Biochimica et Biophysica Acta 1859 (2016) 1269–1280

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Review

Integrator complex and transcription regulation: Recent findings and pathophysiology☆

Monica Rienzo, Amelia Casamassimi *

Department of Biochemistry, Biophysics and General Pathology, Second University of Naples, Via L. De Crecchio 7, 80138 Naples, Italy

ARTICLE INFO

Article history:
Received 18 March 2016
Received in revised form 8 July 2016
Accepted 12 July 2016
Available online 15 July 2016

Keywords:
Integrator complex
Transcription regulation
DNA and RNA metabolism
Development and differentiation
Human diseases

ABSTRACT

In the last decade, a novel molecular complex has been added to the RNA polymerase II-mediated transcription machinery as one of the major components. This multiprotein complex, named Integrator, plays a pivotal role in the regulation of most RNA Polymerase II-dependent genes. This complex consists of at least 14 different subunits. However, studies investigating its structure and composition are still lacking. Although it was originally discovered as a complex implicated in the 3′-end formation of noncoding small nuclear RNAs, recent studies indicate additional roles for Integrator in transcription regulation, for example during transcription pause-release and elongation of polymerase, in the biogenesis of transcripts derived from enhancers, as well as in DNA and RNA metabolism for some of its components. Noteworthy, several subunits have been emerging to play roles during development and differentiation; more importantly, their alterations are likely to be involved in several human pathologies, including cancer and lung diseases.

1. Introduction

RNA Polymerase II (RNAPII)-mediated transcription is a molecular process that is controlled at several steps, from the recruitment and assembly of the entire transcription apparatus, to the initiation, elongation and termination of transcription; moreover, also various RNA maturation events, like splicing, occurring together with transcription, constitute an additional level of complexity in the gene regulation process [1–3]. These multiple regulatory mechanisms involve numerous molecules and molecular factors establishing larger networks. In the last decade, a novel complex has been added as one of the components of the RNAPII-mediated transcription machinery for most regulated genes, named Integrator (INT), which is a multisubunit complex [4]. Although it is still considered to be a metazoan-specific complex in the recent literature, sequence databases indicate that INT orthologues are also present in plants (see *National Center for Biotechnology Information* online database). Conceivably, snRNA promoters of plants are similar in structure to animal snRNAs thus suggesting that the whole mechanism may be conserved in all the multicellular eukaryotes [5–7]. The initial affinity purification of INT identified twelve subunits (IntS1 to IntS12) and demonstrated its association with the C-terminal domain (CTD) of RPB1, the largest subunit of RNAPII [8]. In humans, these proteins are annotated in numerical order on the basis of gel migration, with IntS1 having the largest predicted molecular mass of 244 kDa and IntS12 being the smallest one with 49 kDa [9]. Subsequent proteomic analyses confirmed its composition and identified further subunits as well as new potential cofactors in both *Drosophila* and human genomes. Among them, C12orf11 (also known as ASUN, or Asunder in *Drosophila*) and C15orf44 (also named VWA9, or CG4785 in *Drosophila*) were proven to be functionally associated with INT and were renamed IntS13 and IntS14, respectively [4,10]. By size exclusion chromatography, the whole molecular weight is estimated to be greater than 1 MDa, since most of IntSs molecular weights are >100 kDa [8,4,11]. The most common predicted motifs within INT subunits are α-helical repeats (HEAT, ARM and TPR or VWA domains), suggestive of protein–protein interaction surfaces [4,9] (see Table 1). In this context, IntS9 and IntS11 contain the most identifiable motifs present in this complex [4]. Indeed, IntS9 and IntS11 are homologous to CPSF100 (cleavage and polyadenylation specificity factor 100 kDa subunit) and CPSF73, respectively [12]; these factors are essential for 3′-end cleavage of RNAPII transcribed messenger RNAs (mRNAs) including histone mRNAs since they catalyze cleavage at the polyadenylation site, directed by the polyadenylation signal localized upstream [12–14]. These subunits have been also isolated as a heterodimeric complex (IntS9/IntS11); this specific interaction is mediated by a distinct domain localized at the extremity of IntS9 C-terminus and within the C-terminus of IntS11, adjacent to the site of endonuclease activity required for snRNA processing [15]. Away from these two subunits, the other members of the complex display little similarity with the proteins involved in either poly(A) or histone mRNA 3′-end formation. Nevertheless, another protein–protein interaction is known to occur also between IntS3 and IntS6; indeed, these subunits were shown to participate, together with

☆ Conflict of interest: the authors declare that they have no conflict of interest.

* Corresponding author.
E-mail address: amelia.casamassimi@unina2.it (A. Casamassimi).

http://dx.doi.org/10.1016/j.bbagrm.2016.07.008
1874-9399/

Table 1
Human INTSs and their possible pathophysiological roles.

