Biochemical and mass spectrometry analysis of an HIV-1 ribonucleoprotein export complex

Mizar Francesca Oliva

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

Mizar Francesca OLIVA

Thèse dirigée par **Carlo PETOSA,** Responsable d'équipe**,** CNRS, et codirigée par **Fabienne HANS,** Maître de Conférences**,** UGA

préparée au sein de l' **Institut de Biologie Structurale**
dans l'**École Doctorale Chimie et Sciences du Vivant**

Biochemical and mass spectrometry analysis of an HIV-1 ribonucleoprotein export complex

(Analyse biochimique et par spectrométrie de masse d'un complexe ribonucléoprotéique d'export du VIH-1)

Thèse soutenue publiquement le **23 Mai 2017**,
devant le jury composé de :

Monsieur Marc LAVIGNE
Chargé de recherche, Institut Pasteur de Paris (Rapporteur)
Monsieur Frank SOBOTT
Professeur des universités, Université de Anvers (Rapporteur)
Madame Claire VOURC'H
Professeur des universités, Université Grenoble Alpes (President)
Monsieur Edouard BERTRAND
Directeur de recherche, Institut de Génétique Moléculaire de Montpellier (Examinateur)
Madame Montserrat SOLER LOPEZ
Chercheuse, ESRF, Grenoble (Examinateur)

LIST OF ABBREVIATIONS

IA	Stem IA
IIA	Stem IIA
IIB	Stem IIB
IIC	Stem IIC
IIABC	Stem IIABC
III	Stem III
IV	Stem IV
V	Stem V
AmAc	Ammonium Acetate
ARE	AU rich element
ARM	arginine rich motif
BSA	bovine serum albumin
bp	base pair
CID	collision induced dissociation
CRM1	chromosome region maintenance-1
DAC	diammonium hydrogen citrate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
ESI	electrospray Ionization
FP	fluorescence polarization
GDP	guanosine diphosphate
GFP	green fluorescent protein
GTP	guanosine triphosphate
FWHM	full width at half maximum
HHR	hammerhead rybozyme
HIV	human immunodeficiency virus
HPLC	high performance liquid chromatography
HuR	human antigen R
IMAC	immobilised metal affinity chromatograpy
Kaps	Karyopherins-β
kb	kilo base pairs
LB	Luria-Bertani
LC	liquid chromatography
LRPPRC	leucine-rich pentatricopeptide repeat protein

MALDI	matrix assisted laser ionization
MCS	multi-cloning site
MS	mass spectrometry
NE	nuclear envelope
NES	nuclear export signal
NLS	nuclear localization signal
nMS	native mass spectrometry
NPC	nuclear pore complex
nt	nucleotide
NTRs	nuclear transport receptors
OD	oligomerization domain
ORFs	open reading frames
PCR	polymerase chain reaction
PIC	pre-integration complex
PKI	protein kinase inhibitor
PMSF	phenylmethylsulfonyl fluoride
Q-TOF	quadrupole time-of-flight mass spectrometer
RBD	RNA binding domain
RNA	ribonucleic acid
RNP	ribonucleoprotein
RNPs	ribonucleoprotein particles
RRE	Rev Responsive Element
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension
TEV	Tobacco Etch Virus
TOF	time of flight

ABSTRACT

An important step in the life cycle of human immunodeficiency virus (HIV) is the nuclear export of incompletely spliced viral transcripts, including the replicated viral RNA genome. This process is mediated by the viral RNA-binding protein Rev. In the nucleus, Rev recognizes unspliced and partially spliced viral transcripts by multimerizing on a 350-nucleotide intronic sequence, the Rev-response element (RRE). Rev then recruits the host cell export factor CRM1 and the small GTPase Ran to form the RRE/Rev/CRM1/Ran export complex. Knowledge of the 3D architecture of this ribonucleoprotein complex would provide important insights into how unspliced viral RNA export is achieved. However, the molecular details of this complex are poorly understood. In particular, the stoichiometry of Rev and CRM1 molecules bound to the RRE is still uncertain.

My Ph.D. project aims to investigate the architecture of the RRE/Rev/CRM1/Ran complex. As part of this work, I used biochemical and cell-based assays to characterize the interactions between CRM1 and Rev and between Rev and the RRE. The majority of my efforts focused on investigating these interactions by native mass spectrometry (MS), a powerful method for determining the stoichiometry of macromolecular complexes. I set up protocols for the large-scale preparation of a 66-nucleotide RRE fragment (IIABC) bearing a high-affinity Rev binding site, and adapted these for compatibility with native MS analysis. Because Rev tends to aggregate and precipitate in solution, I engineered a mutant form of Rev (Rev*) to overcome this problem. Analysis of IIABC/Rev* complexes by native gel electrophoresis confirms multimerization of Rev on the RNA. After extensive optimization, I obtained high-quality native MS spectra of these complexes, revealing that IIABC binds up to 6 Rev* monomers. Furthermore, I reconstituted a 4-species complex, IIABC/Rev*/CRM1/Ran, and succeeded in determining its mass and stoichiometry by native MS – a technically challenging task. Additional efforts at analyzing the intact RRE and complexes with wild-type Rev have also yielded informative spectra, while analysis of the intact RRE/Rev/CRM1/Ran holo-complex has had more limited success. These results illustrate the strengths and limitations of native mass spectrometry and its potential for future development as a tool for analyzing ribonucleoprotein complexes.

RÉSUMÉ

Une étape importante du cycle viral du virus de l'immunodéficience humaine (VIH) est l'export nucléaire de transcrits viraux incomplètement épissés, incluant le génome viral ARN. Ce processus fait intervenir la protéine virale de liaison à l'ARN Rev. Dans le noyau, Rev interagit avec les transcrits viraux non épissés et partiellement épissés en s'oligomérisant sur une séquence intronique de 350 nucléotides, appelée Element de Response à Rev (RRE). Rev recrute également le facteur d'export cellulaire CRM1 et la petite GTPase Ran pour former le complexe d'export RRE/Rev/CRM1/Ran. Connaître l'architecture 3D de ce complexe ribonucléoprotéique fournirait des informations utiles pour une meilleure compréhension de l'export des ARN du VIH incomplètement épissés. Cependant, les détails moléculaires de ce complexe sont mal connus ; en particulier, la stœchiométrie des molécules Rev et CRM1 liées au RRE est en discussion.

Mon doctorat vise à étudier l'architecture du complexe RRE/Rev/CRM1/Ran. Dans le cadre de ce travail, j'ai utilisé des essais biochimiques et cellulaires pour caractériser les interactions entre CRM1 et Rev et entre Rev et RRE. La majorité de mes efforts ont porté sur l'étude de ces interactions par spectrométrie de masse (MS) en condition native, une méthode puissante pour déterminer la stœchiométrie de complexes macromoléculaires. J'ai mis en place des protocoles pour la préparation à grande échelle d'un fragment du RRE de 66 nucléotides (IIABC), portant un site de liaison Rev de haute affinité, protocoles que j'ai ensuite adaptés à l'analyse de IIABC par MS en condition native. Comme Rev a tendance à s'agréger et à précipiter en solution, j'ai également conçu une forme mutante de Rev (Rev*) permettant de contourner ces problèmes. L'analyse des complexes IIABC/Rev* par électrophorèse sur gel natif confirme l'oligomérisation de Rev* sur l'ARN. Après d'intenses optimisations, j'ai obtenu des spectres MS en condition native de haute qualité, révélant que IIABC lie jusqu'à 6 monomères Rev*. De plus, j'ai reconstitué un complexe à 4 partenaires IIABC/Rev*/CRM1/Ran et j'ai réussi à déterminer sa masse et sa stœchiométrie par MS en condition native, une tâche techniquement difficile. Des efforts supplémentaires pour analyser le RRE seul et en complexe avec Rev de type sauvage ont également généré des spectres informatifs, alors que l'analyse du complexe intact RRE/Rev/CRM1/Ran a été plus compliquée. Ces résultats illustrent les forces et les limites de la spectrométrie de masse en condition native et son potentiel pour son développement futur en tant qu'outil d'analyse des complexes de ribonucléoprotéines.

1. INTRODUCTION

ABSTRACT

To complete its replication the HIV-1 must export intron-containing viral RNAs from the nucleus to the cytosol. In eukaryotes intron-containing mRNAs are retained in the nucleus and degraded, but HIV-1 exploits a non-canonical pathway to export intron-containing viral mRNAs. The early expressed regulatory viral protein Rev binds on an intronic 350-nt RNA element, the Rev response element (RRE). Rev then recruits the host cell export factor CRM1 and the small GTPase Ran to form the RRE/Rev/CRM1/Ran export complex. It is known that multiple copies of Rev are required for the efficient export of viral mRNAs and that most likely also multiple copies of CRM1 are needed. The molecular details and structure of this RNP are still poorly understood, hindered by its highly dynamicity. In particular the stoichiometry of Rev and CRM1 needed to produce an efficiently exported RNP are still uncertain. The aim of this project is to investigate the stoichiometry of this RNP using native mass spectrometry, an emerging and powerful tool to determine the stoichiometry of macromolecular complexes.

RÉSUMÉ

Pour compléter sa réplication, le VIH-1 doit exporter du noyau au cytosol des ARN viraux contenant des introns. Dans les cellules eukaryotes, les ARN contenant des introns sont retenus dans le noyau et dégradés, mais le VIH-1 utilise une voie d'export non canonique pour exporter les ARN contenant des introns. La protéine virale de régulation Rev, exprimée dans la phase d'infection précoce, se lie à un élément intronique d'ARN de 350-nt, l'élément de réponse à Rev (RRE). Rev recrute alors le facteur d'export cellulaire CRM1 et la petite GTPase Ran pour former le complexe d'export RRE/Rev/CRM1/Ran. Il est connu que plusieurs copies de Rev sont nécessaires pour exporter efficacement ces ARN viraux et que, très probablement, plusieurs copies de CRM1 sont aussi nécessaires. Les détails moléculaires ainsi que la structure de cette RNP ne sont pas encore complètement compris, rendus difficiles par la dynamique du complexe. En particulier, le nombre de Rev et CRM1 nécessaires pour une export efficace de cette RNP est encore incertain. Le but de ce projet est d'étudier la stœchiométrie de cette RNP en utilisant la spectrométrie de masse native, un outil émergeant et puissant pour déterminer la stœchiométrie des complexes macromoléculaires.

1.1 RIBONUCLEOPROTEIN EXPORT MEDIATED BY CRM1

In eukaryotes, the genetic material is confined within the nucleus, separated from the rest of the cell by the nuclear envelope. This compartmentalisation requires nuclear and cytosolic activities to be well coordinated, implying a highly regulated trafficking of macromolecules, such as RNA molecules, between the nucleus and cytosol. In the cell, RNA molecules such as tRNA, mRNA, snRNA and rRNA associate with RNA-binding proteins to form various ribonucleoprotein particles (RNPs), namely, tRNPs, mRNPs, snRNPs, and ribosomal subunits, respectively. RNPs are highly dynamic complexes whose composition and structure vary throughout the lifetime of the particle, depending on the degree of processing of the concerned RNA molecules. During their biogenesis most RNPs travel from the nucleus to the cytosol and vice-versa through the nuclear pore complex (NPC) (Figure 1.1 A) with the help of nuclear transport receptors (NTRs). This first section describes two principal actors involved in nuclear export: the NPC and the karyopherins, with a special emphasis on CRM1 which mediates the export of many RNPs.

1.1.1 The Nuclear Pore Complex: the gateway between nucleus and cytosol

The inner and outer membranes of the nuclear envelope (NE) are fused at multiple sites by a large protein assembly called the nuclear pore complex (NPC) (Figure 1.1 A). NPCs are highly specialized channels that regulate the passage of macromolecules between the nucleus and cytosol. The NPC is one of the largest protein complexes in the cell, with an estimated mass of ~60 MDa in yeast and ~120 MDa in higher eukaryotes (Beck and Hurt, 2017). The NPC is composed of multiple copies (8-64) of approximately 30 different proteins called nucleoporins (Nups), arranged around the central channel with a pseudo-eightfold rotational symmetry. A schematic structure of the NPC is presented in Figure 1.1 A, where different substructures can be distinguished: the inner pore ring, situated at the fused nuclear membranes; the nuclear and cytoplasmic rings, which are anchored to the inner pore ring; and the nuclear basket and cytoplasmic filaments, which are peripheral elements emanating respectively from the nuclear and cytoplasmic rings. The height of the NPC is estimated to be ~95 nm, with an external and internal diameter of ~125 nm and ~55 nm, respectively (von Appen and Beck, 2016).

The total number of Nups present in a single NPC is estimated to be ~500-1000 and they are grouped into distinct subcomplexes (Figure 1.1 B) (Kabachinski and Schwartz, 2015). The NPC scaffold contains two classes of Nups: integral membrane proteins (Figure 1.1 B, yellow boxed Nups) which constitute the core scaffold fused to the membrane, and other scaffold Nups, which are anchored to these to form the different ring structures (Figure 1.1 B, red and blue boxed Nups).

