	6E-04	5E-04	2.23	27.31			
FALSE	High	[K]QVIVQKVR[E]	FALSE	1	1	2	seq1	seq1	1112-		969.6	200		High	High	High	6E-04	9E-05	2.62	13.88			
FALSE	High	[K]QVIVQKREAE[R][A]	FALSE	1	1	1	seq1	seq1	1112-		1455	200		High	Not Found	High	6E-04	0.013	1.71	14.94			
FALSE	High	[R]RDSATSFSLDFGYLR[N]	FALSE	1	1	5	seq1	seq1	1533-		1735	197.2	2.8	High	High	High	6E-04	6E-05	3.66	34.9			
FALSE	High	[R]RETYLCYVVK[R]	1xCar	FALSE	1	1	8	seq1	seq1	1522-	1331	126.9	73.1	High	High	High	6E-04	9E-04	2.65	25.52			
FALSE	High	[R]RETYLCYVVK[R]	1xCar	FALSE	1	1	1	seq1	seq1	1522-	1487	200		High	Not Found	High	6E-04	0.01	3.37	22.51			
FALSE	High	[R]RKFLYQFK[N]	FALSE	1	1	5	seq1	seq1	1506-		1130	112.1	87.9	High	High	High	6E-04	0.003	2.51	21.08			
FALSE	High	[K]RIDPVGACVGM[R][G]	1xCar	FALSE	1	1	4	seq1	seq1	1244-	1331	132	68	High	High	High	6E-04	1E-04	4.6	22.92			
FALSE	High	[K]RIDPVGACVGM[R][G]	1xCar	FALSE	1	1	4	seq1	seq1	1244-	1347	74.7	125.3	High	High	High	6E-04	0.003	3.81	20.14			
FALSE	High	[R]RILLPLYEVDL[R][D]	FALSE	1	1	5	seq1	seq1	1675-		1615	138	62	High	High	High	6E-04	5E-04	1.56	35.22			
FALSE	High	[R]RILLPLYEVDL[R][D]	FALSE	1	1	1	seq1	seq1	1675-		2104	186.2	13.8	High	Peak Found	High	6E-04	0.069	1.8	37.57			
FALSE	High	[K]RRDSATSFSLDFGYLR[N]	FALSE	1	1	1	seq1	seq1	1532-		1891	200		High	Not Found	High	6E-04	0.013	2.81	32.04			
FALSE	High	[R]RWLVVDEVTPQK[E]	FALSE	1	1	4	seq1	seq1	161-7		1571	163.1	36.9	High	High	High	6E-04	7E-04	3.22	29.07			
FALSE	High	[K]SGDFDTR[R]	FALSE	1	1	12	seq1	seq1	153-6		944.4	102.6	97.4	High	High	High	6E-04	2E-06	2.36	25.28			
FALSE	High	[K]SGDFDTRR[W]	FALSE	1	1	12	seq1	seq1	153-6		1101	64.4	135.6	High	High	High	6E-04	0.007	1.25	20.81			
FALSE	High	[R]SKPEMILFLR[I]	FALSE	1	1	13	seq1	seq1	1200-		1363	189.8	10.2	High	High	High	6E-04	5E-05	0.76	34.58			
FALSE	High	[R]SKPEMILFLR[I]	1xOxi	FALSE	1	1	11	seq1	seq1	1200-	1379	145.5	54.5	High	High	High	6E-04	4E-05	2.98	34.87			
FALSE	High	[R]TXAWEGLIHNSVRL[L]	FALSE	1	1	7	seq1	seq1	1655-		1674	31.2	168.8	High	High	High	6E-04	1E-06	3.46	23.29			
FALSE	High	[K]TNDKRIPVGACVGM[R][G]	1xCar	FALSE	1	1	2	seq1	seq1	1240-	1789			No O	High	High	6E-04	1E-06	3.98				
FALSE	High	[K]TNDKRIPVGACVGM[R][G]	1xCar	FALSE	1	1	1	seq1	seq1	1240-	1805			No O	High	Not Found	High	6E-04	0.027	0.61			
FALSE	High	[K]VNRDNISLDLGNNAEAVILR[IE]	FALSE	1	1	3	seq1	seq1	1145-		2196	195.6	4.4	High	High	High	6E-04	6E-06	4.66	33.81			
FALSE	High	[K]VNRDNISLDLGNNAEAVILRED	FALSE	1	1	1	seq1	seq1	1145-		2938	200		High	Not Found	High	6E-04	0.003	3.66	36.03			
FALSE	High	[K]VNRDNISLDLGNNAEAVILRED	1xOxi	FALSE	1	1	1	seq1	seq1	1145-	2954	200		High	Not Found	High	6E-04	3E-04	6.21	34.24			
FALSE	High	[R]VQAVSTELGGER[I]	FALSE	1	1	66	seq1	seq1	1259-		1246	87.1	112.9	High	High	High	6E-04	6E-08	2.31	21.99			
FALSE	High	[K]VREARFAMVVDQFR[E]	1xOxi	FALSE	1	1	1	seq1	seq1	1118-	1722			No O	High	Not Found	Media	0.032	0.873	1.13			
FALSE	High	[R]VRGVLVSVRPEAR[G]	FALSE	1	1	2	seq1	seq1	1179-		1502	43.4	156.6	High	High	High	6E-04	0.011	2.54	20.03			
FALSE	High	[R]VTFWTSWSPCYDCAR[H]	2xCar	FALSE	1	1	13	seq1	seq1	1575-	1936	180.4	19.6	High	High	High	6E-04	2E-06	1.44	38.35			

III. Peptide mapping of Nus-hAID-His

Check	Protein	Master	Unique Sequence ID	Protein Accession	Description	Sequence	FAST	Exp. c	Contaminant	Sum P	# Cover	# Peptides	# PSMs	# Prot	# Unique Peptide	# AAs	MW	fi	calc. n	Score	Cover	# PSM
FALSE	High	Master1	-1651104552391925452	1	seq1	seq1	MNKI>seq1	0	FALSE	723	83	102	1492	102	102		701	79.8	5.26	2973	83	1492
Check	Confide	Annotated Sequence	Modif	Contam	# Prot	# Prot	# PSM	Maste	Positions	h	Modifi	# Theo.	lower band	Upper band	Quan	Nus-hAID lower	Nus-hAID upper band	Confic	Percol	Percol	XCorr	Top A
FALSE	High	[R].WLVVDEVTPQTK.[E]		FALSE	1	1	91	seq1	seq1	[62-7]		1415	132.5	67.5		High	High	High	6E-04	9E-08	2.61	33.22
FALSE	High	[R].WLVVDEVTPQTKHLEAAR.[Y]		FALSE	1	1	7	seq1	seq1	[62-8]		2298	176.3	23.7		High	High	High	6E-04	1E-06	4.28	35.32
FALSE	High	[R].YEDESLNLGDYVEDOIESVTFDI		FALSE	1	1	6	seq1	seq1	[82-10]		2736	197.7	2.3		High	High	High	6E-04	2E-09	2.99	46.63
FALSE	High	[K].YEQEIDVR.[V]		FALSE	1	1	6	seq1	seq1	[39-40]		1052	94.4	105.6		High	High	High	6E-04	9E-06	2.45	21.57
FALSE	High	[R].YISDWDLPGR.[C]		FALSE	1	1	60	seq1	seq1	[561-]		1337	104.9	95.1		High	High	High	6E-04	5E-07	2.31	33.46
FALSE	High	[R].YISDWDLPGR.[V]	1xCar	FALSE	1	1	1	seq1	seq1	[561-]		1816				No Q; Not Found	High	High	6E-04	0.048	1.33	

