ETH zürich

ETH Library

Characterization of complexes required for mammalian transcriptional and post transcriptional gene regulation

Doctoral Thesis

Author(s):
Qiu, Chunhong

Publication date:
2024

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000673495

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

DISS. ETH NO. 29857

Characterization of complexes required for mammalian transcriptional and post-transcriptional gene regulation

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zürich)

presented by

Chunhong Qiu

Master of Natural Science in Biology, USTC China

born on 21.04.1990

accepted on the recommendation of

Prof. Dr. Stefanie Jonas, examiner

Prof. Dr. Nenad Ban, co-examiner

Prof. Dr. Martin Jinek, co-examiner

2024

Abstract

Integrator (INT) is a large protein complex in metazoans, containing at least 16 subunits with a combined molecular weight of more than 1.5 MDa. This modular complex harbors dual enzymatic functionality with a cleavage module that has RNA endonuclease activity, and a phosphatase module with phosphatase activity. INT was initially identified as the 3'-end processing machinery for uridine-rich small nuclear RNAs (UsnRNAs). Later, it was found to be involved in the biogenesis of other non-coding RNAs, such as enhancer RNAs, long non-coding RNAs, and the precursor of micro RNAs etc. More recent studies show that INT plays a role in the premature termination of messenger RNA transcription at promoter-proximal pausing sites. INT recognizes the paused RNA polymerase II (RNAPII) that is bound by negative elongation factors NELF and DSIF, which stall RNAPII, then INT cleaves the nascent RNA and its phosphatase module counteracts RNAPII elongation. INT's function is crucial in all tissue types and developmental stages in metazoans. Disruption of INT subunits or mutations will lead to various genetic diseases with related neuropathological phenotypes. Until now, many studies were focused on INT's function and mechanism. In this thesis, we present our new understanding of the INT cleavage module's biogenesis. In addition, we also explore the reconstitution of INT with paused RNAPII. Furthermore, we present our approach to identify additional binding proteins during UsnRNA 3'-end processing.

Apart from INT biology, this thesis also presents our studies of three newly identified factors, FOCAD, AVEN, and HBS1LV3, which are interactors of the RNA exosome associated superkiller (SKI)-complex. The exosome and its cofactors play important roles in co-translational mRNA surveillance pathways during which defective transcripts are degraded to safeguard cells from aberrant transcripts. The three factors, FOCAD, AVEN, and HBS1LV3 were found to bind the SKI complex directly. In this thesis, we present our studies in understanding the architecture and functions of FOCAD, AVEN, HBS1LV3, and their influence on the SKI complex.

Chapters 3, 5, and 6 focus on INT biology. BRAT1 and WDR73 have been identified as interactors of INTS9 and INTS11 subunits of INT. Mutations and depletion of BRAT1 and WDR73 cause misregulation of INT's function. In Chapter 3, I resolved the cryo-electron microscopy (cryo-EM) structures of INTS9-INTS11-BRAT1-WDR73, INTS9-11-BRAT1, and INTS4-9-11-BRAT1, elucidating how BRAT1 and WDR73 associate with INTS9-11 in the cytoplasm, how INTS4 associates with INT9-11-BRAT1 in the nucleus as an intermediate of cleavage module maturation and how WDR73 and BRAT1 play a role in controlling the endonuclease activity of INTS11. Furthermore, we elucidated the stepwise assembly mechanism of the catalytic module and determined why inositol hexaphosphate is crucial for module assembly by using biochemical approaches. Chapter 5 shows our approach to obtain purified RNAPII to characterize the structure of INT with paused RNAPII. I optimized the purification of RNAPII and carried out parts of the EM data analysis. Chapter 6 shows our approach to find 3'-box binding factors that might assist INT in sequence specific UsnRNA processing. I identified 17 potential proteins and further analyzed their knockdown effect on 3'-end processing.

Chapter 4 focuses on the structural analysis of the SKI-complex cofactors FOCAD, AVEN, HBS1LV3. I resolved the cryo-EM structures of FOCAD, the FOCAD-AVEN complex and the HBS1LV3-SKI complex. These structures show, how HBS1LV3 associates with the SKIC3 subunits of the mammalian SKI complex, that FOCAD forms a dimer and how AVEN binds FOCAD. Furthermore, I analyzed the existence

of FOCAD and AVEN in other species based on a sequence and predicted structural similarities.

In summary, the findings in this thesis contribute to the understanding of the INT biogenesis, the mechanism of INT during the 3'-end processing of UsnRNAs, and the molecular basis of SKI complex function.