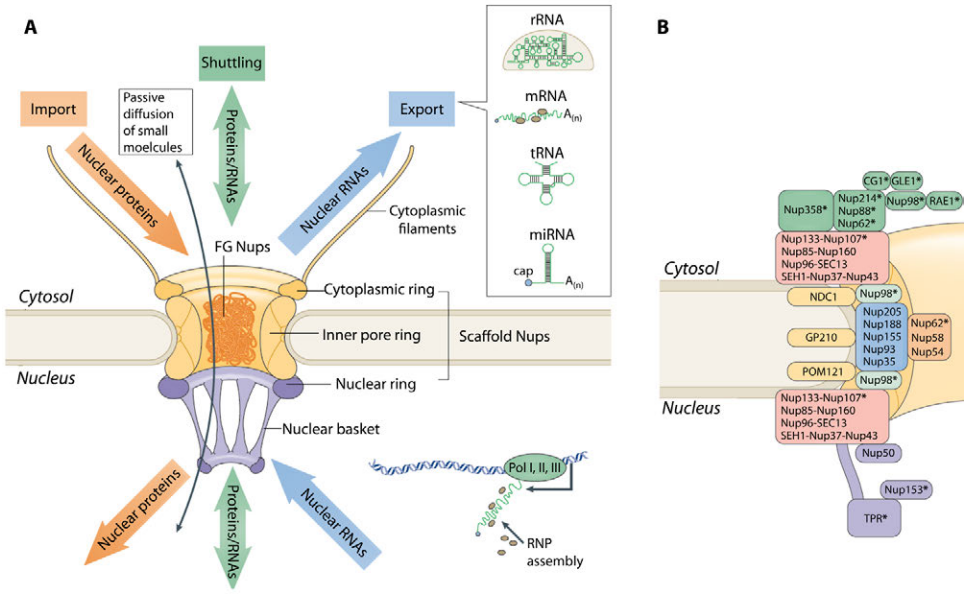


Figure 1.1: NPC architecture, transport functions and molecular composition. A) Principal sub-structures and transport functions. The inner pore ring, cytoplasmic and nuclear rings, cytoplasmic filaments and nuclear basket are indicated. Colored arrows indicate the import of proteins from the cytosol to the nucleus (orange), the shuttling of proteins and RNAs in both directions (green), and the export of RNAs and RNPs from the nucleus to the cytosol (blue). The inset indicates some classes of RNA that are exported. B) Molecular composition and localization of conserved Nups. Colored boxes indicate Nups that localize to: the membrane (yellow), the inner pore ring (blue), both the nuclear and cytoplasmic rings (red), the cytoplasmic filaments (green) and the nuclear basket (purple). Nups implicated in RNA transport are indicated by an asterisk. *(adapted from Beck and Hurt, 2017).*

The channel is composed of so-called FG Nups that mediate the passage of macromolecules across the NPC. FG Nups contain Phe-Gly repeats that are hydrophobic and very flexible domains forming the permeability barrier of the NPC. FG-repeats transiently interact with low affinity with proteins called nuclear transport receptors (NTRs) which escort specific macromolecules across the NPC (Rexach and Blobel, 1995).

1.1.2 KARYOPHERINS: A SUPERFAMILY OF NUCLEAR TRANSPORT RECEPTORS

While ions, metabolites and small molecules (<40 kDa) can freely pass across the NPC, larger macromolecules (>40 kDa, i.e large proteins, protein complexes, tRNA, snRNP, mRNP, ribosomal subunits) must be actively transported by protein carriers referred to as nuclear transport receptors (NTRs). Most NTRs belong to the karyopherin-β (Kap) superfamily, which includes more than 20 members in mammals (Mosammaparast and Pemberton, 2004). Kaps are soluble factors that shuttle between the nucleus and cytosol mediating most transport processes across the NPC by collectively recognizing a wide diversity of cargos, ranging from simple protein and RNA cargos to complex ribonucleoproteins (RNPs) (Figure 1.2) (Güttler et al., 2010; Matsuura, 2016).

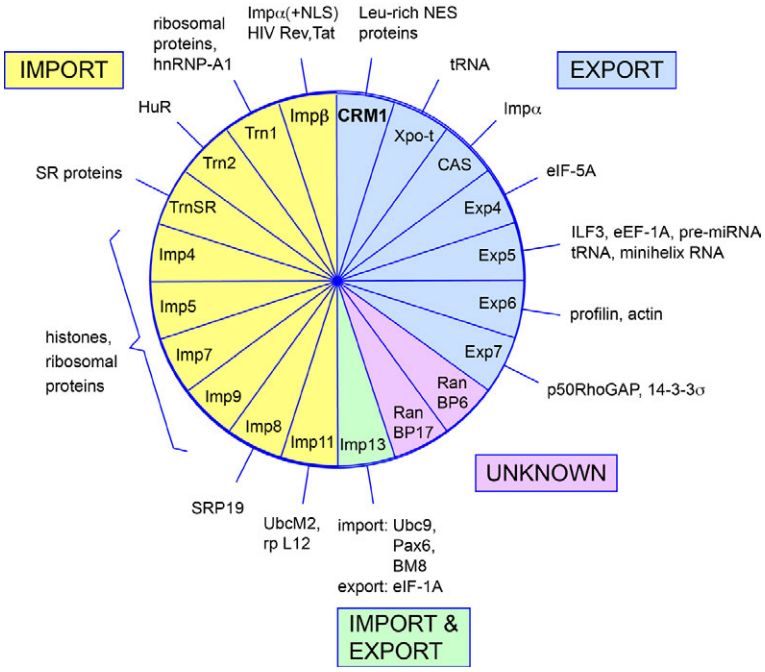


Figure 1.2: The karyopherin-β family in mammals. Colours indicate Kaps which mediate nuclear export (blue), import (yellow) or both import and export (green) as well as Kaps of uncertain function (purple). Selected known cargos are shown next to each Kap. *(adapted from Mosammaparats and Pemberton, 2004).*

Depending on the direction of transport, Kaps are either classified as importins, (transporting cargos from the cytosol to the nucleus), exportins (transporting in the opposite direction); or bidirectional carriers (possessing both functions).

Some Kaps are specialized in the transport of a single cargo or of a restricted class of related cargoes (i.e, Imp4, Trn1, Exp-t, CAS) whereas others transport a wide variety of cargos (i.e., the importin heterodimer Imp-β/ Imp-α, CRM1) (Figure 1.2). The

nucleocytoplasmic transport of many protein cargos depends on the presence of short polypeptide which are recognized by specific importins or exportins. For example, a short lysine-rich stretch of residues constitutes a basic nuclear localization signal (NLS) which is recognized in the cytosol by the importin heterodimer Imp-β/Imp-α, leading to import of the NLS-bearing cargo into the nucleus (Gorlich and Mattaj, 1996). Another motif called nuclear export signal (NES) is recognized by the exportin CRM1 , which mediates the export of NES-bearing cargos to the cytosol. The NES is a short leucine-rich motif (8-15 residues), described in further detail below (see § 1.1.3 for more details concerning NES characteristic). Some proteins (e.g., HIV-1 Rev) contain both a NLS and a NES, allowing then to shuttle between both cellular compartments. The affinity between the importin/exportin and NLS/NES is modulated by other proteins (the most important of these is the small GTPase Ran), allowing cargo binding and release during the transport.

Small RNA cargos (e.g., tRNAs, miRNAs) are recognized by karyopherins such as Exp-t and Exp5 through short nucleotide motifs (Güttler and Gorlich, 2011). Certain larger RNAs (e.g., snRNAs, rRNA) are exported by CRM1 (Okamura et al., 2015). However, CRM1 does not directly bind these RNAs but associates indirectly via one or more adaptor proteins, which specifically recognize the RNA cargo and possess an NES recognized by CRM1. Examples of such CRM1-mediated export are discussed below (§ 1.1.3).

Despite sharing low sequence identity (10-20%) all Kaps show a similar architecture, have similar biophysical properties (e.g., a molecular mass of 90-150 kDa and an isoelectric point of 4.0-5.5) and all interact with FG-Nups and with the small GTPase Ran (Matsuura, 2015). Kaps are composed of 19 to 21 tandem repeats of a ~50 residue motif called the HEAT repeat, which folds as two antiparallel α-helices (helices A and B) connected by a short loop (Figure 1.3 A) (Andrade and Bork, 1995; Cingolani et al., 1999; Chook et al., 1999).

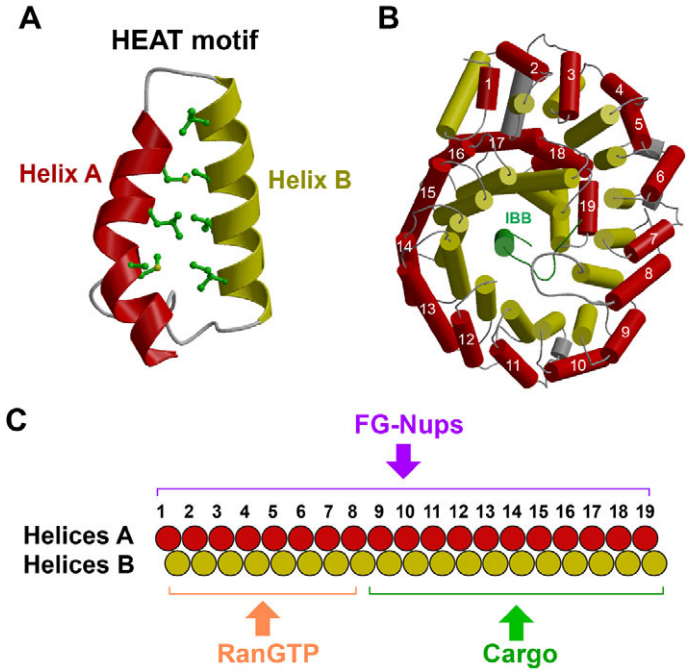


Figure 1.3: **Karyopherin-β architecture.** A) Structure of a HEAT repeat. The antiparallel helices A and B are colored red and gold, respectively. B) Structure of Impβ in complex with the Impβ binding domain (IBB) of Impα (in green). HEAT repeats are numbered on the A helix from N- to C-terminus (*Adapted from Cingolani et al., 1999*). C) Schematic organization of the HEAT repeats showing the regions that interact with Ran^{GTP} (orange), cargo (green) and Nups (magenta).

The array of HEAT repeats form a superhelical structure, also termed a "solenoid", where the A helices define the outer convex surface and the B helices form the inner concave surface (Andrade et al., 2001) (Figure 1.3 B). The conformation of the solenoid differs among Kaps and it opens or closes upon binding the different partners. This flexibility is a common characteristic of Kaps and is critical for their function (Conti et al, 2006). The outer surface of the solenoid interacts with FG Nups, whereas for many Kaps the N- and C-terminal regions of the inner surface interact with Ran and with the cargo, respectively (Figure 1.3 C).

As mentioned above, the interaction of importins and exportins with their cargo is regulated by Ran, a ~25 kDa member of the Ras-related GTP-binding protein superfamily (Joseph, 2006). Like other GTPases, Ran alternates between a GDP-bound and a GTP-bound form (Drivas et al., 1990). Ran crystal structures reveal how the switch between these two forms is associated with a large change in conformation (Figure 1.4 A) (Stewart et al., 1998; Vetter et al., 1999). Specifically, three regions undergo conformational changes

upon nucleotide exchange: switches I and II have partial secondary structure in the GDP-form and melt upon GTP binding; switch III (C-terminal switch) contains a C-terminal helix which interacts with the core domain in the GDP-bound state, whereas it adopts an extended, partly disordered conformation in the GTP-bound state (as indicated by the dotted lines).

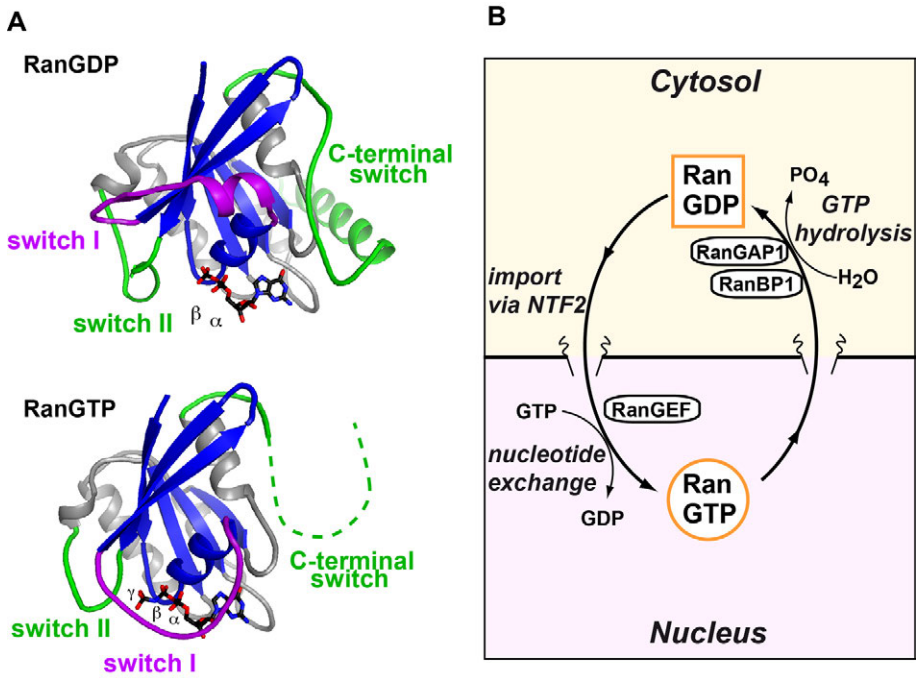
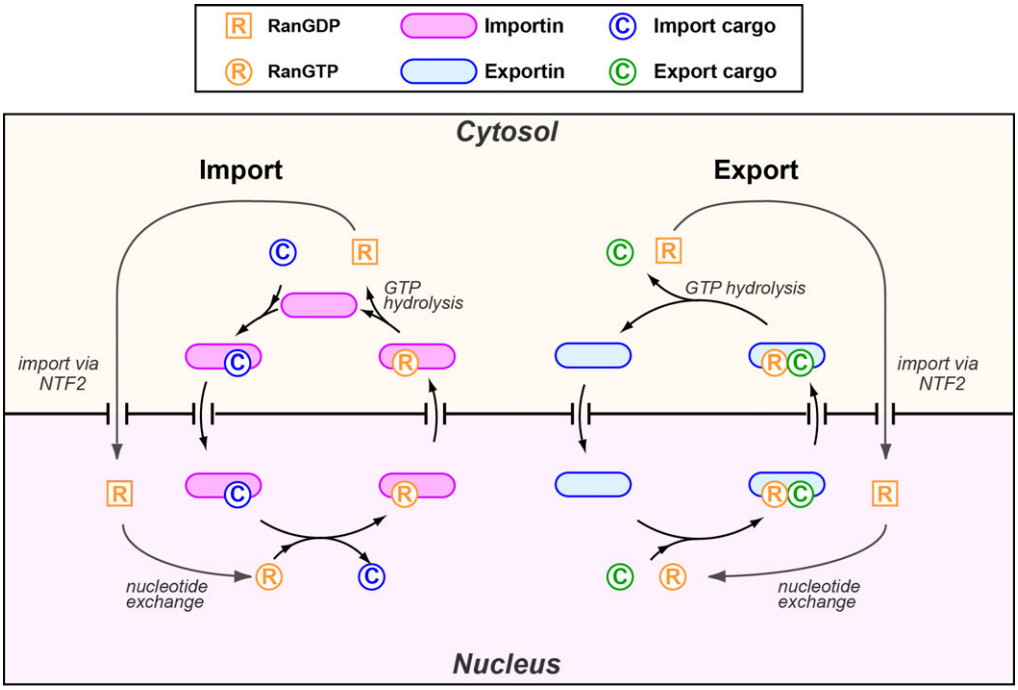