IV. Peptide mapping of His-MBP-hAID

Protein FC	Master	Unique Sequence ID	Prot	Accession	Description	Sequen	Exo. n-val	Contaminant	Sum PEP Score	# Deco	Coverage [%]	# Peptides	# PSMs	# Protein	# Unique I	# AAs	MW [kDa]	
High	Master Pr	6179294056156229354	1	seq1	seq1	MHHH	0	FALSE	1092.671	0	92	115	1563	115	115	616	69.5	
Checked	Confidence	Annotated Sequence	Moc	Master Pr	Positions in Master	Modific	Theo. MH	Abundance MBP AID u	Abundance MB	Quan In	MBP-hAID user	MBP-hAID lower t	Confidenc	Percolato	Percolato	XCorr (b)	Top Apex RT	
FALSE	High	[K] AEPEGLR [R]		seq1	seq1 [539-545]		176.8	23.2	High	High	High	High	High	0.000556	6.07E-06	1.05	14.79	
FALSE	High	[K] AEPEGLRR [L]		seq1	seq1 [539-546]		927.5007	164.5	High	High	High	High	High	0.001018	0.0106%	1.03	13.15	
FALSE	High	[K] AFQDKLYPTWDVAVR [Y]		seq1	seq1 [117-131]		1856.933	78	High	High	High	High	High	0.000556	1.53E-08	3.83	33.04	
FALSE	High	[K] AFQDKLYPTWDVAVRYNGK [L]		seq1	seq1 [117-135]		2319.156	31.7	High	High	High	High	High	0.000556	1.43E-05	1.83	30.68	
FALSE	High	[K] AGLTFLVDLKK [N]		seq1	seq1 [223-233]		1189.719	75.4	High	High	High	High	High	0.000556	3.26E-08	4.26	41.07	
FALSE	High	[K] AGLTFLVDLKK [H]		seq1	seq1 [223-235]		1431.857	106.1	High	High	High	High	High	0.000556	9.07E-06	4.02	36	
FALSE	High	[K] AGLTFLVDLKKHMNADTDYSIAEAFNK [G]	1xO	seq1	seq1 [223-252]		3326.673	200	Not Found	High	Low	0.07895	1	1.09	37.84			
FALSE	High	[R] AGVQAIAMTFK [D]		seq1	seq1 [550-560]		1178.66	193.7	High	High	High	High	High	0.000556	4.57E-07	2.73	31.86	
FALSE	High	[R] AGVQAIAMTFK [D]	1xO	seq1	seq1 [550-560]		1194.655	200	Not Found	High	Not Found	High	High	0.000556	9.74E-05	1.94	28.07	
FALSE	High	[R] AGVQAIAMTFKDYFYCWNTFVNER [T]	1xO	seq1	seq1 [550-575]		3230.508	200	High	High	Not Found	High	High	0.000556	3.59E-05	1.95	42.11	
FALSE	High	[K] AWEGLHENSVR [L]		seq1	seq1 [579-589]		1292.628	194.4	High	High	High	High	High	0.000556	8.38E-08	3	18.63	
FALSE	High	[K] AWEGLHENSVR [Q]		seq1	seq1 [579-592]		1653.846	200	High	High	Not Found	High	High	0.000556	5.15E-06	3.26	20.58	
FALSE	High	[K] DKP.GAVALK [S]		seq1	seq1 [329-338]		1001.62	94.5	High	High	High	High	High	0.000556	2.09E-05	1.88	19.58	
FALSE	High	[K] DKP.GAVALKSYEEELAKDPR [I]		seq1	seq1 [329-349]		2329.491	50.9	High	High	High	High	High	0.000556	0.000918	1.71	25.93	
FALSE	High	[K] DLLPNPK [T]		seq1	seq1 [153-160]		893.5091	115.9	High	High	High	High	High	0.000556	2.15E-06	2.5	24.89	
FALSE	High	[K] DLLPNPKTWEIPALDK [E]		seq1	seq1 [153-170]		2076.101	78.2	High	High	High	High	High	0.000556	1.26E-05	2.29	34.28	
FALSE	High	[K] DLLPNPKTWEIPALDKELK [A]		seq1	seq1 [153-173]		2446.323	58.3	High	High	High	High	High	0.000556	2.92E-06	5.59	33.24	
FALSE	High	[R] DSATSFSLDGLR [N]		seq1	seq1 [455-468]		1578.743	121	High	High	High	High	High	0.000556	7.62E-09	3	38.4	
FALSE	High	[R] DSATSFSLDGLR [N]		seq1	seq1 [455-470]		1820.881	84.7	High	High	High	High	High	0.000556	4.8E-06	3.13	32.52	
FALSE	High	[K] DTGKIVTEHPDK [L]		seq1	seq1 [63-75]		1438.754	68.4	High	High	High	High	High	0.000556	0.001833	1.02	16.89	
FALSE	High	[K] DTGKIVTEHPDKLEEK [F]		seq1	seq1 [63-79]		1938.018	83.8	High	High	High	High	High	0.000556	3.76E-09	4.45	18.82	
FALSE	High	[K] DVGVDNAGAK [A]		seq1	seq1 [213-221]		945.4636	71	High	High	High	High	High	0.000556	4.81E-06	2	15.25	
FALSE	High	[K] DVGVDNAGAKGLTFLVDLKK [N]		seq1	seq1 [213-233]		2116.165	89.8	High	High	High	High	High	0.000556	1.8E-07	3.78	42.83	
FALSE	High	[K] DVGVDNAGAKGLTFLVDLKK [H]		seq1	seq1 [213-235]		2358.303	129	High	High	Peak Found	High	High	0.000556	0.000242	1.64	40.17	
FALSE	High	[K] DVFYCWNTFVNER [T]	1xO	seq1	seq1 [561-575]		2079.865	196	High	High	High	High	High	0.000556	1.13E-08	3.74	33.91	
FALSE	High	[K] EFLNLLTDEGLEAVNK [D]		seq1	seq1 [311-328]		2097.039	91.4	High	High	High	High	High	0.000556	1.9E-11	4.84	37.84	
FALSE	High	[K] EFLNLLTDEGLEAVNKDKPLGAVALK [S]		seq1	seq1 [311-338]		3089.641	72.6	High	High	High	High	High	0.000556	1.47E-09	8.17	34.6	
FALSE	High	[K] ELAKEFLNLLTDEGLEAVNK [D]		seq1	seq1 [307-328]		2538.297	111.9	High	High	High	High	High	0.000556	1.54E-07	2.52	37.23	
FALSE	High	[K] ELAKEFLNLLTDEGLEAVNKDKPLGAVALK [S]		seq1	seq1 [307-338]		3530.899	96.2	High	High	High	High	High	0.000556	4.89E-07	7.83	35.18	
FALSE	High	[R] ETYLCYVVK [R]	1xO	seq1	seq1 [444-452]		1174.581	101.3	High	High	High	High	High	0.000556	5.14E-08	1.59	27.69	
FALSE	High	[R] ETYLCYVVK [R]	1xO	seq1	seq1 [444-453]		1330.682	75.3	High	High	High	High	High	0.000556	8.83E-05	2.05	23.6	
FALSE	High	[R] ETYLCYVVK [D]	1xO	seq1	seq1 [444-454]		1486.783	80.8	High	High	High	High	High	0.000556	0.000463	1.56	20.91	
FALSE	High	[K] FEKDTGKIV [V]		seq1	seq1 [60-67]		937.4989	19.7	High	High	High	High	High	0.000556	1.25E-05	1.69	13.16	
FALSE	High	[K] FEKDTGKIVTEHPDK [L]		seq1	seq1 [60-75]		1842.96	71.2	High	High	Peak Found	High	High	0.000556	0.00025	3.91	16.63	
FALSE	High	[R] FGAYAGSGLLAETPPDK [A]		seq1	seq1 [100-116]		1766.896	103.6	High	High	High	High	High	0.000556	6.25E-10	5.23	33.32	
FALSE	High	[R] FGAYAGSGLLAETPPDK [L]		seq1	seq1 [100-121]		2356.182	103.9	High	High	High	High	High	0.000556	9.83E-09	7.31	32.77	
FALSE	High	[R] FGAYAGSGLLAETPPDKAFQDK [Y]		seq1	seq1 [100-131]		3608.811	87.7	High	High	High	High	High	0.000556	4.06E-05	2.71	39.17	
FALSE	High	[K] FLYQFK [N]		seq1	seq1 [429-434]		845.4556	88	High	High	High	High	High	0.000556	1.2E-06	1.67	26.56	
FALSE	High	[K] FLYQFK [N]		seq1	seq1 [429-437]		1214.668	88.2	High	High	High	High	High	0.000556	1.34E-05	2.22	22.98	
FALSE	High	[K] FLYQFK [N]		seq1	seq1 [429-440]		1599.879	84.6	High	High	High	High	High	0.000556	0.003348	3.49	24.96	
FALSE	High	[K] FPGVAATGGDPIFAVADR [F]		seq1	seq1 [80-99]		2213.077	118.3	High	High	High	High	High	0.000556	1.26E-11	4.31	34.36	
FALSE	High	[K] GEIMPNIQMSAFWYAVR [T]		seq1	seq1 [360-377]		2110.025	187.1	High	High	High	High	High	0.000556	2.36E-08	5.04	41.8	
FALSE	High	[K] GEIMPNIQMSAFWYAVR [T]	1xO	seq1	seq1 [360-377]		2126.02	163.3	High	High	High	High	High	0.000556	8.58E-08	2.81	37.01	
FALSE	High	[K] GEIMPNIQMSAFWYAVR [T]	2xO	seq1	seq1 [360-377]		2142.015	105.4	High	High	High	High	High	0.000556	1.34E-05	4.37	34.5	
FALSE	High	[K] GEIMPNIQMSAFWYAVRTAVINAASGR [G]	1xO	seq1	seq1 [360-387]		3066.529	182.1	High	High	Peak Found	High	High	0.000556	0.000757	2.19	42.66	
FALSE	High	[K] GEIMPNIQMSAFWYAVRTAVINAASGR [Q]	2xO	seq1	seq1 [360-387]		3082.524	134	High	High	Peak Found	High	High	0.000556	0.000145	1.81	41.65	
FALSE	High	[K] GETAMTINGPWAWNSIDTSK [V]		seq1	seq1 [253-272]		2179.012	158.9	High	High	High	High	High	0.000556	4.75E-08	4.34	35.66	
FALSE	High	[K] GETAMTINGPWAWNSIDTSK [V]	1xO	seq1	seq1 [253-272]		2195.007	88.2	High	High	High	High	High	0.000556	8.1E-08	1.8	33.29	
FALSE	High	[K] GSGGGGGGLENLYFGSDSLMNR [F]	1xO	seq1	seq1 [404-428]		2573.216	150.8	High	High	High	High	High	0.000556	0.02585	0.8378	0.68	30.54
FALSE	High	[R] GNPNLSLR [I]		seq1	seq1 [518-525]		870.4792	99.2	High	High	High	High	High	0.000556	1.65E-06	1.05	18.86	
FALSE	High	[R] GNPNLSLR [R]		seq1	seq1 [518-530]		1458.838	113.7	High	High	High	High	High	0.000556	0.001381	1.36	27.16	
FALSE	High	[K] GQSPKPEVGVLSGAINAASP [E]		seq1	seq1 [285-306]		2139.156	87.8	High	High	High	High	High	0.000556	1.14E-07	3.66	27.85	