Zusammenfassung
Integrator (INT) ist ein großer Protein-Komplex in Metazoen, welcher mindestens 16 Untereinheiten mit einer Gesamtmasse von mehr als 1,5 Mda besitzt. Dieser modulare Komplex übernimmt zwei Funktionen: eine mit einem Spaltungsmodul, das RNA-Endonuklease-Aktivität aufweist, und einem Phosphatase-Modul mit Phosphatase-Aktivität. INT wurde ursprünglich als eine 3'-Endverarbeitungsmaschinerie für uridinreiche kleine, nukleare RNAs (UsnRNAs) identifiziert. Später stellte sich heraus, dass INT an der Biogenese anderer, nicht-kodierender RNAs beteiligt ist, wie z.B. Enhancer-RNAs, langer nicht-kodierender RNAs und dem Vorläufer von Mikro-RNAs, usw. Neuere Studien zeigen, dass INT eine Rolle bei der vorzeitigen Beendigung der Transkription von messenger-RNAs an Loci in der Nähe der Promotor-region spielt. INT erkennt eine pausierende RNA-Polymerase II (RNAPII), die von den negativen Elongationsfaktoren NELF und DSIF gebunden ist und somit RNAPII blockieren; dann schneidet INT die entstehende RNA und sein Phosphatase-Modul wirkt dem Elongationsmechanismus der RNAPII entgegen. Die Funktion von INT ist in allen Gewebetypen und Entwicklungsstadien bei Metazoen von wichtiger Bedeutung. Ein Defekt oder Mutationen in INT-Untereinheiten führen zu verschiedenen genetischen Krankheiten und resultieren in neuropathologischen Phänotypen. Bis jetzt konzentrierten sich viele Studien auf die Funktion und den Mechanismus von INT. In dieser Dissertation präsentieren wir unser neues Verständnis der Biogenese des Spaltungsmoduls von INT. Darüber hinaus erforschen wir auch die in vitro Rekonstitution von INT mit pausierter RNAPII. Weiterhin stellen wir unseren Ansatz, zusätzliche Bindungsproteine während der UsnRNA-3'-Endverarbeitung zu identifizieren, vor.

Abgesehen von der Biologie von INT präsentiert diese Dissertation auch unsere Studien zu drei neu identifizierten Faktoren, FOCAD, AVEN und HBS1LV3, welche Interaktoren des RNA-Exosom-assoziierten Superkiller (SKI)-Komplexes sind. Das Exosom und seine Cofaktoren spielen wichtige Rollen in der ko-translationalen mRNA-Überwachung, bei denen defekte Transkripte abgebaut werden, um Zellen vor aberranten Transkripten zu schützen. Es wurde festgestellt, dass die drei Faktoren FOCAD, AVEN und HBS1LV3 direkt mit dem SKI-Komplex interagieren. In dieser Dissertation beschreiben wir unsere Studien zum Verständnis der Architektur und Funktionen von FOCAD, AVEN, HBS1LV3, sowie deren Einfluss auf den SKI-Komplex.

Die Kapitel 3, 5 und 6 konzentrieren sich auf die Biologie von INT. BRAT1 und WDR73 wurden als Interaktoren der INTS9- und INTS11-Untereinheiten von INT identifiziert. Mutationen und eine Depletion von BRAT1 und WDR73 führen zu einer Fehlregulierung der Funktion von INT. In Kapitel 3 zeige ich die kryo-elektronenmikroskopischen (Kryo-EM)-Strukturen von INTS9-INTS11-BRAT1-WDR73, INTS9-11-BRAT1 und INTS4-9-11-BRAT1 und kläre darüber auf, wie BRAT1 und WDR73 mit INTS9-11 im Zytoplasma assoziieren. Ferner zeige ich, wie INTS4 mit INT9-11-BRAT1 im Kern als Zwischenprodukt der Spaltungsmodulreifung assoziiert ist und wie WDR73 und BRAT1 eine Rolle bei der Kontrolle der

Endonuklease-Aktivität von INTS11 spielen. Darüber hinaus haben wir den schrittweisen Aufbauechanismus des katalytischen Moduls aufgeklärt und mittels biochemischer Ansätze festgestellt, warum Inositolhexaphosphat für die Modulmontage entscheidend ist. Kapitel 5 zeigt unseren Ansatz zur Gewinnung gereinigter RNAPII zur Charakterisierung der Struktur von INT mit pausierter RNAPII. In meiner Arbeit habe die Reinigung von RNAPII optimiert und Teile der EM-Datenanalyse durchgeführt. Kapitel 6 zeigt unseren Ansatz, 3'-Box-Bindungsfaktoren zu finden, die INT bei der sequenzspezifischen UsnRNA-Verarbeitung unterstützen könnten. Ich habe 17 potenzielle Proteine identifiziert und ihre Knockdown-Wirkung auf die 3'-Endverarbeitung weiter analysiert.

Kapitel 4 konzentriert sich auf die strukturelle Analyse der Cofaktoren des SKI-Komplexes FOCAD, AVEN, HBS1LV3. Die Kryo-EM-Strukturen von FOCAD, dem FOCAD-AVEN-Komplex und dem HBS1LV3-SKI-Komplex konnten gelöst werden. Diese Strukturen zeigen, wie HBS1LV3 mit den SKIC3-Untereinheiten des SKI-Komplexes von Säugern assoziiert ist, dass FOCAD ein Dimer bildet und wie AVEN an FOCAD bindet. Darüber hinaus habe ich die Existenz von FOCAD und AVEN in anderen Arten auf der Grundlage von Sequenz- und vorhergesagten strukturellen Ähnlichkeiten analysiert.

Zusammenfassend tragen die Ergebnisse dieser Dissertation zum Verständnis der Biogenese von INT, des Mechanismus von INT während der 3'-Endverarbeitung von UsnRNAs und der molekularen Grundlagen über die Funktion des SKI-Komplexes bei.