Figure 1.4: **Ran structures and regulation of the nucleotide state.** A) Structure of Ran in the GDP- and GTP-bound states (PDB codes 1BYU and 1IBR, respectively). The core domain (G domain, colored in blue and grey) folds as a central six-stranded β-sheet surrounded by peripheral α-helices and loops, the latter containing several residues or motifs highly conserved among guanine nucleotide binding proteins. Regions that undergo large conformational changes are highlighted in magenta (switch I) and green (switch II and C-terminal tail). B) The Ran nucleotide state is regulated by accessory factors, which are asymmetrically distributed across the nuclear membrane. (*adapted from Petosa, 2012).*

In vivo the transition between the Ran^{GDP} and Ran^{GTP} states is regulated by accessory factors (Figure 1.4 B). The exchange of GDP with GTP is regulated by a guanine nucleotide exchange factor, RanGEF (called RCC1 in mammals), whereas GTP hydrolysis is accelerated by a GTPase activating protein, RanGAP, and by additional co-factors called Ran-binding proteins (RanBPs) (Bischoff and Ponstingl, 1991; Bischoff et al., 1994). An additional factor involved in Ran regulation is NTF2, which shuttles Ran^{GDP} from the cytosol to the nucleus (Ribbeck et al., 1998; Smith et al., 1998). These regulators are

asymmetrically distributed: RanGEF (RCC1) is chromatin bound and thus nuclear (Nemergut et al., 2001), whereas RanGAP and RanBP1 are cytoplasmic (Mahajan et al., 1997; Richards et al., 1996) and RanBP2 (Nup358) is localized on the cytoplasmic ring of the NPC (Yokoama et al., 1995). In the cytosol, GTP hydrolysis by Ran is greatly enhanced by the presence of RanGAP and RanBPs. In contrast, in the nucleus, RanGEF promotes the switch from the Ran^{GDP} to Ran^{GTP} state. This results in an asymmetric distribution of Ran (the so-called "Ran^{GTP} gradient"), with the GTP- and the GDP-bound forms prevailing in the nucleus and cytosol, respectively (Kalab et al., 2002).

The Ran^{GTP} gradient controls the directional transport of cargos across the NPC (Görlich et al., 1996; Izaurralde et al., 1997). Both importins and exportins have high affinity only for Ran^{GTP} (Figure 1.5). The directionality of transport is ensured by the fact that the binding of importins to cargo and to Ran^{GTP} is mutually exclusive, whereas the binding of exportins to cargo and Ran^{GTP} is cooperative.



1.5: Directional karyopherin-mediated transport of cargos across the NPC. Schematic import (left) and export (right) pathways mediated by karyopherins.

More specifically, importins bind their cargo in the cytosol and move across the NPC as a binary complex (Rexach and Blobel, 1995). In the nucleus Ran^{GTP} competes with the cargo, causing its release, and the importin-Ran^{GTP} complex moves back to the cytosol

(Cook et al., 2007). Conversely, exportins bind their cargo in the nucleus together with Ran^{GTP}, forming a ternary complex that crosses the NPC (Kutay et al., 1997; Fornerod et al., 1997b). In the cytosol RanGAP and RanBP1 stimulate the GTPase activity of Ran, triggering its dissociation from the karyopherins (Bischoff and Görlich, 1997; Lounsbury and Macara, 1997). As a consequence importins can bind a new cargo, whereas exportins release their cargos and translocate back to the nucleus for a new cycle of export. Combined with the nucleocytoplasmic gradient of Ran^{GTP}, this difference in binding mode (mutually exclusive versus cooperative) ensures the correct movement of cargos across the NPC.

1.1.3 CRM1: A VERSATILE EXPORTIN

CRM1 (chromosome region maintenance 1, Exportin-1 or XPO1) is the major and most versatile nuclear export receptor in the cell, delivering a broad range of NES-containing cargos across the NPC (Fornerod et al., 1997a; Xu et al., 2012; Fu et al., 2013; Monecke et al., 2014; Fung et al., 2014). Like other exportins CRM1 binds to Ran^{GTP} and to the NES of the cargo in a cooperative manner: Ran^{GTP} increases the affinity of CRM1 for the cargo (and *vice versa*) by ~500-1000 fold (Paraskeva et al., 1999; Petosa et al., 2004). However, in contrast to other known exportins, which bind the cargo through the inner surface of the HEAT-repeat solenoid, CRM1 binds cargos via its outer surface (Dong et al., 2009). As there is no direct interaction between Ran^{GTP} and the cargo, the cooperative formation of the cargo-CRM1-Ran^{GTP} ternary complex is allosterically regulated by CRM1 structural elements. After the cargo-CRM1-Ran^{GTP} ternary complex is assembled in the nucleus, it translocates across the NPC through interactions with the Nups (Fornerod et al., 1997b; Bernad et al., 2004; Port et al., 2015) and dissociates in the cytosol.

CRM1 is a 124 kDa protein composed of 21 HEAT repeats (abbreviated H1-21) which form a ring-shaped solenoid (Figure 1.6) (Monecke et al., 2014; Fung et al., 2014). Several structural features play an important role in CRM1 function: the overall conformation of the HEAT-repeat solenoid; helix B of H21 which corresponds to the C-terminal helix; a large, acidic loop that connects helices A and B of H9; and a hydrophobic groove located on the outer surface between H11 and H12 where the NES binds.

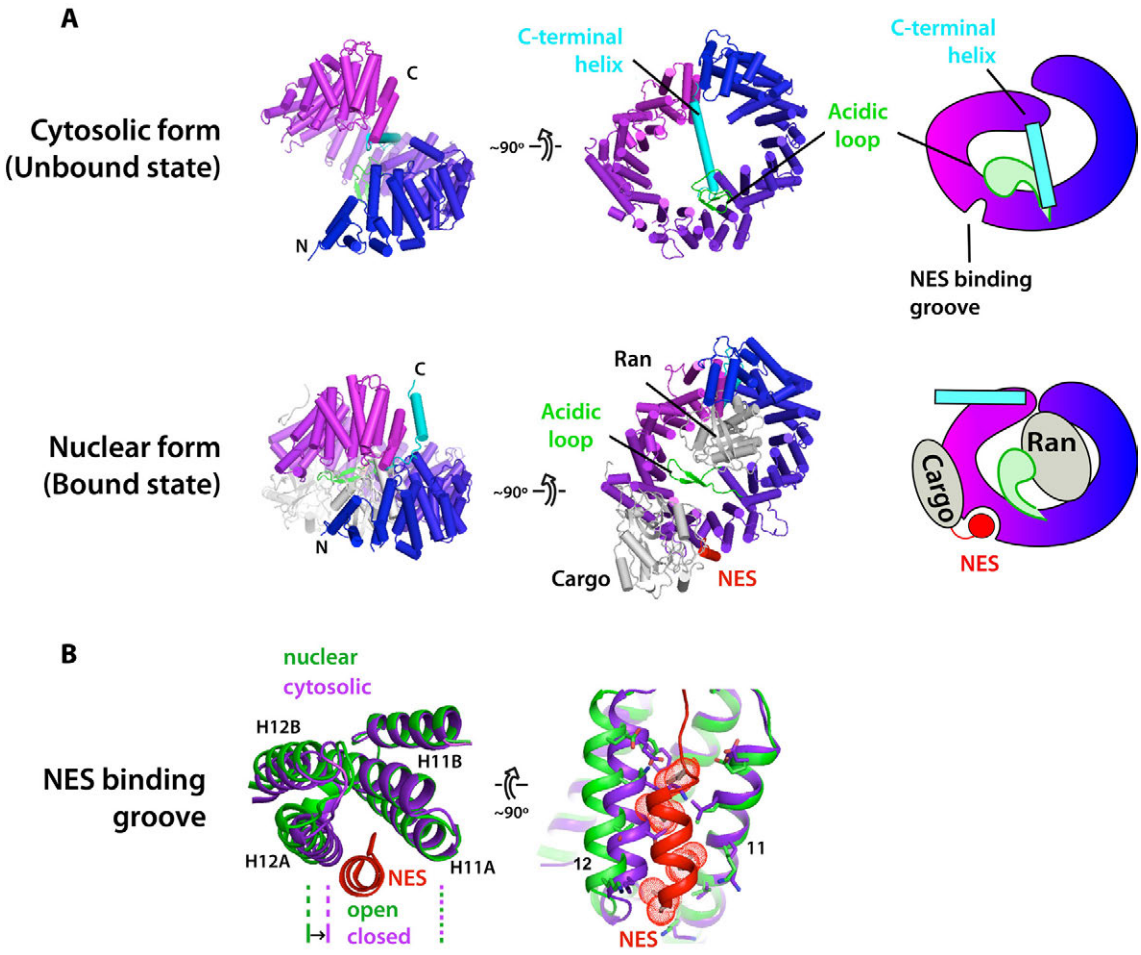


Figure 1.6: CRM1 structures and conformational changes. A) Cytosolic and nuclear conformations of CRM1. The HEAT repeats are colored from blue (N-terminus) to magenta (C-terminus). The C-terminal helix and acidic loop are in cyan and green, respectively. Ran^{GTP} and the cargo are colored in grey. In its cytosolic form (PDB 4FGV) the CRM1 HEAT repeats adopt a more extended conformation with the C-terminal helix in *cis* and a retracted acidic loop. In the nuclear state (PDB 3GJX) the HEAT repeats are more compact, the C-helix is in *trans* and the acidic loop in an elongated conformation. B) Comparison of the NES binding groove in the nuclear (magenta) and cytosolic (green) conformations. The NES peptide of the cargo is in red. In the closed (cytosolic) conformation the NES residues would clash with the A helix of HEAT repeat 12 (H12A).

The Protein Data Bank (PDB) currently contains ~40 structures of CRM1 in the unbound state and bound to various partner proteins, revealing a conserved architecture across different orthologs (human, mouse, fungi). These structures provide a series of snapshots during the export cycle: unbound CRM1, CRM1-cargo, CRM1-Ran-cargo, CRM1-Ran-cargo-Nup, CRM1-Ran-RanBP1, CRM1-inhibitor. The structures show that CRM1 adopts different conformations in the cytosolic (unbound) and nuclear (bound)

states (Figure 1.6 A). In the unbound state, CRM1 adopts an extended conformation with the N- and C-terminal HEAT repeats apart from each other (Saito et al., 2013; Monecke et al., 2013; Dian et al., 2013). The C-terminal helix adopts the so-called *trans* conformation, crossing the ring to contact the B helices of H9 to H12. The acidic loop adopts a retracted conformation which interacts with the B helices of H11 and H12. The solenoid, C-terminal helix and acidic loop conformations stabilize the NES-binding groove in a narrow conformation that is sterically incompatible with the binding of an NES peptide (Figure 1.6 B). In the Ran^{GTP}- and cargo-bound state (Figure 1.6 A), the CRM1 solenoid is more compact, with the N- and C-terminal HEAT repeats in close proximity (Monecke et al., 2009). The C-terminal helix is positioned on the outer surface of the CRM1 ring (*cis* conformation). The acidic loop adopts an elongated conformation which interacts with Ran^{GTP} and with residues in the B helices of H12 to H15. These altered conformations of the solenoid, acidic loop and C-terminal helix stabilize a wider conformation of the NES-binding groove, able to accommodate the NES peptide.

The key structures that have helped to elucidate the function of CRM1 during the export cycle are schematically illustrated in Figure 1.7. Upon first entering the nucleus CRM1 (Figure 1.7, state A), the retracted conformation of the acidic loop stabilizes the NES-binding groove in the narrow conformation. CRM1 can then (weakly) interact either with Ran^{GTP} or with a cargo (states A' and A"). Both intermediates have been crystallized showing how CRM1 features re-arrange upon binding of one of the two partners. In the CRM1-cargo complex the solenoid is more compact than in the unbound state, the acidic loop is unstructured and the C-terminal helix is in the *trans* conformation (Dong, 2009). This conformation has an enhanced affinity for Ran^{GTP} because the unstructured loop can more easily capture Ran^{GTP}. In the CRM1-Ran^{GTP} structure the solenoid is even more compact than in the CRM1-cargo complex (Güttler et al., 2010). Ran^{GTP} binds to CRM1 on the N-terminal HEAT repeats (with interactions spanning from H1 to H9) as well as with the C-terminal HEAT repeats (H17 to H19). The acidic loop interacts with Ran^{GTP}, locking it at its interaction site, whereas the C-terminal helix adopts a *cis* conformation. All these conformational changes cause the widening of the NES groove, facilitating NES binding.