IV. Peptide mapping of His-MBP-hAID

Checked	Protein FC	Master	Unique Sequence ID	Prot	Accession	Description	Sequen	Exp. o-vak	Contaminant	Sum PEP Score	# Decon.	Coverage [%]	# Peptides	# PSMs	# Protein	# Unique	# AAs	MW [kDa]	Top Apex
FALSE	High	Master Pr	6179294056156229354	1	seq1	seq1	MHHH	0	FALSE	1092.671	0		92	115	1863	115	616	69.5	
Checked	Confidenc	Annotated Sequence	Mod	Master Pr	Positions in Master	Modific	Theo. MH	Abundance	MBP AID u	Abundance MB	Quan Ir	MBP-hAID usser	MBP-hAID lower f	Confidenc	Percolato	Percolato	XCorr	Idv	Top Apex
FALSE	High	[K] GQSPKPVYVQVSGAINAASPNWELAK [E]		seq1	seq1		2580.414	56.7	143.3	High	High	High	High	0.000556	3.21E-11	9.23	26.91		
FALSE	High	[K] GQSPKPVYVQVSGAINAASPNWELAK [E]		seq1	seq1		2580.414	56.7	143.3	High	High	High	High	0.000556	9.93E-06	3.11	38.77		
FALSE	High	[K] GRRETYLYVVK [R]	1xO	seq1	seq1		1543.805	46.5	153.5	High	High	High	High	0.004458	0.2061	1.87	20.31		
FALSE	High	[K] GYNGLAEVGK [K]		seq1	seq1		1007.516	117.3	82.7	High	High	High	High	0.000556	4.27E-05	2.52	20.42		
FALSE	High	[K] GYNGLAEVGK [F]		seq1	seq1		1135.611	101.2	98.8	High	High	High	High	0.000556	0.000434	2.51	16.56		
FALSE	High	[K] HMNADTYSIAEAFNK [G]		seq1	seq1		1897.839	151.3	48.7	High	High	High	High	0.000556	2.55E-09	6.51	26.41		
FALSE	High	[K] HMNADTYSIAEAFNK [G]	1xO	seq1	seq1		1913.833	123.9	76.1	High	High	High	High	0.000556	1.87E-06	5.99	25.28		
FALSE	High	[K] HMNADTYSIAEAFNKGETAMTINGPWAWSF		seq1	seq1		4057.833	192.9	7.1	High	High	High	High	0.000556	4.46E-08	10.06	37.47		
FALSE	High	[K] HMNADTYSIAEAFNKGETAMTINGPWAWSF	1xO	seq1	seq1		4073.828	194.9	5.1	High	High	Peak Found	High	0.000556	5.88E-07	6.17	34.27		
FALSE	High	[K] HMNADTYSIAEAFNKGETAMTINGPWAWSF	2xO	seq1	seq1		4089.823	141.2	58.8	High	Peak Found	High	High	0.000556	1.49E-05	2.27	33.79		
FALSE	High	[R] HVADFLR [G]		seq1	seq1		857.4628	99.7	100.3	High	High	High	High	0.000556	2.48E-07	2.34	19.73		
FALSE	High	[R] HVADFLRGNPNLSR [I]		seq1	seq1		1708.924	41.1	158.9	High	High	High	High	0.000556	6.15E-07	3.16	25.59		
FALSE	High	[R] HVADFLRGNPNLSRIFR [L]		seq1	seq1		2297.263	169.9	30.1	High	Peak Found	High	High	0.000556	0.000447	1.97	30.3		
FALSE	High	[R] KAEPEGLR [R]		seq1	seq1		899.4945	95.9	104.1	High	Peak Found	High	High	0.000556	2.68E-05	1.24	13.15		
FALSE	High	[R] KFEKDTGK [V]		seq1	seq1		1065.594	200	200	High	High	High	High	0.000556	1.07E-05	2.87	13.13		
FALSE	High	[R] KFLVQPK [N]		seq1	seq1		973.5506	98.5	101.5	High	High	High	High	0.000556	3.66E-06	2.73	21.98		
FALSE	High	[R] KFLVQPK [W]		seq1	seq1		1428.4371	242.763	81.8	118.4	High	High	High	0.000556	4.12E-06	3.25	19.98		
FALSE	High	[R] LAATMENAQK [G]		seq1	seq1		1076.54	144.2	55.8	High	High	High	High	0.000556	9.36E-08	3.1	15.36		
FALSE	High	[R] LAATMENAQK [G]	1xO	seq1	seq1		1092.535	70.7	129.3	High	High	High	High	0.000556	1.05E-06	2.69	13.14		
FALSE	High	[R] LAATMENAQKGEIMPNPQMSAPWYAVR [I]		seq1	seq1		3167.547	200		High	Not Found	High	High	0.000556	0.002051	1.4	37.8		
FALSE	High	[R] LAATMENAQKGEIMPNPQMSAPWYAVR [I]	1xO	seq1	seq1		3183.542	161.1	38.9	High	Peak Found	High	High	0.000556	3.08E-06	3.16	33.77		
FALSE	High	[R] LAATMENAQKGEIMPNPQMSAPWYAVR [I]	7xO	seq1	seq1		3199.537	200		High	Not Found	High	High	0.000556	0.000311	1.01	32.88		
FALSE	High	[R] LAATMENAQKGEIMPNPQMSAPWYAVR [I]	3xO	seq1	seq1		3215.532	80.8	119.2	High	High	High	High	0.000556	0.000241	1.15	30.94		
FALSE	High	[K] IEEGKLVIWINGDK [G]		seq1	seq1		1613.89	101.3	98.7	High	High	High	High	0.000556	1.23E-07	5.14	29.13		
FALSE	High	[K] IEEGKLVIWINGDKYNGLAEVGK [K]		seq1	seq1		2602.388	65.6	134.4	High	High	High	High	0.000556	9.63E-10	7.54	30.32		