Name	Other gene names	Human chromosome	RefSeq locus	Functional domains	Pathophysiological role
INTS1	INT1; NET28	7p22.3	NM_001080453	DUF3677	Development at early blastocyst stage Hematopoietic development
INTS2	INT2; KIAA1287	17q23.2	NM_020748	–	Gastric cancer peritoneal carcinomatosis
INTS3	INT3; SOSSA; SOSS-A; C1orf60; C1orf193	1q21.3	NM_023015	DUF2356	DNA damage response and maintenance of genome stability Overexpressed in hepatocellular carcinoma
INTS4	INT4; MGC16733, MST093	11q14.1	NM_033547	ARM HEAT	Development
INTS5	INT5; KIAA1698	11q12.3	NM_030628	–	Hematopoietic development Postmenopausal osteoporosis
INTS6	INT6; HDB; DBI-1; DDX26; DICE1; DDX26A; Notchl2	13q14.3	NM_001039937	VWA ISDCC	Adipogenesis Overgrowth Participates in the DNA damage response Together with its pseudogene INTS6P1 plays key roles in the pathogenesis of hepatocellular carcinoma Childhood B-precursor acute lymphoblastic leukemia
INTS7	INT7; C1orf73	1q32.3	NM_015434	ARM	DNA damage response Multiple roles in development
INTS8	INT8; C8orf52	8q22.1	NM_017864	TPR	Putative biomarker for gastric cancer Mutated in peripheral T cell lymphoma
INTS9	INT9; RC74; CPSF2L	8p21.1	NM_001145159	β-lactamase β-CASP	Possible role in malignancies
INTS10	INT10; C8orf35	8p21.3	NM_018142	–	Nicotine dependence (GWAS) Childhood B-precursor acute lymphoblastic leukemia
INTS11	INT11; RC68; RC-68; CPSF3L; CPSF73L	1p36.33	NM_017871	β-lactamase β-CASP RMMBL	Adipogenesis Hematopoietic development Possible role in malignancies
INTS12	INT12; PHF22; SBBI22	4q24	NM_001142471	PHD	Lung function and pulmonary diseases
INTS13 (ASUN)	Asunder; C12orf11; GCT1; NET48; Mat89Bb; SPATA30	12p12.3	NM_018164	COIL	Spermatogenesis and oogenesis Cell division
INTS14 (VWA9)	C15orf44; VWA9; CG4785	15q22.31	NM_001207058	VWA	Elevated expression in SV40-immortalized cells, cancer cells, and NSCLC tissues

Abbreviations: ARM, armadillo like repeats; COIL, coiled coil domain; DUF, domain of unknown function; ISDCC, INTS6/SAGE1/DDX26B/CT45 C terminus; PHD, plant homeodomain finger; TPR, tetratricopeptide repeats; b-lactamase/b-CASP; VWA, von Willebrand type A like domain.

Nucleic Acid Binding Proteins (NABPs), to the formation of other protein complexes involved in DNA repair [16,17].

However, to date most of the INT subunits are yet to be fully characterized and structural studies depicting the complete architecture of this complex are still lacking. Moreover, studies designed to understand the exact role of the different INT subunits and possible modules in the various biological processes, where they are involved, are also needed to gain insights about INT function.

2. INT functions in RNAPII-mediated transcription: first and last evidences

Since its first description, INT has been implicated in the 3′-end formation of noncoding uridine-rich small nuclear RNA (snRNA) [4,8]. However, in the last years, both genome-wide analyses and the increasing number of studies describing its possible role in developmental defects and diseased states upon dysfunction (see also below) have promptly indicated broader functions for this complex extending its role to other aspects of transcriptional regulation, which are described below.

2.1. INT function in snRNA processing

In addition to protein-coding genes, mammalian RNAPII also transcribes diverse genes for non-coding RNAs, including the best-characterized genes encoding the spliceosomal U1 and U2 snRNAs. Uridine-rich snRNAs are ubiquitously expressed and are essential for the removal of introns, proper expression of histone mRNA and biosynthesis of rRNA [9,15,18]. These genes differ from protein-coding genes in their unique promoter structure, mechanism of 3′-end formation and transcription termination. Indeed, snRNAs are characterized by a relatively basic gene structure: no TATA box, no introns, no polyadenylation [9,15,18]. A snRNA promoter typically contains two characteristic elements: a distal sequence element (DSE) that recruits the transcription factors Oct1 and Sp1, and a proximal sequence element (PSE) that is bound by a pentameric factor, the snRNA activating protein complex (SNAPC), in addition to the general transcription factors (TFIIA, TFIIB, TFIIE and TFIIF) [18–21] (Fig. 1). Moreover, the 3′-end formation is distinctive in the different classes of RNAPII transcripts, poly(A)$^+$ mRNA, histone mRNA and snRNA [9,15,20,22,23]. The 3′-end of snRNA genes contains a consensus sequence (GTTTN$_{0-3}$AAARNNAGA), called 3′-box sequence, located 9–19 nucleotides downstream of their 3′-end [24]. As regard to the involved mechanism, the initial work could not determine whether the 3′-end was formed by transcription termination or by co-transcriptional cleavage although the authors favored the first hypothesis [22,23,25]. This open question has remained until the discovery of INT complex, when cleavage has become the preferred interpretation as a result of the homology to the polyadenylation factors [8,12]. Indeed, according to the current accepted view, once INT complex is recruited to snRNA promoters, it associates with the RNAPII-CTD and cleaves the nascent pre-snRNA as the 3′-box is transcribed and recognized, probably through an interaction between an INT subunit with the stem–loop at the end of the snRNA [9,15]. Thus, similarly to other RNAPII transcripts, the two events, capping of the 5′-end and cleavage of the 3′-end, are supposed to occur co-transcriptionally [9,15]. Importantly, INT is likely to cleave the pre-snRNA using its catalytic RNA endonuclease in the IntS9/IntS11 heterodimeric cleavage factor of the complex [4]. As mentioned above, this notion is essentially based on their high sequence similarity to CPSF73 and CPSF100, which are known to be involved in the cleavage reaction of other RNAPII transcripts, since INT has not yet been isolated and observed to work *in vitro* [12–15,26]. Knockdown experiments of these subunits, which are members of the β-CASP family, in both human and *Drosophila* cells, abolished proper 3′-end formation of snRNAs [8,13,26]. Moreover, a model was recently proposed where the major difference between the control of transcription termination in human