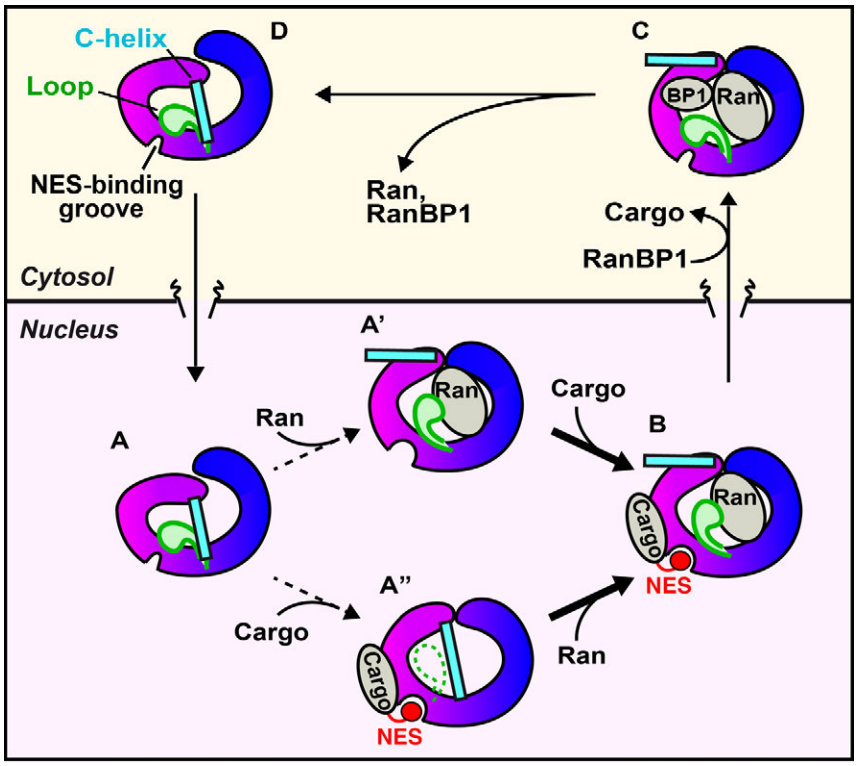


Figure 1.7: Overview of conformational changes in CRM1 during the export cycle. The cartoons show the key structural features of CRM1: the ring shape (blue to magenta), the C-terminal helix (cyan), the acidic loop (green) and the NES-binding groove. Ran^{GTP} and the cargo are colored in grey, the NES peptide in red.

Once the ternary complex is assembled in the nucleus (state B) it crosses the NPC, facilitated by interactions with FG-Nups (Port et al., 2015). In the cytosol the disassembly of the complex is accelerated by RanBPs that, in cooperation with RanGAP, stimulate the GTPase activity of Ran. The structure of the CRM1-Ran^{GTP}-RanBP1 complex shows how RanBP1 destabilizes the interactions between Ran^{GTP} and the acidic loop of CRM1, inducing the latter to adopt a retracted conformation (state C in Figure 1.7) (Koyama and Matsuura, 2010). As a consequence the NES groove is closed and the cargo is released. GTP hydrolysis induces Ran to dissociate from CRM1, freeing the exportin to re-enter the nucleus and start a new export cycle (state D in Figure 1.7). Two aspects make CRM1 a highly versatile exportin. First, the position of the NES groove on the outer, rather than the inner, surface means that CRM1 can sterically accommodate cargos having a wide diversity of 3D structures. Second, the modular nature of the NES motif means that any protein bearing this signal can be targeted for CRM1-mediated export.

The NES motif was first identified in the HIV-1 regulatory protein Rev (Fisher et al., 1995) and in protein kinase inhibitor (PKI) (Wen et al., 1995). Since then the analysis of many different NES peptides has led to the definition of a general NES sequence: a pattern of 5 hydrophobic residues (Φ), numbered from 0 to 4, with a general consensus of Φ^0-X-Φ^1-X_3-Φ^2-X_{2-3}-Φ^3-X-Φ^4 where X is any residue (Güttler et al., 2010). Very recently this consensus has been extended with the identification of NESs that binds in the opposite polypeptide direction (Fung et al., 2015). Atomic resolution structures of different NES peptides bound to CRM1, show that the Φ residues dock into five hydrophobic pockets within the NES-binding groove (Figure 1.8). Depending on the non-hydrophobic spacer residues the NESs of different cargos adopt slightly distinct conformations to adapt to the NES-binding groove (Dong et al., 2009; Güttler et al., 2010). For example the Rev NES, compared to PKI NES, compensates for having fewer spacer residues by adopting a more stretched structure (Figure 1.8).

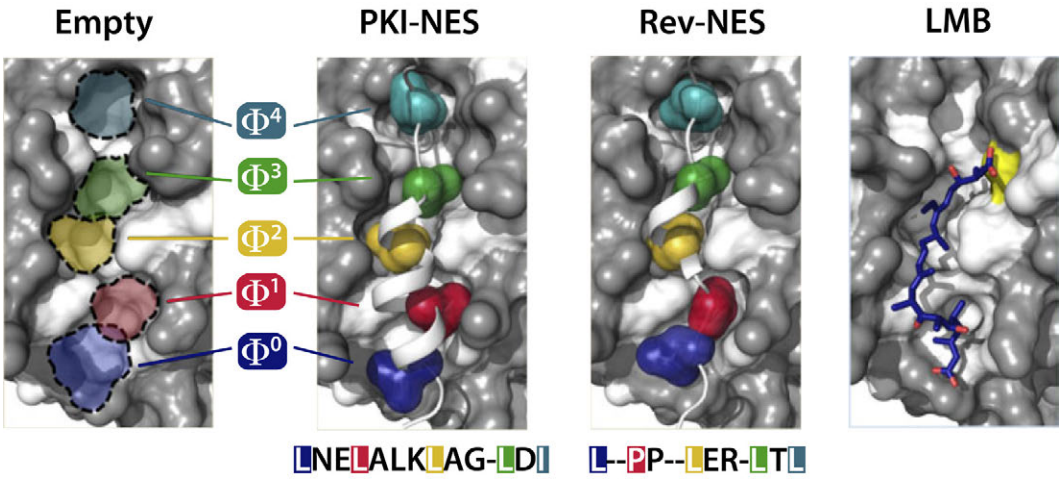


Figure 1.8: NES-binding groove of CRM1. A surface representation of the NES-binding groove is shown with hydrophobic residues colored in white and charged residues in grey. The NES peptide is omitted from the leftmost panel for clarity. The three other panels show the groove occupied by a PKI-NES or Rev-NES peptide or by leptomycin B (LMB) (PDB codes 3NC1, 3NBZ and 4AHT, respectively). Φ residues of the NES consensus and the corresponding hydrophobic pockets occupied in the groove are color coded (Φ^0 in blue; Φ^1 in red; Φ^2 in yellow; Φ^3 in green and Φ^4 in teal) (*adapted from Monecke et al., 2014*)

Finally, CRM1 export activity is inhibited by leptomycin B (LMB), initially discovered as an antifungal antibiotic, and by related compounds. LMB acts as a suicide inhibitor by forming a covalent adduct to a cysteine residue (Cys528 in human CRM1) located within

the NES groove (Kudo et al., 1999), thereby competing with cargo binding by sterically hindering NES recognition.

1.1.4 RNP EXPORT MEDIATED BY CRM1

Besides its activity in protein export, CRM1 is also essential for the export of certain ribonucleoproteins (RNPs). These include spliceosomal small nuclear RNPs (snRNPs), a subset of mRNPs, 40S and 60S ribosomal subunits, and some viral RNPs (Okamura et al., 2015) (Figure 1.9). As CRM1 does not directly bind to RNA, various adaptor proteins are required for export. These adaptors contain a NES that recruits CRM1 and can interact with the target RNA by recognizing specific RNA motifs. An important adaptor protein in the context of this thesis is the HIV-1 protein Rev, which tethers CRM1 to viral RNAs by recognizing the Rev Responsive Element (RRE) (described in detail in section 1.2). Examples of cellular adaptor proteins are given below.

The snRNPs contain spliceosomal snRNAs (U1, U2, U4 and U5) of ~ 100 nt long that are required for the assembly and functionality of the spliceosome in the nucleus. After a first step of maturation these snRNPs are exported to the cytosol before returning to the nucleus and assembling into the spliceosome. Their export requires the adaptor protein PHAX, which only upon phosphorylation is able to recruit CRM1 (Figure 1.9).

The bulk of cellular mRNPs are exported by the Tap/p15 heterodimer. This hetero-complex was discovered as the transport factor required for the export of intron-containing transcripts of the simple retrovirus Mason-Pfizer monkey virus and further identified to mediate bulk mRNA export (Segref et al., 1997; Gruter et al., 1998). CRM1 is involved in the export of specific mRNAs whose expression is highly regulated, such as certain mRNAs coding for oncogenes (e.g., eIF4E) and cytokines (e.g., Tumor Necrosis Factor) (Cullen, 2003; Natalizio and Wente, 2013). These RNAs contain specific sequence elements such as the AU rich element (ARE) or the 4E-SE RNA element. So far three adaptor proteins that recruit CRM1 on these mRNAs have been identified: the human antigen R (HuR), which recognizes specific AREs in the mRNA and recruits CRM1 together with other two proteins (pp32 and APRIL); the leucine-rich pentatricopeptide repeat protein (LRPPRC), which recognizes the 4E-SE RNA element; and Nxf3, a paralog of TAP which lacks the Nup interaction domain and which bridges CRM1 to currently unknown RNA elements.

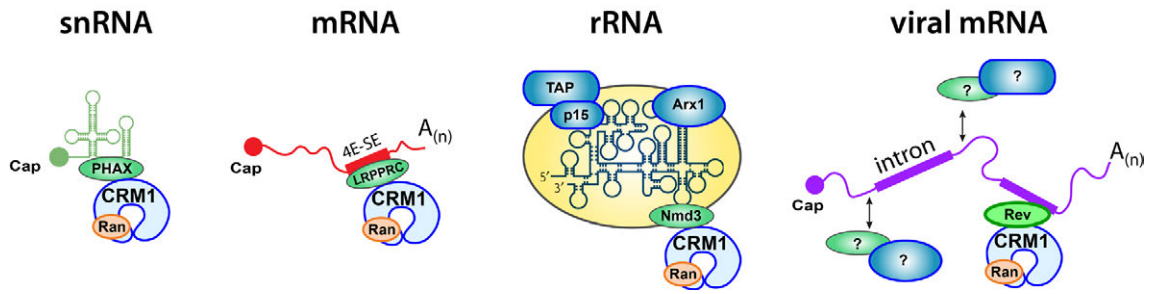


Figure 1.9: RNA export mediated by CRM1. Different classes of RNA that require CRM1 for nuclear export are snRNAs, mRNAs, rRNAs and viral mRNAs. CRM1 and other export factors are colored in blue and adaptor proteins in green. As the RNA size increases (from left to right) multiple export factors are recruited onto the RNA or RNP cargo.

Another important observation is that as the RNA size of a given RNP increases, the cooperation of multiple export factors is required to translocate such large complexes across the NPC. For example the efficient export of the 60S ribosomal subunit, which contains three RNA molecules of 120, 4700 and 160 nt associated with 49 proteins, requires at least three different export factors, including CRM1 (Kohler and Hurt, 2007).

We will focus in the next section on the export of HIV-1 viral transcripts. Many complex retroviruses export intron-containing viral transcripts to the cytosol and some of these exploit the CRM1 pathway. It is likely that such large RNPs might require multiple transport factors to translocate across the NPC.

1.2 HIV EXPLOITS CRM1 TO EXPORT INTRON-CONTAINING VIRAL RNAS

1.2.1 THE EXPORT OF INTRON-CONTAINING VIRAL RNAS OF HIV-1 IS REGULATED BY REV

Human immunodeficiency virus type 1 (HIV-1) is the prototype of complex retroviruses that encode not only for the conventional retroviral Gag, Pol and Env proteins but also for auxiliary and regulatory proteins such as Tat, Nef and Rev (Peterlin and Trono, 2013). An overview of the viral cycle is shown in Figure 1.10.

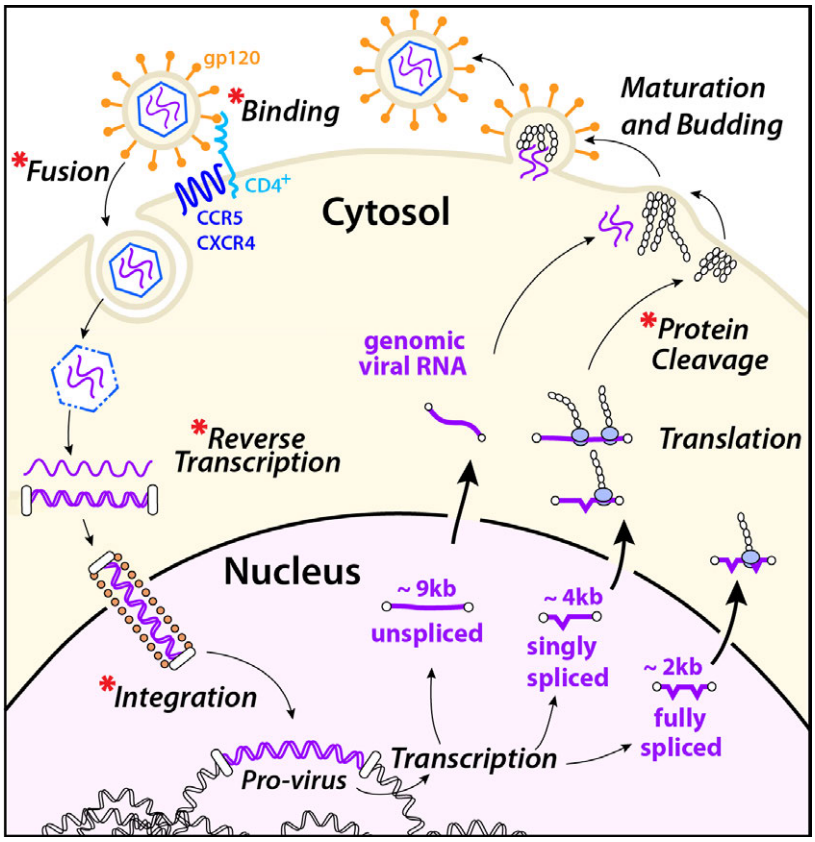


Figure 1.10: HIV-1 infection cycle. Schematic representation of the key steps of the HIV-1 replication cycle: binding, fusion, reverse transcription, integration, transcription, translation, protein cleavage, budding and maturation. Steps that are targeted by antiretroviral drugs are highlighted with a red asterisk.