FALSE	High	[K] IEEKPPQVAATGGDPHFVWADR [F]		seq1	seq1		2712.342	120.3	79.7	High	High	High	High	0.000556	1.92E-10	8.01	32.52		
FALSE	High	[R] IFTARLYVCDRK [A]	1xO	seq1	seq1		1718.868	200		High	Not Found	High	High	0.000241	0.06277	1.16	22.87		
FALSE	High	[K] LIAYPIAEVALSLYLNK [D]		seq1	seq1		1891.094	134.6	65.4	High	High	High	High	0.000556	1.13E-09	4.06	44.65		
FALSE	High	[K] LIAYPIAEVALSLYLNKLNPPK [T]		seq1	seq1		2765.585	126	74	High	High	High	High	0.000556	1.02E-08	6.39	43.76		
FALSE	High	[K] LIAYPIAEVALSLYLNKLNPPKWEIPALDK [I]		seq1	seq1		3948.177	200		Not Found	High	High	High	0.000556	1.29E-05	1.8	43.73		
FALSE	High	[R] ILLIPEVDLDR [D]		seq1	seq1		1458.82	196.3	3.7	High	High	High	High	0.000556	2.83E-09	2.71	37.42		
FALSE	High	[R] ILLIPEVDLDR [T]		seq1	seq1		1948.054	196.8	3.2	High	High	High	High	0.000556	1.08E-07	2.5	38.67		
FALSE	High	[R] ILLIPEVDLDR [L]		seq1	seq1		2332.291	200		High	Not Found	High	High	0.000556	2.25E-05	1.29	45.29		
FALSE	High	[K] LVIWINGDK [G]		seq1	seq1		1057.604	121.5	78.5	High	High	High	High	0.000556	2.69E-08	2.36	29.7		
FALSE	High	[K] LVIWINGDKYNGLAEVGK [K]		seq1	seq1		2046.102	124.1	75.9	High	High	High	High	0.000556	1.54E-07	3.75	30.92		
FALSE	High	[K] LVIWINGDKYNGLAEVGK [F]		seq1	seq1		2174.197	78.1	121.9	High	High	High	High	0.000556	3.88E-05	3.21	27.59		
FALSE	High	[R] LYFCDRK [K]	1xO	seq1	seq1		1002.435	118.6	81.4	High	High	High	High	0.000556	1.13E-07	1.68	22.61		
FALSE	High	[R] LYFCDRK [A]	1xO	seq1	seq1		1130.53	116.2	83.8	High	High	High	High	0.000556	2.38E-05	1.76	16.29		
FALSE	High	[R] LYFCDRKAEPEGLR [R]	1xO	seq1	seq1		1882.912	163.1	36.9	High	High	High	High	0.000556	4.47E-05	2.65	20.89		
FALSE	High	[K] LYFTVDVAVR [Y]		seq1	seq1		1122.1311	1267.647	98.8	101.2	High	High	High	0.000556	1.46E-07	1.77	34.42		
FALSE	High	[K] LYFTVDVAVR [NGK] [L]		seq1	seq1		1729.87	36.5	163.5	Peak Found	High	High	High	0.000556	0.000288	2.85	30.91		
FALSE	High	[R] MCHVVELLRL [Y]	1xO	seq1	seq1		1357.705	134.6	65.4	High	High	High	High	0.000556	5.81E-08	3.28	28.61		
FALSE	High	[K] NMHMNADTYSIAEAFNK [G]		seq1	seq1		2139.976	134.7	65.3	High	High	High	High	0.000556	1.86E-06	5.5	23.12		
FALSE	High	[K] NMHMNADTYSIAEAFNK [G]	1xO	seq1	seq1		2155.971	100.3	99.7	High	High	High	High	0.000556	2E-05	4.74	22.1		
FALSE	High	[K] NMHMNADTYSIAEAFNKGETAMTINGPWAWSF		seq1	seq1		4299.971	193.8	6.2	High	Peak Found	High	High	0.000556	1.58E-05	3.55	34.94		
FALSE	High	[K] NMHMNADTYSIAEAFNKGETAMTINGPWAWSF	1xO	seq1	seq1		4315.966			No Out	High	Not Found	Medium	0.04178	1	0.76			
FALSE	High	[R] NMGVVELLRL [Y]	1xO	seq1	seq1		1599.842	64.1	135.9	High	High	High	High	0.000556	6.89E-06	3.09	24.66		
FALSE	High	[R] QTVDEALK [D]		seq1	seq1		903.4782	87.9	112.1	High	High	High	High	0.000556	2.16E-06	1.42	17		
FALSE	High	[R] RDSATSFDFGLR [N]		seq1	seq1		1734.845	127.5	72.5	High	High	High	High	0.000556	1.02E-05	4.81	32.11		
FALSE	High	[R] RETLYLVVVK [R]	1xO	seq1	seq1		1330.682	108.8	91.2	High	High	High	High	0.000556	6.84E-05	3.59	23		
FALSE	High	[R] RETLYLVVVK [R]		seq1	seq1		1486.783	58.2	141.8	High	High	High	High	0.000556	0.000382	2.89	20.18		
FALSE	High	[R] RETLYLVVVK [N]		seq1	seq1		1122.652	19	181	High	High	High	High	0.000556	0.000018	2.02	18.46		
FALSE	High	[R] RILLPEVDLDR [D]		seq1	seq1		1614.921	197.3	2.7	High	High	High	High	0.000556	1.61E-06	2.78	32.37		
FALSE	High	[R] RILLPEVDLDR [T]		seq1	seq1		2104.155	197.7	2.3	High	High	High	High	0.000556	2.14E-06	4.27	34.55		
FALSE	High	[K] RRDSATSFDFGLR [N]		seq1	seq1		1890.946	74	126	High	High	High	High	0.000556	5.36E-05	3.12	29.1		