The infectious virion binds to the host cell through the glycoprotein gp120, which interacts with the CD4 receptor and co-receptor (CXCR4 or CCR5) present on the host cell surface. Fusion of the viral and cellular membrane results in release of the viral proteins and capsid into the cytosol. The uncoating of the capsid facilitates reverse transcription of

the single strand RNA viral genome into double strand DNA by the viral enzyme reverse transcriptase. Linear DNA is then imported into the nucleus as a nucleic acid-protein complex (pre-integration complex, PIC) and incorporated into host chromatin as a proviral DNA genome by the viral enzyme integrase. The production of new viral particles requires the proviral DNA to be transcribed by the host machinery. Three different classes of viral transcripts (~9 kb 4 kb and ~2 kb) are produced, undergo maturation and are exported to the cytosol. Here they are translated into all the viral proteins and further processed by the viral enzyme protease for the correct assembly of new viral particles.

Four main steps of the viral cycle are target of drugs that are currently used to treat the HIV-1 infection (Figure 1.10). The currently approved therapy uses a combination of these drugs and is defined Highly Active Anti-Retroviral Therapy (HAART) (Arts and Hazuda, 2012).

Entry inhibitors. One class of inhibitors are peptides homolog to the leucine zipper domain of the viral protein gp41. The peptides interact with this viral protein interfering with the change of conformation required during the process of viral entry (Kilby et al., 1998). Another class of inhibitors are antagonists of the CCR5 chemokine receptor. These molecules bind the trans-membrane domain of the CCR5, inducing and stabilizing a conformation of the receptor that is not recognized anymore by the viral envelope (Tsamis et al., 2003).

Reverse transcriptase inhibitors. Reverse transcriptase was the first HIV-1 enzyme to be exploited for antiretroviral drug discovery. This enzyme is the target for two classes of antiretroviral agents: the nucleoside reverse transcriptase inhibitors (NRTIs) and the non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs are nucleoside analogs that lack the 3'-hydroxyl group at the sugar moiety, preventing the formation of a 3'-5'-phosphodiested bond with the incoming 5'-nucleoside triphosphates, resulting in termination of the growing viral DNA chain (Hart et al., 1992). NNRTIs bind on the RT inducing a change of conformation that affects the substrate binding site, resulting in a reduced polymerase activity (Spence et al., 1995).

Integrase inhibitors. The integrase inhibitors (INIs) block integration of the HIV-1 DNA into the host cell genome, by specifically inhibiting the strand transfer step during which the viral and host DNA are fused. This class of drugs interact with the two essential

magnesium ion cofactors present in the active site of the integrase as well as with the viral DNA (Sato et al., 2006).

Protease inhibitors. These agents block proteolysis of the viral polyproteins that is required for the production of infectious viral particles. They are large peptide-like compounds and require the co-administration of a boosting agent to inhibit their metabolism and enhance drug levels (Hsu et al., 1998).

The use of HAART greatly reduced morbidity and mortality related to HIV-1 infection and AIDS. However the virus continues to evolve and escape and therefore new treatments are needed. Nearly all the viral processes that are distinct from the cellular life cycle are pontentially suitable for screening and designing inhibitors, such as for instance the export of the non-spliced viral RNAs.

HIV-1 encodes up to 15 different proteins starting from a very small genome of ~10 kb (Frankel and Young, 1998) (Figure 1.11 A). During the replication all the HIV-1 viral proteins are produced thanks to transcriptional and translational mechanisms such as the presence of overlapping open reading frames (ORFs), the use of alternative start codons, the production of polyproteins that are subsequently cleaved by the viral protease, and the presence of alternative splicing sites (Figure 1.11 and https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html). The genome contains two introns and a key regulatory sequence called the Rev response element, located within the second of these (Malim et al., 1989; Kjems et al., 1991).

The proviral DNA is transcribed by the host machinery as a ~9 kb unspliced transcript, corresponding to the entire viral genome, further processed to generate the other two classes of viral RNAs: fully spliced transcripts of ~2 kb and singly spliced transcripts of ~4 kb that retain the second intron (Figure 1.11 B). Fully spliced transcripts contain early expressed genes, encoding for regulatory and accessory viral proteins, notably the protein Rev (Cullen and Malim, 1991). Singly spliced and un-spliced transcripts contain late expressed genes, encoding for structural and enzymatic viral proteins.

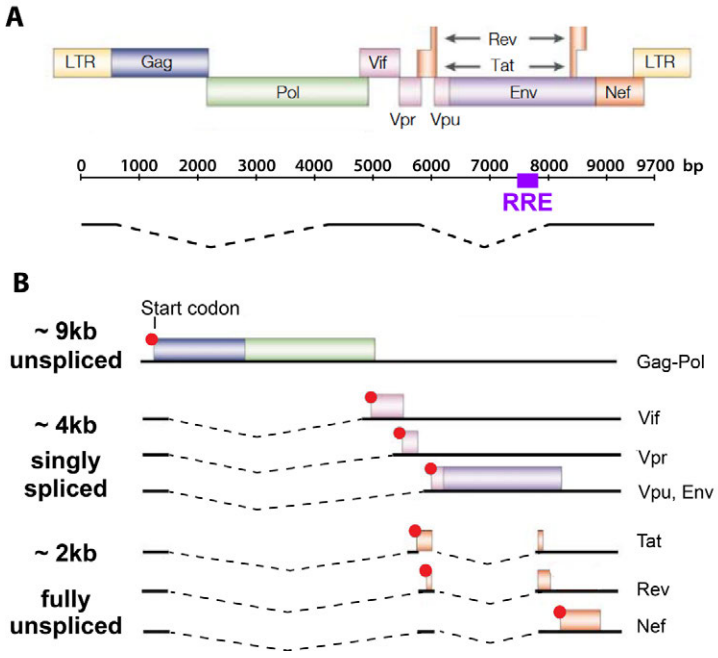


Figure 1.11: HIV-1 genome and different classes of viral transcripts. A) The HIV-1 genome contains regulatory untranslated regions at the 5' and 3' ends (LTRs) and 9 different genes (colored boxes) for a total of 15 open reading frames. The genome length in base pairs (bp) is indicated as a continuous black line with the position of the RRE highlighted in magenta. The scheme underneath indicates the position of the two introns, depicted as dotted lines. B) Three classes of viral transcript are produced in the host cell. Each transcript is depicted as a continuous black line with the encoded protein boxed (listed at right), with the start codon of the open reading frame indicated as a red dot (*adapted from Peterlin and Trono, 2013*).

The three classes of transcripts are exported through different pathways and expressed at different stages of the viral cycle (Figure 1.12). The transition between early and late gene expression is regulated by Rev which allows the export of the ~9 kb and ~4 kb viral RNAs (Figure 1.12). Initially only the ~2 kb transcripts are exported through the host cellular Tap/p15 pathway and encodes for the early expressed protein Rev. Rev is then imported to the nucleus where it binds to the RRE, present on the ~4 kb and ~ 9kb viral transcripts. Rev recruits CRM1 and its cofactor Ran^{GTP}, forming a RNPs competent for export.

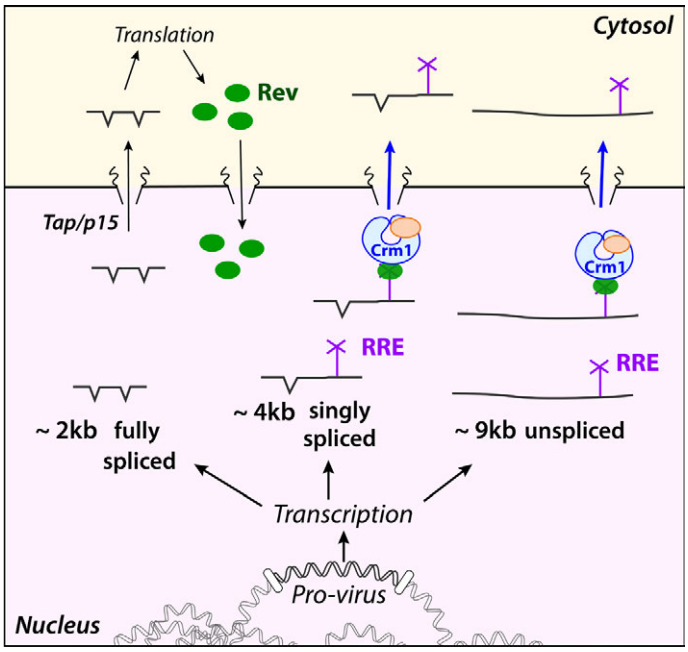


Figure 1.12: Rev mediates the export of intron-containing viral RNAs. The proviral DNA is transcribed into three classes of viral RNAs. Fully spliced viral mRNAs (~2 kb) are delivered to the cytosol by the Tap/p15 export pathway, where they encode for regulatory and accessory proteins, notably Rev (in green). Rev is imported into the nucleus where it binds the RRE (magenta cross), present on singly spliced (~4 kb) and unspliced (~9 kb) viral RNA transcripts. Rev recruits CRM1 (in blue) and Ran^{GTP} (in orange), which export the Rev-bound viral transcripts to the cytosol.

1.2.1.1 Rev oligomerization on the RRE is critical for export

Shortly after the discovery of Rev (Sodroski et al., 1986) and the identification of the RRE (Malim et al., 1989) it was shown that more than one Rev molecule is recruited on the RRE (Heaphy et al., 1991; Malim and Cullen, 1991; Kjems et al., 1991). The RRE is a highly structured RNA element of ~350-nt containing multiple binding sites for Rev (Figure 1.13). The initial binding event is sequence specific and occurs in a region of the RRE called stem-loop IIB (IIB), which is the highest affinity binding site for Rev on the RRE (Bartel et al., 1991; Tiley et al., 1992; Iwai et al., 1992). After the first Rev molecule binds, the protein oligomerizes cooperatively along the RRE (Mann et al., 1994; Kjems et al., 1991; Daly et al., 1993; Zemmel et al., 1996). Indeed, the *in vitro* binding affinity of Rev for the RRE is 500-fold higher than that estimated for the isolated IIB (Daugherty et al., 2008). Significantly, an early study showed that the rate of export of intron-containing viral RNAs *in vivo* is proportional to the number of Rev molecules bound to the RRE (Mann et al, 1994). In this

study a series of RRE truncation fragments were used to compare Rev oligomerization *in vitro* and its activity *in vivo*. Each truncation reduced both the Rev-RRE complex stoichiometry formed *in vitro* and the rate of export *in vivo*, with a minimal RRE length of ~270 nt defined to be essential to fully retain the Rev response *in vivo*.

These findings led to the hypothesis that the RRE acts as a molecular rheostat that senses the concentration of Rev during viral infection (Mann et al., 1994): when Rev concentration is too low (early gene expression phase) binding occurs only on IIB and the Rev-RRE complex is not competent for being exported; as the concentration of Rev increases, cooperative oligomerization on RRE occurs, allowing for its nuclear export and increasing the strength of the *in vivo* response (late gene expression phase) (Figure 1.13). The ability of Rev to oligomerize on the RRE is thus crucial for the efficient export of intron-containing HIV-1 transcripts. Indeed Rev mutants defective in oligomerization severely compromise RRE export (Thomas et al., 1998; Edgcomb et al., 2008; Daugherty et al., 2008).

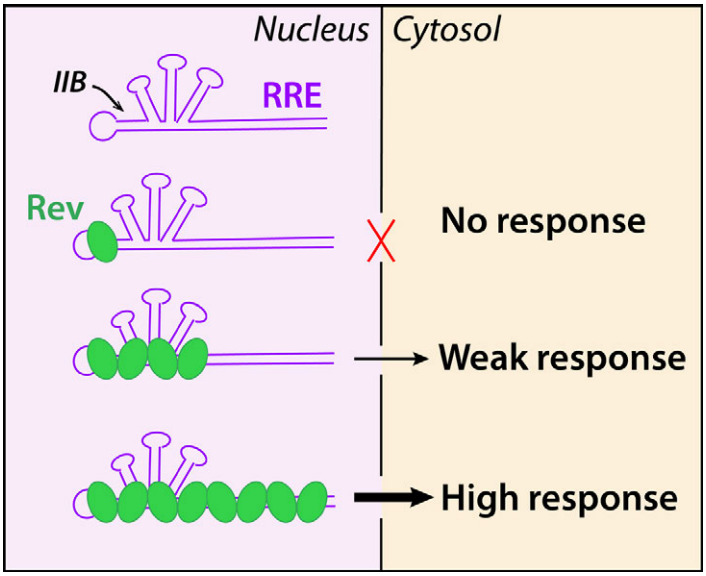


Figure 1.13: Schematic representation of the Rev response *in vivo*. At low Rev concentration, Rev binds the RRE only on IIB and is insufficient to trigger RRE export. As the Rev concentration increases cooperative binding occurs along the RRE with an *in vivo* response that is proportional to the number of Rev molecules bound (*adapted from Mann et al, 1994*).

Why Rev oligomerization is critical for efficient export is unclear. One possibility is that efficient export requires a tight and specific RRE-Rev interaction that is achieved through

cooperative binding. An alternative explanation is that multiple copies of CRM1 are needed for efficient export, recruited through the Rev oligomer. These two scenarios are not mutually exclusive.

How many Rev molecules bound to the RRE are required for efficient export and how many CRM1 molecules are needed to export such large RNPs is still unknown. Over the years different *in vitro* techniques have been employed to determine the stoichiometry of the RRE-Rev complex. These have identified the number of Rev molecules per RRE to vary from 4 to 13 (Table 1.1).

Length of RRE (nt)	Stoichiometry (Rev:RRE)	Nucleotides per Rev monomer	Method	Reference
242	6:1	40.3	SEC	Daugherty et al., 2010a
244	8:1	30.5	NFBA EMSA	Holland et al., 1990 Wingfield et al., 1991
244	4:1	61.0	SPR	Van Ryk et al., 1999
244	5:1	48.8	NFBA	Jeong et al., 2000
280	8:1	35.0	NFBA EMSA	Cook et al., 1991 Daly et al., 1993
351	8:1	43.9	EMSA	Mann et al., 1994
351	13:1	27.0	EMSA	Brice et al., 1999
370	13:1	28.5	AFM	Pallesen et al., 2009

Table 1.1 Estimates of Rev/RRE stoichiometry reported by different groups. NFBA: nitrocellulose filter binding assay; EMSA: electrophoretic mobility shift assay; SPR: surface plasmon resonance; AFM: atomic force microscopy; SEC: size exclusion chromatography (*adapted from Vercruysse and Daelemans, 2013*).

Like many RNA-binding proteins, Rev can interact with RNA non-specifically, which may explain why the different studies have led to a broad range of proposed stoichiometry. In addition, the tendency of Rev to oligomerize in solution, even in the absence of RNA, may compromise the formation of specific RRE-Rev complexes *in vitro*. Consequently, the number of Rev molecules that directly and specifically interact with the RRE *in vitro* is still uncertain. Unfortunately even less is known about the RRE-Rev stoichiometry *in vivo* due to the lack of assays to monitor and quantify Rev oligomerization in cells.

1.2.2 At least two copies of CRM1 mediate the export of intron-containing HIV-1 RNPs

The affinity of the Rev NES for CRM1 is weaker compared to that of other cellular cargos (Pareskeva et al., 1999), yet the virus recruits the host export machinery with high efficiency. Several studies suggest that the Rev-RRE complex recruits at least two molecules of CRM1. First, *in vivo* studies showed that at least two Rev NESs are required for the efficient export of viral RNPs, suggesting the possible recruitment of more than one copy of CRM1 (Hoffmann et al., 2012; Aligeti et al., 2014). Second, a recent structural study proposed a low resolution electron microscopy (EM) model of an *in vitro* reconstituted RRE-Rev-CRM1-Ran^{GTP} particle in which two molecules of CRM1 are clearly bound to the Rev-RRE complex (Booth et al., 2014). The EM reconstruction shows an hourglass-like shape in which a CRM1/Ran^{GTP} dimer clearly occupies most of the map density. The RRE-Rev complex is proposed to be located in the extra density that is positioned asymmetrically on one surface of the CRM1 dimer (Figure 1.14 A). This study identified a previously uncharacterized dimerization interface specific of human CRM1. This interface is formed by a patch of residues forming a continuous surface spanning from H5 to H10 (Figure 1.14 B, C). Third, in support of a CRM1 dimer, *in vivo* studies identified the residues present in the dimer interface to be species-specific determinants required for the efficient export of HIV-1 intron-containing transcripts (Sherer et al., 2011; Elinav et al., 2012). These residues are specific to higher primates and their substitution into murine CRM1 allows the latter to export RRE-containing transcripts as efficiently as human CRM1.

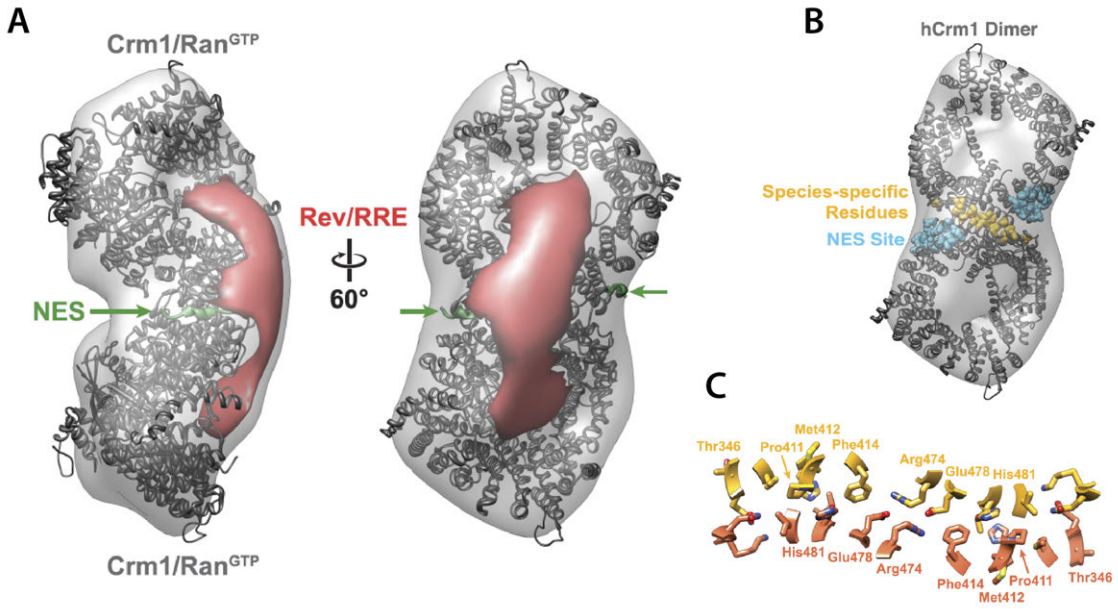


Figure 1.14: EM model of a RRE-Rev-CRM1-Ran^{GTP} complex. A) EM map of the RRE-Rev-CRM1-Ran^{GTP} complex is shown as a semi-transparent surface. The fitted CRM1-Ran^{GTP} dimer crystal structure is shown as a grey ribbon diagram. The extra density highlighted in red is attributed to the Rev-RRE complex. B) The dimer interface of CRM1 and the NES binding grooves (called NES site) are indicated in yellow and cyan respectively. B) The residues of CRM1 forming the dimer interface are shown as isolated sticks. Yellow and orange distinguish the residues on two different CRM1 molecules. The interface is stabilised by both salt bridges and hydrophobic interactions (*From Booth et al., 2014*).

In conclusion, Rev oligomerization on the RRE might be a way to increase the local concentration of NES and therefore facilitate the recruitment of CRM1, whose affinity for the RRE-Rev RNP is further increased by dimer formation. All these recent findings confirm the possibility that more than one CRM1 molecule may be needed for the efficient export of HIV-1 viral RNPs. Therefore, the uncertainty about stoichiometry is not restricted to the RRE-Rev complex but extends to the entire RRE-Rev-CRM1-Ran^{GTP} complex.

1.2.2 Rev: a small protein with multiple functional domains

Rev is a ~13 kDa protein of 116 residues comprising three functional domains (Hammarskjold and Rekosh, 2011; Rausch and Grice, 2015). The N-terminal region contains two functional domains: the nuclear localization signal and RNA binding domain

(NLS/RBD), also known as the arginine rich motif (ARM), and the oligomerization domain (OD) (Figure 1.15 A). The ARM spans residues 34 to 50 and contains several arginine residues involved in RNA binding. The OD consists of two regions (OD_1 and OD_2) which flank the ARM and contain hydrophobic residues involved in Rev-Rev interactions. The integrity of these two domains allows Rev to specifically recognize IIB and to oligomerize along the RRE. The C-terminal region contains the NES (residues 74-83) through which Rev binds to CRM1 (Fisher et al., 1995).

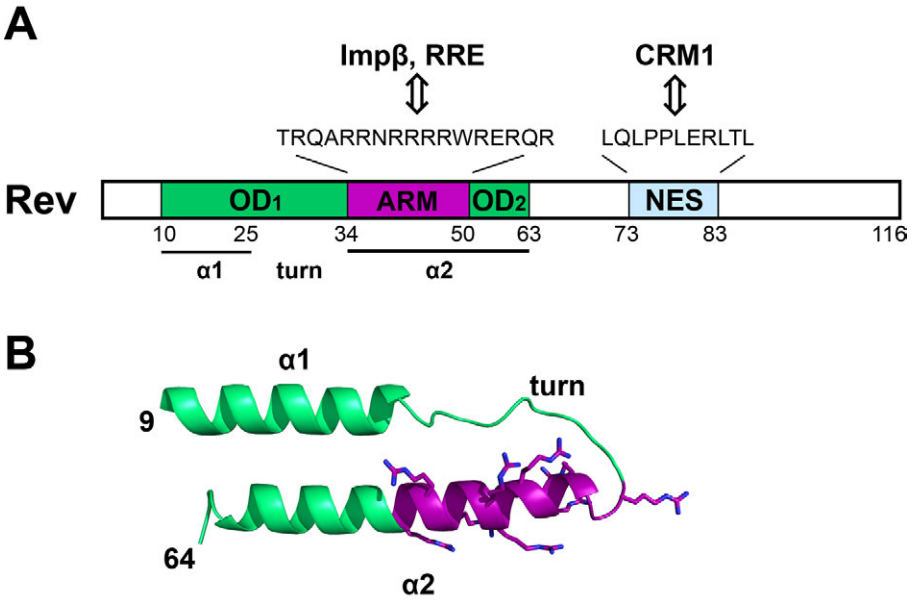


Figure 1.15: Rev functional domains and crystal structure. A) Rev primary structure is represented as a rectangular box in which the OD is colored green, the ARM in magenta and the NES motif in light blue. Residue positions delimiting functional and structural domains are specified below and the extended ARM and NES peptides sequences are indicated above. B) The N-terminal domain of Rev (residues 9-64) adopts a helix-turn-helix structure (PDB ID 2X7L). Domains are colored as in A) with arginine residues in the ARM shown as sticks.

The tendency of Rev to oligomerize *in vitro* in combination with its highly basic domain had hindered structural and functional studies for a long time due to low protein solubility and/or the formation of filaments and insoluble aggregates (Wingfield et al., 1991; Cole et al., 1993). The crystal structure of the N-terminal domain of Rev was solved independently by two groups, revealing that residues 9 to 65 fold as an antiparallel helix-turn-helix (α1-turn-α2) domain (Figure 1.15 B) (DiMattia et al., 2010; Daugherty et al., 2010b). The two helices adopt a nearly parallel orientation, stabilized by both hydrophobic and electrostatic

interactions. The OD_1 region forms the α1-helix and turn whereas the ARM domain and OD_2 region are continuously arranged to form the α2-helix, such that the entire OD defines a flat two-sided structure. Even if DiMattia and coworkers (2010) crystallized full-length Rev, only the N-terminal domain structure was visible in the electron density map, indicating that the C-terminal region is intrinsically disordered.

1.2.2.1 Rev oligomerization interfaces

Long before the Rev crystal structure was solved, an elegant mutagenesis study had located the OD of Rev on two different surfaces of the protein, designated as the "head" (H) and "tail" (T) surfaces (Jain and Belasco, 2001). This study demonstrated that Rev oligomerization had a polarity, with each OD surface able to interact only with its counterpart on a second Rev molecule. Accordingly, a model was proposed whereby Rev oligomerization involved only the formation of symmetric tail-to-tail (T/T) or head-to-head (H/H) interactions. Moreover, Jain and Belasco (2001) showed that the T interface was involved in the dimerization of Rev, whereas the H interface mediated the formation of higher-order oligomers. This model was confirmed by the two crystal structures published in 2010, in which two distinct Rev dimers were observed by association through the H/H interface (DiMattia et al., 2010) or through the T/T interface (Daugherty et al., 2010b), with each interface formed by two mutually exclusive sets of hydrophobic residues (Figure 1.16 A, B). In addition, a recent structural study identified a third group of residues, located within the interhelical turn, as forming a third oligomerization interface, designated the C/C interface (DiMattia et al., 2016) (Figure 1.16 C).

In all these structures the two Rev monomers adopt a 'V' or 'X' shape. The ARM motifs are far apart from each other and point in opposite directions in the H/H and T/T dimers, whereas they are in close proximity in the C/C dimer. The precise crossing angle between Rev monomers (and hence the overall conformation of the dimer) can vary significantly because the hydrophobic interface can tolerate important variations. This conformational flexibility is probably crucial for the efficient oligomerization of Rev on the RRE (discussed further in section 1.2.5).

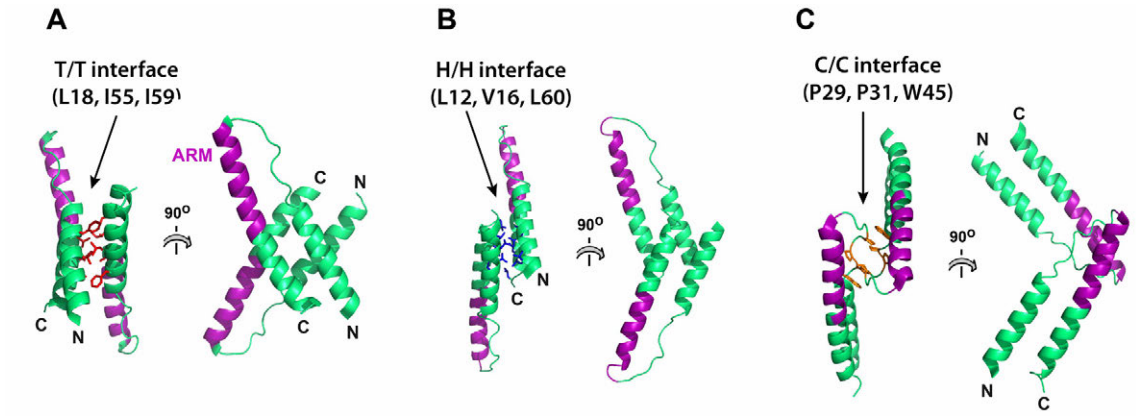


Figure 1.16: Oligomerization interfaces of Rev. Three different Rev crystal structures are shown. Residues involved in dimer formation are shown as sticks, some of which are indicated above. A) T/T interface formed by residues L18, F21, L22, I55, I59 (PDB ID 3LPH). B) H/H interface formed by residues L12, L13, V16, I19, L60 and L64 (PDB ID 2X7L). C) C-C interface formed by residues P28, P29, P31 and W45 (PDB ID 5DHV). Note how the two Rev ARM (in magenta) of each dimer adopt different relative orientations depending on the Rev-Rev interface involved.