IV. Peptide mapping of His-MBP-hAID

Checked	Protein FC	Master	Unique Sequence ID	Prot	Accession	Description	Sequen	Exo. o-val	Contaminant	Sum PEP Score	# Decon	Coverage [%]	# Peptides	# PSMs	# Protein	# Unique	# AAs	MW (kDa)
FALSE	High	Master Pr	6179294056156229354	1	seq1	seq1	MHHHH	0	FALSE	1092.671	0	92	115	1863	115	115	616	69.5
Checked	Confidenc	Annotated Sequence	Moc	Master Pr	Positions in Master	Modific	Theo. MH	Abundance MBP AID u	Abundance MB	Quan Ir	MBP-hAID upper	MBP-hAID lower	Confidenc	Percolato	Percolato	XCorr f/v	Top Apex R	
FALSE	High	[K].SALMFNLEQPPFTWPLIAADGGYAFK.[Y]		seq1	seq1 [178-203]		2950.448	97.8	102.2	High	High	High	High	0.000556	8.49E-09	3.72	46.13	
FALSE	High	[K].SALMFNLEQPPFTWPLIAADGGYAFK.[Y]	1xO	seq1	seq1 [178-203]		2966.443	77.7	122.3	High	High	High	High	0.000556	7.38E-08	2.28	45.12	
FALSE	High	[K].SYEEELAK.[D]		seq1	seq1 [339-346]		968.4571	100.7	99.3	High	High	High	High	0.000556	9.34E-05	1.73	18.4	
FALSE	High	[K].SYEEELAKDPR.[I]		seq1	seq1 [339-349]		1336.638	82.1	117.9	High	High	High	High	0.000556	1.82E-09	3.52	18.49	
FALSE	High	[R].TAVINAASGR.[Q]		seq1	seq1 [378-387]		959.5269	82.7	117.3	High	High	High	High	0.000556	1.94E-08	2.19	16.18	
FALSE	High	[R].TAVINAASGRQTVDEALK.[D]		seq1	seq1 [378-395]		1843.987	70.2	129.8	High	High	High	High	0.000556	1.79E-05	3.41	22.39	
FALSE	High	[R].TFKAWEGLHENSVR.[L]		seq1	seq1 [576-589]		1673.839	174.9	25.1	High	High	High	High	0.000556	6.43E-07	3.34	20.91	
FALSE	High	[R].TFKAWEGLHENSRLR.[Q]		seq1	seq1 [576-592]		2030.057			Not Found	High	Not Found	High	0.000556	0.004074	1.18		
FALSE	High	[K].TWEIIPALDK.[E]		seq1	seq1 [161-170]		1201.61	91.3	108.7	High	High	High	High	0.000556	1.42E-08	1.8	29.48	
FALSE	High	[K].TWEIIPALDKELK.[A]		seq1	seq1 [161-173]		1571.832	87.4	112.6	High	High	High	High	0.000556	8.27E-08	1.71	29.06	
FALSE	High	[K].TWEIIPALDKELKAK.[G]		seq1	seq1 [161-175]		1770.964	35.1	164.9	High	High	High	High	0.000556	0.000445	2.61	26.2	
FALSE	High	[K].VNYGVTVLPTFFK.[G]		seq1	seq1 [273-284]		1337.746	82.4	117.6	High	High	High	High	0.000556	6.28E-07	2.22	31.61	
FALSE	High	[K].VNYGVTVLPTFFKGGPSKPFVGLV[SAGINAASPNK		seq1	seq1 [273-306]		3457.884	91.2	108.8	High	High	High	High	0.000556	2.1E-07	6.04	32.69	
FALSE	High	[K].VNYGVTVLPTFFKGGPSKPFVGLV[SAGINAASPNK		seq1	seq1 [273-310]		3899.143	88.9	111.1	High	High	High	High	0.000556	0.000699	3.66	31.66	
FALSE	High	[K].VTVEHPDKLEEK.[F]		seq1	seq1 [68-79]		1423.743	86.5	113.5	High	High	High	High	0.000556	9.47E-06	3.94	14.83	
FALSE	High	[K].VTVEHPDKLEEKFPQVAATGGDPDHFVAHDR.[I]		seq1	seq1 [68-99]		3617.802	115.9	84.1	High	High	High	High	0.000556	4.07E-09	8.3	30.17	
FALSE	High	[R].VTWFTSWSPCYDCAR.[H]		2xCi	seq1 [496-510]		1935.815	110.4	89.6	High	High	High	High	0.000556	1.54E-07	3.79	35.37	
FALSE	High	[K].VDIKDVGVDNAGAK.[A]		seq1	seq1 [209-222]		1464.733	104.9	95.1	High	High	High	High	0.000556	4.74E-07	3.67	21.17	
FALSE	High	[K].YENGYDIK.[D]		seq1	seq1 [204-212]		1129.552	76.8	123.2	High	High	High	High	0.000556	9.86E-07	3.08	14.93	
FALSE	High	[K].YENGYDIKDVGDNAGAK.[A]		seq1	seq1 [204-222]		2055.998	61.4	138.6	High	High	High	High	0.000556	2.28E-08	5.18	19.57	
FALSE	High	[R].YISDWLDPGR.[C]		seq1	seq1 [482-492]		1336.617	96.4	103.6	High	High	High	High	0.000556	6.38E-07	3.28	30.54	
FALSE	High	[R].YISDWLDPGR.[V]		1xCi	seq1 [482-495]		1815.812		200	High	High	High	High	0.000556	7.54E-06	1.97	28.1	
FALSE	High	[R].YNGKLIAYPIAVEALSLIYNKDLLNPPK.[T]		seq1	seq1 [132-160]		3227.808	103.1	96.9	High	High	High	High	0.000556	8.5E-06	4.02	41.53	

References

1. Maizels N. (2005), Immunoglobulin gene diversification. *Annu Rev Genet*, 39, 23–46 (<https://doi.org/10.1146/annurev.genet.39.073003.110544>)
2. Oettinger M.A., Schatz D.G., Gorka C., Baltimore D. (1990), RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, 248, 1517–23 (<https://doi.org/10.1126/science.2360047>)
3. Petersen-Mahrt S.K., Harris R.S., Neuberger M.S. (2002), AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*, 418, 99–103. (<https://doi.org/10.1038/nature00862>)
4. Muramatsu M., Kinoshita K., Fagarasan S., Yamada S., Shinkai Y., Honjo T. (2000), Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, 102, 553–63. ([https://doi.org/10.1016/s0092-8674\(00\)00078-7](https://doi.org/10.1016/s0092-8674(00)00078-7))
5. Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O., et al. (2000), Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*, 102, 565–75 ([https://doi.org/10.1016/s0092-8674\(00\)00079-9](https://doi.org/10.1016/s0092-8674(00)00079-9))
6. Rajewsky K. (1996) Clonal selection and learning in the antibody system. *Nature*, 381, 751–8 (<https://doi.org/10.1038/381751a0>)
7. Victora G.D., Nussenzweig M.C. Germinal centers. (2012), *Annu Rev Immunol*, 30, 429–57 (<https://doi.org/10.1146/annurev-immunol-020711-075032>)
8. Choudhary M., Tamrakar A., Singh AK., Jain M., Jaiswal A., Kodgire P. (2018), AID Biology: A pathological and clinical perspective. *Int Rev Immunol*, 37, 37–56 (<https://doi.org/10.1080/08830185.2017.1369980>)

9. Neuberger M.S., di Noia J.M., Beale R.C.L., Williams G.T., Yang Z., Rada C. (2005), Somatic hypermutation at A.T pairs: polymerase error versus dUTP Incorporation, *Nat Rev Immunol*, 5, 171–8. (<https://doi.org/10.1038/nri1553>)
10. Rada C., Williams G.T., Nilsen H., Barnes D.E., Lindahl T., Neuberger M.S. (2002), Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol*, 12, 1748–55 ([https://doi.org/10.1016/s0960-9822\(02\)01215-0](https://doi.org/10.1016/s0960-9822(02)01215-0))
11. Wilson T.M., Vaisman A., Martomo S.A., Sullivan P., Lan L., Hanaoka F., et al. (2005) MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. *J Exp Med*, 201, 637–45, (<https://doi.org/10.1084/jem.20042066>)
12. Stavnezer J., Guikema J.E.J., Schrader C.E. Mechanism and regulation of class switch recombination. (2008), *Annu Rev Immunol*, 26, 261–92 (<https://doi.org/10.1146/annurev.immunol.26.021607.090248>)
13. Rada C., di Noia J.M., Neuberger M.S. (2004), Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell*, 16, 163–71 (<https://doi.org/10.1016/j.molcel.2004.10.011>)
14. Guikema J.E.J., Linehan E.K., Tsuchimoto D., Nakabeppu Y., Strauss P.R., Stavnezer J., et al. (2007), APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. *J Exp Med*, 204, 3017–26 (<https://doi.org/10.1084/jem.20071289>)
15. Stavnezer J., Schrader C.E. (2014), IgH chain class switch recombination: mechanism and regulation. *J Immunol*, 193, 5370–8 (<https://doi.org/10.4049/jimmunol.1401849>)
16. Chaudhuri J., Alt F.W. (2004), Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol*, 4, 541–52 (<https://doi.org/10.1038/nri1395>)

17. Methot S.P., di Noia JM. (2017), Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv Immunol*, 133, 37–87 (<https://doi.org/10.1016/bs.ai.2016.11.002>)
18. Boboila C., Alt FW., Schwer B. (2012), Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks, *Adv Immunol*, 116, 1–49. (<https://doi.org/10.1016/B978-0-12-394300-2.00001-6>)
19. Pannunzio N.R., Li S., Watanabe G., Lieber MR. (2014), Non-homologous end joining often uses microhomology: implications for alternative end joining. *DNA Repair*, 17, 74–80 (<https://doi.org/10.1016/j.dnarep.2014.02.006>)
20. Lieber M.R. (2010), The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, 79, 181–211. (<https://doi.org/10.1146/annurev.biochem.052308.093131>)
21. Bransteitter R., Pham P., Calabrese P., Goodman MF. (2004), Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase, *J Biol Chem*, 279, 51612–21 (<https://doi.org/10.1074/jbc.M408135200>)
22. Shinkura R., Ito S., Begum N.A., Nagaoka H., Muramatsu M., Kinoshita K., et al. (2004), Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol*, 5, 707–12 (<https://doi.org/10.1038/ni1086>)
23. Iacobucci I., Lonetti A., Messa F., Ferrari A., Cilloni D., Soverini S., et al. (2010), Different isoforms of the B-cell mutator activation-induced cytidine deaminase are aberrantly expressed in BCR-ABL1-positive acute lymphoblastic leukemia patients. *Leukemia*, 24, 66–73. (<https://doi.org/10.1038/leu.2009.197>)
24. van Maldegem F., Jibodh RA., van Dijk R., Bende R.J., van Noesel C.J.M. (2010), Activation-induced cytidine deaminase splice variants are defective because of the lack of structural support for the catalytic

site. J Immunol, 184, 2487–91
(<https://doi.org/10.4049/jimmunol.0903102>)

25. Wu X., Darce J.R., Chang S.K., Nowakowski G.S., Jelinek D.F. (2008), Alternative splicing regulates activation-induced cytidine deaminase (AID): implications for suppression of AID mutagenic activity in normal and malignant B cells, *Blood*, 112, 4675–82
(<https://doi.org/10.1182/blood-2008-03-145995>)

26. Chelico L., Pham P., Goodman MF. (2009), Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G, *Philos Trans R Soc Lond B Biol Sci*, 364, 583–93
(<https://doi.org/10.1098/rstb.2008.0195>)

27. Larijani M., Frieder D., Basit W., Martin A. (2005), The mutation spectrum of purified AID is similar to the mutability index in Ramos cells and in *ung(-/-)msh2(-/-)* mice, *Immunogenetics*, 56, 840–5
(<https://doi.org/10.1007/s00251-004-0748-0>)

28. Pham P., Bransteitter R., Petruska J., Goodman MF. (2003), Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation, *Nature*, 424, 103–7
(<https://doi.org/10.1038/nature01760>)

29. Shapiro G.S., Aviszus K., Murphy J., Wysocki L.J. (2002), Evolution of Ig DNA sequence to target specific base positions within codons for somatic hypermutation, *J Immunol*, 168, 2302–6
(<https://doi.org/10.4049/jimmunol.168.5.2302>)

30. Nagaoka H., Tran T.H., Kobayashi M., Aida M., Honjo T. (2010), Preventing AID, a physiological mutator, from deleterious activation: regulation of the genomic instability that is associated with antibody diversity, *Int Immunol*, 22, 227–35
(<https://doi.org/10.1093/intimm/dxq023>)

31. Tran T.H., Nakata M., Suzuki K., Begum N.A., Shinkura R., Fagarasan S., et al. (2010), B cell-specific and stimulation-responsive

enhancers derepress *Aicda* by overcoming the effects of silencers, *Nat Immunol*, 11, 148–54 (<https://doi.org/10.1038/ni.1829>)

32. Yadav A., Oлару A., Saltis M., Setren A., Cerny J., Livák F. (2006), Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene, *Mol Immunol*, 43, 529–41 (<https://doi.org/10.1016/j.molimm.2005.05.007>)

33. Dedeoglu F., Horwitz B., Chaudhuri J., Alt FW., Geha R.S. (2004), Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB, *Int Immunol*, 16, 395–404 (<https://doi.org/10.1093/intimm/dxh042>)

34. Gonda H., Sugai M., Nambu Y., Katakai T., Agata Y., Mori K.J., et al. (2003), The balance between Pax5 and Id2 activities is the key to AID gene expression, *J Exp Med*, 198, 1427–37 (<https://doi.org/10.1084/jem.20030802>)

35. Park S.R., Zan H., Pal Z., Zhang J., Al-Qahtani A., Pone E.J., et al. (2009), HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation, *Nat Immunol*, 10, 540–50 (<https://doi.org/10.1038/ni.1725>)

36. Kohlhaas S., Garden O.A., Scudamore C., Turner M., Okkenhaug K., Vigorito E. (2009), Cutting edge: the Foxp3 target miR-155 contributes to the development of regulatory T cells, *J Immunol* 2009, 182, 2578–82 (<https://doi.org/10.4049/jimmunol.0803162>)

37. Vigorito E., Perks K.L., Abreu-Goodger C., Bunting S., Xiang Z., Kohlhaas S., et al. (2007), microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells, *Immunity*, 27, 847–59 (<https://doi.org/10.1016/j.immuni.2007.10.009>)

38. Thai T.H., Calado D.P., Casola S., Ansel K.M., Xiao C., Xue Y., et al. (2007), Regulation of the germinal center response by microRNA-155, *Science*, 316, 604–8 (<https://doi.org/10.1126/science.1141229>)

39. Dorsett Y., McBride K.M., Jankovic M., Gazumyan A., Thai T.H., Robbiani D.F., et al. (2008), MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation, *Immunity*, 28, 630–8 (<https://doi.org/10.1016/j.immuni.2008.04.002>)
40. de Yebenes V.G., Belver L., Pisano D.G., Gonzalez S., Villasante A., Croce C., et al. (2008), miR-181b negatively regulates activation-induced cytidine deaminase in B cells, *J Exp Med*, 205, 2199–206 (<https://doi.org/10.1084/jem.20080579>)
41. Borchert G.M., Holton N.W., Larson ED. (2011), Repression of human activation induced cytidine deaminase by miR-93 and miR-155, *BMC Cancer*, 11, 347 (<https://doi.org/10.1186/1471-2407-11-347>)
42. Hasler J., Rada C., Neuberger MS. (2011), Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A), *Proc Natl Acad Sci U S A*, 108, 18366–71 (<https://doi.org/10.1073/pnas.1106729108>)
43. Aoufouchi S., Faili A., Zober C., D’Orlando O., Weller S., Weill J-C., et al. (2008), Proteasomal degradation restricts the nuclear lifespan of AID, *J Exp Med*, 205, 1357–68 (<https://doi.org/10.1084/jem.20070950>)
44. Brar S.S., Watson M., Diaz M. (2004), Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks, *J Biol Chem*, 279, 26395–401 (<https://doi.org/10.1074/jbc.M403503200>)
45. Ito S., Nagaoka H., Shinkura R., Begum N., Muramatsu M., Nakata M., et al. (2004), Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1, *Proc Natl Acad Sci U S A*, 101, 1975–80 (<https://doi.org/10.1073/pnas.0307335101>)
46. McBride K.M., Barreto V., Ramiro A.R., Stavropoulos P., Nussenzweig M.C. (2004), Somatic hypermutation is limited by CRM1-

dependent nuclear export of activation-induced deaminase, *J Exp Med*, 199, 1235–44 (<https://doi.org/10.1084/jem.20040373>)

47. Betz A.G., Milstein C., Gonzalez-Fernandez A., Pannell R., Larson T, Neuberger MS. (1994), Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region, *Cell*, 77, 239–48 ([https://doi.org/10.1016/0092-8674\(94\)90316-6](https://doi.org/10.1016/0092-8674(94)90316-6)) 48. Peters A., Storb U. (1996), Somatic hypermutation of immunoglobulin genes is linked to transcription initiation, *Immunity*, 4, 57–65 ([https://doi.org/10.1016/s1074-7613\(00\)80298-8](https://doi.org/10.1016/s1074-7613(00)80298-8))

49. Parsa J.Y., Ramachandran S., Zaheen A., Nepal R.M., Kapelnikov A., Belcheva A, et al. (2012), Negative supercoiling creates single-stranded patches of DNA that are substrates for AID-mediated mutagenesis. *PLoS Genet*, 8:e1002518. (<https://doi.org/10.1371/journal.pgen.1002518>)

50. Kodgire P., Mukkavar P., Ratnam S., Martin T.E., Storb U. (2013), Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID, *J Exp Med*, 210, 1481–92 (<https://doi.org/10.1084/jem.20121523>)

51. Meng F.L., Du Z., Federation A., Hu J., Wang Q., Kieffer-Kwon K-R., et al. (2014), Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability, *Cell*, 159, 1538–48 (<https://doi.org/10.1016/j.cell.2014.11.014>)

52. Qian J., Wang Q., Dose M., Pruett N., Kieffer-Kwon K-R., Resch W., et al. (2014), B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity, *Cell*, 159, 1524–37 (<https://doi.org/10.1016/j.cell.2014.11.013>)

53. Pefanis E., Wang J., Rothschild G., Lim J., Chao J., Rabadan R., et al. (2014), Noncoding RNA transcription targets AID to divergently transcribed loci in B cells, *Nature*, 514, 389–93 (<https://doi.org/10.1038/nature13580>)