1.2.3 Structural features of the Rev Response Element

The RRE is a ~350-nt long sequence located in the second HIV-1 intron. RNA folding prediction software combined with enzymatic and chemical probing *in vitro* and *in vivo* have revealed the RRE to be highly structured, comprising 5-stem loops, numbered from I to V, which protrude from a central junction in a 4-way or 5-way fashion (Mann et al., 1994; Charpentier et al., 1997). A general secondary structure of the RRE, with some variations, is shown in Figure 1.17. Stem I is the longest with double strand segments interrupted by internal loops or bulges of variable size, some of which are rich in purines. Stem II is attached to the central junction through stem IIA (IIA) which bifurcates into two substructures called stem IIB (IIB) and stem IIC (IIC). IIA, IIB and IIC arrange around a 3-way junction called junction ABC. Stems III, IV and V (III, IV and V) are the most variable part of the RRE and have been reported to adopt different conformations: all separated to form a 5-way structure (Kjems et al., 1991; Sherpa et al., 2015); forming a hybrid III/IV separated from V in a 4-way structure (Mann et al., 1994; Charpentier et al., 1997; Legiewicz et al., 2008); or forming a hybrid IV/V separated from III to yield an alternative 4-way structure (Bai et al., 2014).

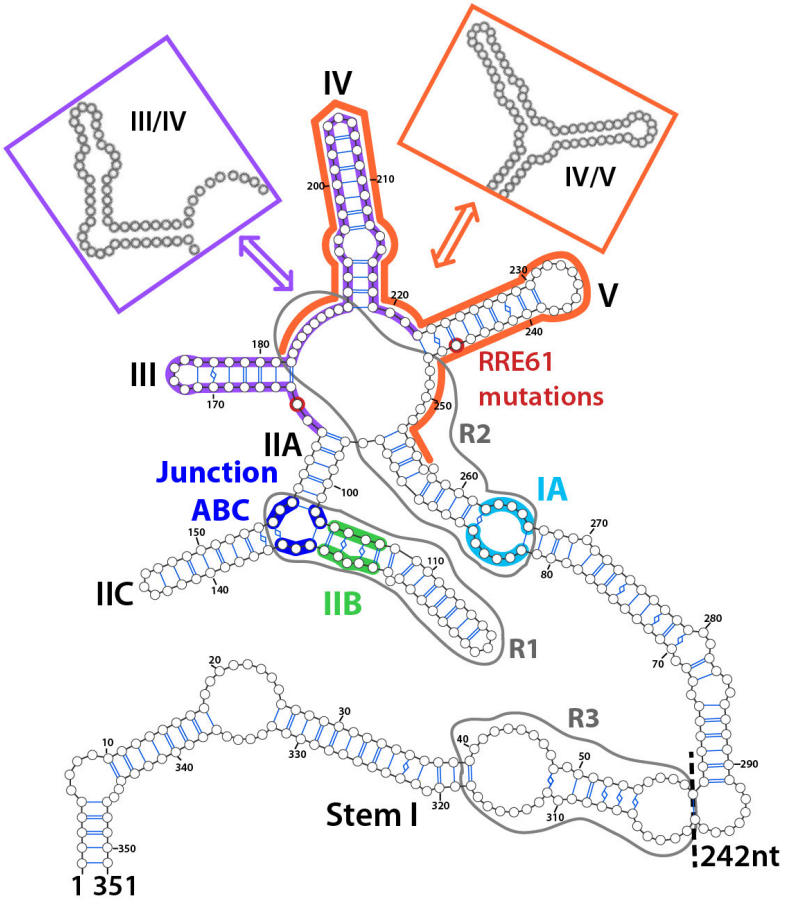


Figure 1.17: RRE secondary structure and Rev binding sites. The RRE is composed of five stem-loops, numbered from I to V, arranged around a 4- or 5-way junction. Alternative conformations of III and IV (in magenta) and IV and V (in orange) are shown in the insets. The minimal RRE sequence identified to retain export activity is ~240-nt long (delimited by the vertical dashed line). Sequence specific Rev binding sites are colored: IIB in green; junction ABC in blue and stem IA (IA) in cyan. Three regions (R1, R2, R3) that show the highest protection upon Rev binding in SHAPE experiments are circled in grey. RRE61 mutations G165A and G254A are highlighted in red (*Adapted from Fernandes et al., 2016*).

The high degree of flexibility of the RRE makes its study with high resolution structural techniques very challenging. Two recent studies of the RRE by small angle X-ray scattering (SAXS), which provides low-resolution information, had shed light on its overall 3D organization (Figure 1.18). The first study showed that a ~230-nt RRE fragment assumes a topology resembling the letter A, with the loop regions of III/IV and V located close together at the apex of the A-like shape, whereas the IIB and IA are located at the opposite end and separated by ~55 Å (Fang et al., 2013). In the second study, a combination of SAXS and SHAPE (Selective 2'-hydroxyl acylation analyzed by primer extension) analysis was

used to study the ~350 nt RRE and proposed a model with the elongated part of stem I folding back towards the core of the central junction through long distance tertiary interactions (Bai et al., 2014). The use of this second model to propose assembly models for the RRE/Rev complex is discussed further below (§ 1.2.5).

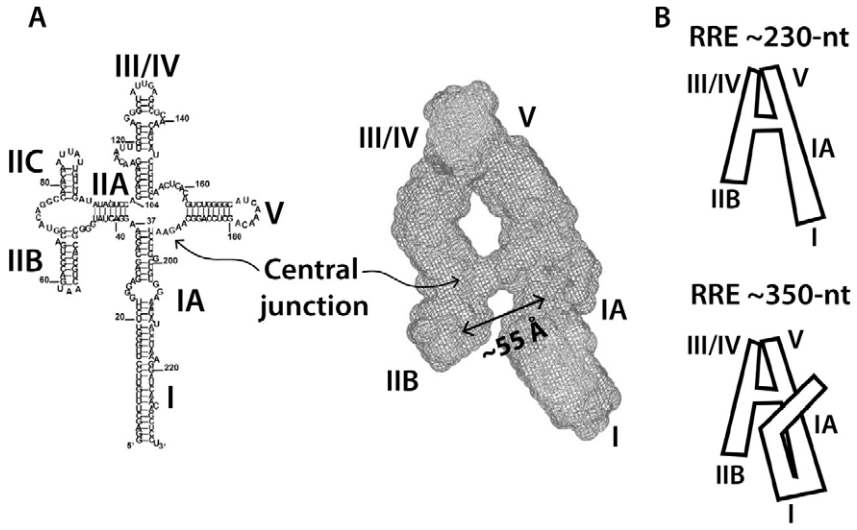


Figure 1.18: 3D structure models of the RRE. A) The secondary structure and SAXS envelope (in grey) of the ~230-nt RRE truncation with the position of the different stem loops indicated. B) Shapes determined by SAXS of a ~230-nt (Fang et al., 2013) and ~350-nt (Bai et al., 2014) RRE are depicted as cartoons

1.2.4 Molecular basis of Rev-RRE interactions

Many efforts have been made to characterize the Rev binding sites on the RRE at the molecular level and to define the assembly pathway of the RRE/Rev complex. The binding sites of Rev on the RRE identified so far are highlighted in Figure 1.17. II B, junction ABC and IA are sequence-specific sites; the first two of these have been structurally characterized (Battiste et al., 1996; Jayaraman et al., 2014) and the third by a mutagenesis approach (Daugherty et al., 2008). Three main regions of the RRE, designated R1, R2 and R3, have also been identified to contain Rev binding sites, as they showed the greatest protection in time-resolved SHAPE experiments upon the addition of Rev (Bai et al., 2014; Figure 1.17). R1 corresponds to IIB and the junction ABC site, R2 to the central junction and the nearby purine-rich IA site and R3 to a purine-rich region located at the center of stem I. Importantly, R3 also contains the area involved in long-distance tertiary interactions that help stem I fold back toward the core junction (Bai et al., 2014). All the Rev binding sites are purine-rich sequences and many are predicted to form internal loops

or bulges. So far no binding site has been located on III, IV and V, which, interestingly, are the most flexible portions of the RRE.

IIB is the highest affinity-binding site of Rev on the RRE. Early studies showed that the α-helical ARM of Rev recognizes a specific sequence of the IIB characterized by non-canonical Watson-Crick base pairs (Bartel et al., 1991; Tiley et al., 1992; Iwai et al., 1992). The structure of a synthetic ARM peptide (residues 33 to 55) bound to a 34-nt RNA fragment, recapitulating IIB, was solved by NMR (Battiste et al., 1996). This structure revealed how two non canonical base pairings (G:G and G:A) and a bulged uridine widen the RNA major groove so as to facilitate the accommodation of the α-helical ARM (Figure 1.19 A). Four Rev residues (R35, R39, N40 and R44) make base-specific contacts, while other ARM residues, notably arginines, interact non-specifically with the phosphate backbone (Figure 1.20 B). Mutations that affect the ARM residues or the bases involved in the interaction drastically reduce the affinity of Rev for the RRE, demonstrating the sequence specificity of the IIB binding site (Tan et al., 1993).

The crystal structure of a Rev dimer bound to a 39-nt RNA, containing the IIB extended to mimic the IIABC junction, revealed the molecular details of Rev interaction at this site (Jayaraman et al., 2014). The Rev monomers cross each other in a 'V' shape with one ARM binding to IIB and the other binding to the ABC junction site (Figure 1.19 B). Surprisingly, the interactions between the ARM and the ABC junction involve different Rev residues than those involved in IIB recognition. Most of the ARM-RNA contacts are made via the phosphate backbone, with only R43 and R44 making base-specific interactions (Figure 1.20 B). Similarly, the characterization of the binding site of Rev on IA by mutagenesis suggested a third mode of recognition of the RNA involving a distinct group of Rev arginine residues (R38, R41, R46) (Daugherty et al., 2008) (Figure 1.20 B).

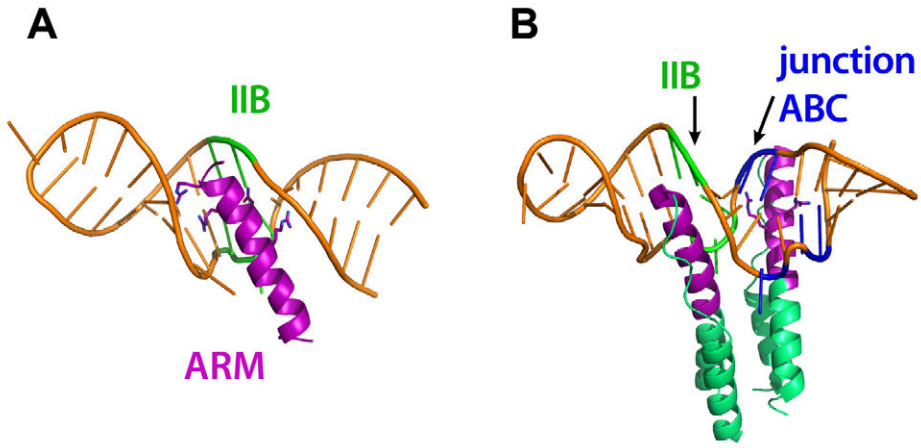


Figure 1.19: Structurally characterized Rev binding sites on the RRE. A) NMR structure of the Rev ARM (magenta) bound to IIB (orange). Rev arginine residues and IIB nucleotides involved in specific contacts are shown as sticks and colored in green, respectively (Battiste et al., 1996; PDB ID 1ETF). B) Crystal structure of a T/T Rev dimer in complex with an elongated (39-nt) IIB (orange) containing SLB II and the adjacent junction ABC binding site of Rev. ARM and OD are colored in magenta and green, respectively. Rev arginine residues and junction ABC nucleotides making base specific contacts are shown as sticks and colored in blue, respectively (Jayaraman et al., 2014; PDB ID 4PMI).

1.2.5 Rev/RRE is a dynamic viral RNP

Crystal structures of Rev alone and in complex with RNA have demonstrated the high flexibility of both the Rev-Rev and Rev-RNA interfaces (Figures 1.16 and 1.19 B). The OD interfaces of Rev are highly plastic and can drastically re-arrange upon RNA binding. Importantly, as mentioned above, Rev multimerization on the RRE has a polarity: the second Rev molecule binds through the hydrophobic T-T interface, followed by a series of H-H, T-T and C-C hydrophobic interactions, the order of which is unclear (Jain and Belasco, 2001; DiMattia et al., 2010; Daugherty et al., 2010b; Jayaraman et al., 2014; DiMattia et al., 2016). These interfaces have a high degree of flexibility as shown by the large variation in crossing angle observed in different crystal structures (100-140°; DiMattia et al., 2016) (Figure 1.20 A; see also section 1.2.2.1).

Rev interaction with different RRE binding sites is also highly versatile: different arginines of the Rev ARM are involved in base specific interactions with the IIB, ABC junction and IA binding sites (Figure 1.20 B). The helical nature of the ARM, decorated by multiple arginines, allows Rev to adapt to each binding site. It is conceivable that base specific interactions on other uncharacterized RRE binding sites would involve different

combinations of ARM residues. This plasticity allows each incoming monomer of Rev to adopt the best conformation to dock to the next available binding site on the RRE and may explain why it has been so difficult to identify other sequence-specific secondary binding sites.