54. Arents G., Burlingame R.W., Wang B.C., Love W.E., Moudrianakis EN. (1991), The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix., *Proc Natl Acad Sci U S A*, 88, 10148–52 (<https://doi.org/10.1073/pnas.88.22.10148>)
55. Kodgire P., Mukkavar P., North J.A., Poirier M.G., Storb U. (2012), Nucleosome stability dramatically impacts the targeting of somatic hypermutation, *Mol Cell Biol*, 32, 2030–40 (<https://doi.org/10.1128/MCB.06722-11>)
56. Aida M., Hamad N., Stanlie A., Begum N.A., Honjo T. (2013), Accumulation of the FACT complex, as well as histone H3.3, serves as a target marker for somatic hypermutation, *Proc Natl Acad Sci U S A*, 110, 7784–9 (<https://doi.org/10.1073/pnas.1305859110>)
57. Yu G., Zhang Y., Gupta V., Zhang J., MacCarthy T., Duan Z., et al. (2021), The role of HIRA-dependent H3.3 deposition and its modifications in the somatic hypermutation of immunoglobulin variable regions, *Proc Natl Acad Sci U S A*, 118, (<https://doi.org/10.1073/pnas.2114743118>)
58. Belotserkovskaya R., Oh S., Bondarenko V.A., Orphanides G., Studitsky V.M., Reinberg D. (2003), FACT facilitates transcription-dependent nucleosome alteration, *Science*, 301, 1090–3 (<https://doi.org/10.1126/science.1085703>)
59. Schwabish M.A., Struhl K. (2004), Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II, *Mol Cell Biol*, 24, 10111–7 (<https://doi.org/10.1128/MCB.24.23.10111-10117.2004>)
60. Ahmad K., Henikoff S. (2002), The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly, *Mol Cell*, 9, 1191–200 ([https://doi.org/10.1016/s1097-2765\(02\)00542-7](https://doi.org/10.1016/s1097-2765(02)00542-7))

61. Sherwood P.W., Tsang S.V., Osley M.A. (1993), Characterization of HIR1 and HIR2, two genes required for regulation of histone gene transcription in *Saccharomyces cerevisiae*, *Mol Cell Biol*, 13, 28–38 (<https://doi.org/10.1128/mcb.13.1.28-38.1993>)
62. Xiong C., Wen Z., Yu J., Chen J., Liu C.P., Zhang X., et al. (2018), UBN1/2 of HIRA complex is responsible for recognition and deposition of H3.3 at cis-regulatory elements of genes in mouse ES cells, *BMC Biol*, 16, 110 (<https://doi.org/10.1186/s12915-018-0573-9>)
63. Chaudhuri J., Khuong C., Alt FW. (2004), Replication protein A interacts with AID to promote deamination of somatic hypermutation targets, *Nature*, 430, 992–8 (<https://doi.org/10.1038/nature02821>)
64. Conticello S.G., Ganesh K., Xue K., Lu M., Rada C., Neuberger M.S. (2008), Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNB1, *Mol Cell*, 31, 474–84 (<https://doi.org/10.1016/j.molcel.2008.07.009>)
65. Nowak U., Matthews A.J., Zheng S., Chaudhuri J. (2011), The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA, *Nat Immunol*, 12, 160–6 (<https://doi.org/10.1038/ni.1977>)
66. Singh A.K., Tamrakar A., Jaiswal A., Kanayama N., Agarwal A., Tripathi P., et al. (2019), Splicing regulator SRSF1-3 that controls somatic hypermutation of IgV genes interacts with topoisomerase 1 and AID, *Mol Immunol*, 116, 63–72, (<https://doi.org/10.1016/j.molimm.2019.10.002>)
67. Kuhny M., Forbes L.R., Çakan E., Vega-Loza A., Kostiuk V., Dinesh R.K., et al. (2020), Disease-associated CTNNB1 mutation impairs somatic hypermutation by decreasing nuclear AID, *J Clin Invest*, 130, 4411-4422 (<https://doi.org/10.1172/JCI131297>)
68. Krainer A.R., Conway G.C., Kozak D. (1990), The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by

activating proximal sites, *Cell*, 62, 35–42 ([https://doi.org/10.1016/0092-8674\(90\)90237-9](https://doi.org/10.1016/0092-8674(90)90237-9))

69. Das S., Krainer A.R. (2014), Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer, *Mol Cancer Res*, 12, 1195–204 (<https://doi.org/10.1158/1541-7786.MCR-14-0131>)

70. Graveley B.R. (2000), Sorting out the complexity of SR protein functions. *RNA*, 6, 1197–211 (<https://doi.org/10.1017/s1355838200000960>)

71. Kanehiro Y., Todo K., Negishi M., Fukuoka J., Gan W., Hikasa T., et al. (2012), Activation-induced cytidine deaminase (AID)-dependent somatic hypermutation requires a splice isoform of the serine/arginine-rich (SR) protein SRSF1, *Proc Natl Acad Sci U S A*, 109, 1216–21 (<https://doi.org/10.1073/pnas.1120368109>)

72. Singh A.K., Tamrakar A., Jaiswal A., Kanayama N., Kodgire P. (2020), SRSF1-3, a splicing and somatic hypermutation regulator, controls transcription of IgV genes via chromatin regulators SATB2, UBN1 and histone variant H3.3, *Mol Immunol*, 119, 69–82 (<https://doi.org/10.1016/j.molimm.2020.01.005>)

73. Dobрева G., Dambacher J., Grosschedl R. (2003), SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin mu gene expression, *Genes Dev*, 17, 3048–61 (<https://doi.org/10.1101/gad.1153003>)

74. Kobayashi M., Aida M., Nagaoka H., Begum N.A., Kitawaki Y., Nakata M, et al. (2009), AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination, *Proc Natl Acad Sci U S A*, 106, 22375–80 (<https://doi.org/10.1073/pnas.0911879106>)

75. Kobayashi M., Sabouri Z., Sabouri S., Kitawaki Y., Pommier Y., Abe T., et al. (2011), Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic

hypermethylation, *Proc Natl Acad Sci U S A*, 108, 19305–10 (<https://doi.org/10.1073/pnas.1114522108>)

76. Zhang H., Gan H., Wang Z., Lee J.H., Zhou H., Ordog T., et al. (2017), RPA Interacts with HIRA and Regulates H3.3 Deposition at Gene Regulatory Elements in Mammalian Cells, *Mol Cell*, 65, 272–84 (<https://doi.org/10.1016/j.molcel.2016.11.030>)

77. Ishikawa T., Krzysko K.A., Kowalska-Loth B., Skrajna A.M., Czuby A., Girstun A., et al. (2012), Activities of topoisomerase I in its complex with SRSF1, *Biochemistry*, 51, 1803–16 (<https://doi.org/10.1021/bi300043t>)

78. Roy A., Kucukural A., Zhang Y. (2010), I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc*, 5, 725–38 (<https://doi.org/10.1038/nprot.2010.5>)

79. Yang J., Yan R., Roy A., Xu D., Poisson J., Zhang Y. (2015), The I-TASSER Suite: protein structure and function prediction, *Nat Methods*, 12, 7–8 (<https://doi.org/10.1038/nmeth.3213>)

80. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., et al. (2021), Highly accurate protein structure prediction with AlphaFold, *Nature*, 596, 583–9 (<https://doi.org/10.1038/s41586-021-03819-2>)

81. Lyskov S., Gray J.J. (2008), The RosettaDock server for local protein-protein docking, *Nucleic Acids Res*, 36, 233-8 (<https://doi.org/10.1093/nar/gkn216>)

82. Case D.A., Cheatham T.E 3rd, Darden T, Gohlke H, Luo R., Merz K.M.J., et al. (2005), The Amber biomolecular simulation programs, *J Comput Chem*, 26, 1668–88 (<https://doi.org/10.1002/jcc.20290>)

83. Maier J.A., Martinez C., Kasavajhala K., Wickstrom L., Hauser K.E., Simmerling C. (2015), ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J Chem Theory Comput*, 11, 3696–713 (<https://doi.org/10.1021/acs.jctc.5b00255>)

84. Price D.J., Brooks C.L. (2004), A modified TIP3P water potential for simulation with Ewald summation, *J Chem Phys*, 121, 10096–103 (<https://doi.org/10.1063/1.1808117>)
85. Kräutler V., van Gunsteren W.F., Hünenberger P.H. (2001), A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, *J Comput Chem*, 22, 501–8 ([https://doi.org/https://doi.org/10.1002/1096-987X\(20010415\)22:5<501::AID-JCC1021>3.0.CO;2-V](https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V))
86. Darden T.A., York D.M., Pedersen L.G. (1993), Particle mesh Ewald: An $O(N \log N)$ method for Ewald sums in large systems, *Journal of Chemical Physics*, 98, 10089–92
87. Pastor R.W., Brooks B.R., Szabo A. (1988), An analysis of the accuracy of Langevin and molecular dynamics algorithms, *Mol Phys*, 65, 1409–19 (<https://doi.org/10.1080/00268978800101881>)
88. Roy R., Mishra A., Poddar S., Nayak D., Kar P. (2022), Investigating the mechanism of recognition and structural dynamics of nucleoprotein-RNA complex from Peste des petits ruminants virus via Gaussian accelerated molecular dynamics simulations, *J Biomol Struct Dyn*, 40, 2302–15 (<https://doi.org/10.1080/07391102.2020.1838327>)
89. Roe D.R., Cheatham T.E. (2013), PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J Chem Theory Comput*, 9, 3084–95 (<https://doi.org/10.1021/ct400341p>)
90. Kollman Peter. (1993), Free energy calculations: Applications to chemical and biochemical phenomena, *Chem Rev*, 93, 2395–417 (<https://doi.org/10.1021/cr00023a004>)
91. Gohlke H., Kiel C., Case D.A. (2003), Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes, *J Mol Biol*, 330, 891–913 ([https://doi.org/10.1016/s0022-2836\(03\)00610-7](https://doi.org/10.1016/s0022-2836(03)00610-7))
92. Suvarma K., Biswas D., Pai M.G., Acharjee A., Bankar R., Palanivel V. (2021), Proteomics and Machine Learning Approaches Reveal a Set

of Prognostic Markers for COVID-19 Severity With Drug Repurposing Potential, *Front Physiol*, 12, 652799 (DOI:10.3389/fphys.2021.652799)

93. Yewdell W.T., Kim Y., Chowdhury P., Lau C.M., Smolkin R.M., Belcheva K.T., et al. (2020), A Hyper-IgM Syndrome Mutation in Activation-Induced Cytidine Deaminase Disrupts G-Quadruplex Binding and Genome-wide Chromatin Localization, *Immunity*, 53, 952-970 (e11. <https://doi.org/10.1016/j.immuni.2020.10.003>)

94. Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. (1984), Molecular dynamics with coupling to an external bath, *J Chem Phys*, 81, 3684–90 (<https://doi.org/10.1063/1.448118>) 95. di Noia J., Neuberger M.S. (2002), Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*, 419, 43–8 (<https://doi.org/10.1038/nature00981>)

96. Romanello M., Schiavone D., Frey A., Sale J.E. (2016), Histone H3.3 promotes IgV gene diversification by enhancing formation of AID-accessible single-stranded DNA, *EMBO J*, 35, 1452–64 (<https://doi.org/10.15252/embj.201693958>)

97. Lyskov S., Chou F.C., Conchúir S.Ó., Der B.S., Drew K., Kuroda D., et al. (2013), Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE), *PLoS One*, 8:e63906.

98. Kawaguchi Y., Nariki H., Kawamoto N., Kanehiro Y., Miyazaki S., Suzuki M., et al. (2017), SRSF1-3 contributes to diversification of the immunoglobulin variable region gene by promoting accumulation of AID in the nucleus, *Biochem Biophys Res Commun*, 485, 261–6 (<https://doi.org/10.1016/j.bbrc.2017.02.097>)

99. Albesiano E., Messmer B.T., Damle R.N., Allen S.L., Rai K.R., Chiorazzi N. (2003), Activation-induced cytidine deaminase in chronic lymphocytic leukemia B cells: expression as multiple forms in a dynamic, variably sized fraction of the clone, *Blood*, 102, 3333–9 (<https://doi.org/10.1182/blood-2003-05-1585>)

100. Pasqualucci L., Guglielmino R., Houldsworth J., Mohr J., Aoufouchi S., Polakiewicz R., et al. (2004), Expression of the AID protein in normal and neoplastic B cells, *Blood*, 104, 3318–25 (<https://doi.org/10.1182/blood-2004-04-1558>)
101. Pasqualucci L., Bhagat G., Jankovic M., Compagno M., Smith P., Muramatsu M., et al. (2008), AID is required for germinal center-derived lymphomagenesis, *Nat Genet*, 40, 108–12 (<https://doi.org/10.1038/ng.2007.35>)
102. Pasqualucci L., Neumeister P., Goossens T., Nanjangud G., Chaganti R.S., Küppers R., et al. (2001), Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas, *Nature*, 412, 341–6 (<https://doi.org/10.1038/35085588>)
103. Rossi D., Berra E., Cerri M., Deambrogi C., Barbieri C., Franceschetti S., et al. (2006), Aberrant somatic hypermutation in transformation of follicular lymphoma and chronic lymphocytic leukemia to diffuse large B-cell lymphoma, *Haematologica*, 91, 1405–9 (PMID: 17018394.)
104. Jin W., Niu Z., Xu D., Li X. (2012), RBM5 promotes exon 4 skipping of AID pre-mRNA by competing with the binding of U2AF65 to the polypyrimidine tract, *FEBS Lett*, 586, 3852–7 (<https://doi.org/10.1016/j.febslet.2012.09.006>)
105. Zheng S., Vuong B.Q., Vaidyanathan B., Lin J.Y., Huang F.T., Chaudhuri J. (2015), Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA, *Cell*, 161, 762–73 (<https://doi.org/10.1016/j.cell.2015.03.020>)
106. Qiao Q., Wang L., Meng F.L., Hwang J.K., Alt F.W., Wu H. (2017), AID Recognizes Structured DNA for Class Switch Recombination, *Mol Cell*, 67, 361–373 (<https://doi.org/10.1016/j.molcel.2017.06.034>)

107. Saldanha R., Mohr G., Belfort M., Lambowitz A.M. (1993), Group I and group II introns, *FASEB J*, 7, 15–24 (<https://doi.org/10.1096/fasebj.7.1.8422962>)
108. Tourasse N.J, Kolstø A.B. (2008), Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution, *Nucleic Acids Res*, 36, 4529–48 (<https://doi.org/10.1093/nar/gkn372>)
109. Ramiro A.R., Stavropoulos P., Jankovic M., Nussenzweig M.C. (2003), Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand, *Nat Immunol*, 4, 452–6 (<https://doi.org/10.1038/ni920>)
110. Ricketts M.D., Dasgupta N., Fan J, Han J., Gerace M., Tang Y., et al. (2019), The HIRA histone chaperone complex subunit UBN1 harbors H3/H4- and DNA-binding activity, *J Biol Chem*, 294, 9239–59 (<https://doi.org/10.1074/jbc.RA119.007480>)
111. Craig S.P., Muralidhar M.G., McKerrow J.H., Wang C.C. (1989), Evidence for a class of very small introns in the gene for hypoxanthine-guanine phosphoribosyltransferase in *Schistosoma mansoni*, *Nucleic Acids Res*, 17, 1635–47 (<https://doi.org/10.1093/nar/17.4.1635>)
112. Russell C.B., Fraga D., Hinrichsen R.D. (1994), Extremely short 20-33 nucleotide introns are the standard length in *Paramecium tetraurelia*, *Nucleic Acids Res*, 22, 1221-12251. (<https://doi.org/10.1093/nar/22.7.1221>)
113. Lew B.M., Mills K.V., Paulus H. (1998), Protein splicing in vitro with a semisynthetic two-component minimal intein, *J Biol Chem*, 273, 15887–90 (<https://doi.org/10.1074/jbc.273.26.15887>)
114. Lockless S.W., Muir T.W. Traceless protein splicing utilizing evolved split inteins. *Proc Natl Acad Sci U S A* 2009;106:10999–1004. (<https://doi.org/10.1073/pnas.0902964106>)

115. Top O., Milferstaedt S.W.L., van Gessel N., Hoernstein S.N.W., Özdemir B., Decker E.L., et al. (2021), Expression of a human cDNA in moss results in spliced mRNAs and fragmentary protein isoforms, *Commun Biol*, 4, 964 (<https://doi.org/10.1038/s42003-021-02486-3>)