Thus, the different binding modes of the Rev ARM on the RNA ensure that a single small protein yield a highly specific RNP, with a diversity of protein-RNA contacts typical of multi-domain or hetero-oligomeric complexes (Daugherty et al., 2008).

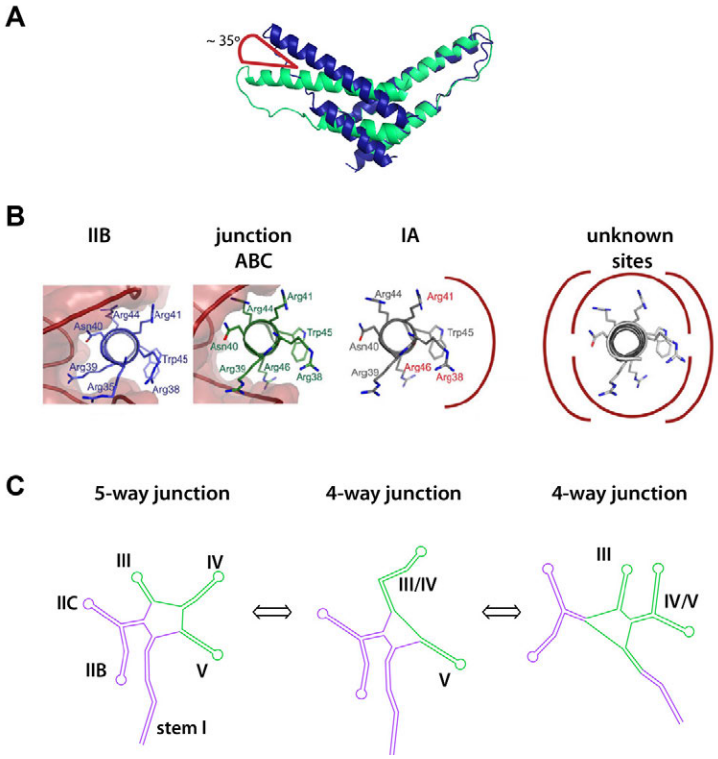


Figure 1.20: RRE-Rev is a highly dynamic RNP. A) Flexibility of the Rev-Rev interaction. Two crystal structures of Rev dimers (in blue and green) interacting through the H-H interface are superposed on one monomer to show the variability in the crossing angle (in red). B) Flexibility of the Rev-RNA interaction. The interaction between the Rev ARM and the different binding sites on the RRE (IIB, junction ABC and IA) is characterized by a different positioning of the Rev ARM in the major groove. C) Flexibility of the RRE. In the schematic representation of the RRE (part of the stem I is omitted for simplicity) the domains that adopt different conformations are highlighted in green.

Finally, the RRE itself is highly dynamic and can indeed adopt different conformations with most of the variability related to the III, IV and V (Figure 1.20 C). Studies suggest that the RRE is an active scaffold which plays an important role in driving the correct oligomerization of Rev. Initial evidence comes from the fact that Rev oligomerization is

directional and defined by the proper spacing of flexible regions along the RRE, required to locally distort the major grooves and facilitate the accommodation of the Rev ARMs (Zemmel et al., 1996). Other studies showed that the structural organization of the RRE is crucial for the formation of an RNP that is actively exported *in vivo*. For example, replacing the RRE by an array of IIB elements does not yield the same export efficiency *in vivo*, despite the ability of this array to accommodate multiple copies of Rev *in vitro* (Kjems et al., 1993; Symensna et al., 1999). The importance of the RRE structure is underlined by mutations that confer resistance to a Rev NES mutant (M10) unable to export viral mRNAs (Malim et al, 1989). Two silent mutations on the RRE (G164A and G254A, designated RRE61, see Figure 1.17) rescue the export activity of the Rev M10 mutant (Hamm et al., 1999). SHAPE analysis of the RRE61 showed that these mutations stabilize the RRE in a 5-way structure suggested to be the conformation responsible to rescue export activity (Legiewicz et al., 2008). A recent study confirmed that the structural variability observed for III, IV and V influences the efficiency of export activity *in vivo,* supporting the hypothesis that the 5-way RRE structure is indeed more efficient (Sherpa et al., 2015).

1.2.6 RRE/Rev/CRM1/Ran^{GTP} : an architecture to be discovered

Based on the biochemical and structural data available so far, certain key steps of the RRE-Rev assembly have been defined. The initial binding event always occurs on IIB and is sequence specific. This first binding event is the nucleation site from which other Rev molecules are recruited cooperatively on the RRE through a series of Rev-Rev and Rev-RNA interactions. Kinetic studies using single molecule fluorescence spectroscopy showed that Rev oligomerization on the RRE proceeds stepwise with the binding of one Rev molecule at a time (Pond et al., 2009). Despite recent advances in our knowledge of Rev-Rev and Rev-RNA interactions, the exact sequence of events involved in Rev oligomerization, the position of individual Rev molecules and the 3D structure of the entire complex remain unknown.

An RRE-Rev assembly model was proposed based on SAXS analysis of the full-length RRE and time-resolved SHAPE analysis of the RRE in the presence and absence of Rev (Bai et al., 2014). In the proposed induced-fit model, at least 3 Rev dimers assemble sequentially at Regions 1-3 (R1, R2 and R3) of the RRE (Figure 1.21 A).

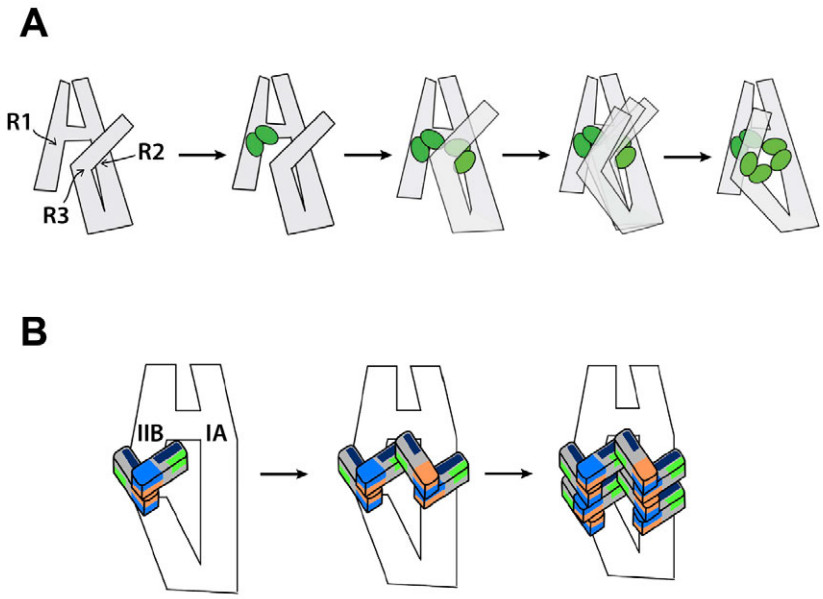


Figure 1.21: RRE-Rev assembly models. A) Induced-fit model. Rev is represented as a green oval and the RRE in grey, with R1, R2 and R3 positions indicated (Bai et al., 2014). The binding of the first two Rev dimers is fast and induces a conformational change of the RRE-Rev complex that renders the R3 region accessible for binding a third Rev dimer. B) Update of the induced-fit model proposed by DiMattia et al., 2016. Rev is represented as a parallelepiped with the T-T, H-H and C-C interfaces colored in light blue, orange and green respectively and the ARMs in dark blue. In the model the C-C interface is proposed to be involved in the check-point step during assembly and to bridge Rev dimers across the two RRE legs.

A first Rev dimer binds at R1 quickly followed by a second dimer at R2. The binding of a Rev dimer at R1 is consistent with the first nucleation event on IIB and the binding of a second Rev molecule on the ABC junction as observed in the crystal structure of Jayaraman et al, 2014. In addition the proximity (~55 Å) between IIB in R1 and 1A in R2 in the 3D model of the RRE (Fang et al., 2013; Bai et al, 2014) (Figure 1.18) would explain the rapid binding of the second Rev dimer on R2. Rev binding on R1 and R2 are proposed to destabilize the tertiary interactions between stem I and the central junction, inducing a conformational change that makes R3 accessible for accommodating a third Rev dimer. This last binding event is estimated to be slower than the first two. It was proposed that the initial binding of the two Rev dimers constitute a checkpoint for specificity, sampling the best RRE-Rev change of conformation for the binding of the third dimer.

With the recent discovery of a third (C-C type) Rev oligomerization interface (Figure 1.16 C), an updated model was proposed to explain plausible Rev-Rev interactions involved in the positioning of different Rev dimers along the RRE (Figure 1.21 B) (DiMattia

et al., 2016). This model proposes the C-C interface to form a bridge between Rev dimers across distant legs of the RRE. Specifically, a first T-T-type Rev dimer would bind at R1, a second H-H-type dimer would bind at R2, and the two Rev dimers would interact via a C-C-type interface. This last step is suggested to correspond to the specificity checkpoint proposed by the induced-fit model. Then, multiple Rev dimers would oligomerize along both legs of the RRE.

Notably, however, the updated RRE/Rev assembly model does not specify the exact number of Rev molecules that bind to the RRE. Moreover, the model does not clarify how many CRM1/Ran^{GTP} molecules are recruited and where these molecules are positioned. This is partially due to the highly flexible structure of the C-terminal region of Rev, where the NES is located. The RRE/Rev complex presents an array of NESs that could in theory recruit an equal number of CRM1 molecules. However, the large size of CRM1 may sterically limit the number bound. For these reasons, more knowledge on the architecture and stoichiometry of the RRE/Rev/CRM1/Ran^{GTP} is needed.

1.3 AIM OF THIS WORK

The principal objective of my PhD thesis has been to elucidate the architecture of the RRE/Rev/CRM1/Ran^{GTP} complex and more specifically to investigate its stoichiometry. The high flexibility of this RNP makes it very challenging to study by structural techniques such as X-ray crystallography or EM, for which a relatively homogenous sample is required. We decided to investigate the stoichiometry of the complex using native mass spectrometry (native MS), a technique which in recent years has become increasingly useful as a tool to complement other structural methods. Three main advantages of this technique were exploited during this study: the possibility to analyze intact complexes by preserving weak non-covalent interactions; the possibility to analyze polydisperse samples (i.e., with multiple oligomerization states); the high sensitivity of native MS as it requires low (μM) sample concentration.

Our goal was to determine the stoichiometry *in vitro* of the RRE/Rev/CRM1/Ran^{GTP} complex. This task was anticipated to be extremely challenging for several technical reasons: the highly dynamic nature and polydispersity of the complex; the difficulty to obtain soluble wild-type Rev protein at high purity and yield; the production of RNA-containing complexes compatible with native MS. Furthermore, the complexity of the task was amplified by the fact that at the time I began my PhD no one else in our lab was actively working with RNA. Furthermore, no one in our institute had ever performed native mass spectrometry on RNA-containing samples. This project was thus an adventure into unchartered territory. Although we did not completely achieve the ambitious goal which we had initially set for ourselves, we nevertheless learned a great deal about the analysis of RNPs by native mass spectrometry, as well as a thing or two about the RRE-Rev export complex. In the end, our results regarding the stoichiometry of the full RNP export complex are highly encouraging. Furthermore, as shown in the following chapters, we demonstrated that RNPs can be successfully analyzed by native MS, which represents a major advance in the field.

2. EXPERIMENTAL PROCEDURES

ABSTRACT

During this study the RRE/Rev/CRM1/Ran export complex was *in vitro* assembled and reconstituted. CRM1, Ran and Rev proteins were recombinantly expressed and full length RRE and a 66-nucleotide RRE fragment (IIABC) were *in vitro* transcribed. All the components were purified in large-scale using common biochemical procedures. Electrophoretic Mobility Shift Assay was used to set up and optimize *in vitro* assembly of the complex. In order to analyze such viral RNPs by native mass spectrometry, both protein and RNA purification procedures as well as assembly conditions were extensively adapted and optimized.

RÉSUMÉ

Au cours de cette étude, le complexe d'export RRE/Rev/CRM1/Ran a été assemblé et reconstitué *in vitro*. Les protéines CRM1, Ran et Rev recombinantes ont été produites exprimées par recombinaison alors que le RRE entier et un fragment du RRE de 66 nucleotides (IIABC) ont été transcrits *in vitro*. Tous les composants ont été purifiés à grande échelle en utilisant des procédures biochimiques communes. Le test de mobilité électrophorétique a été utilisé pour mettre en place et optimiser l'assemblage *in vitro* du complexe. Afin d'analyser ces RNP viraux par spectrométrie de masse native, les procédures de purification de protéines et d'ARN ainsi que les conditions d'assemblage ont été largement adaptées et optimisées.

2.1 PROTEINS AND RNA CONSTRUCTS

2.1.1 CRM1 AND RAN CONSTRUCTS

Prokaryotic and eukaryotic vectors for expressing CRM1 and Ran were already available in the lab and are listed in Table 2.1. Point and deletion mutations were generated using a QuikChange® Site-Directed Mutagenesis Kit (*Stratagene*).

Prokaryotic Expression Vector	pQE60-CRM1, * pROEX RAN^{Q69L}
Eukaryotic Expression Vector	pSG5-CRM1-GFP, *

Table 2.1: Prokaryotic and eukaryotic expression vectors coding for CRM1 and RAN. * mutants of CRM1 (M583A/K590A, N495R, ΔLoop, K693E/R753E, V430A/L431A/V432A,) are also cloned in pQE60 and pSG5 vectors.

2.1.2 HIV-1 REV CONSTRUCT

Previous studies have shown that the expression, solubility and purification of Rev could be facilitated by fusion to the acidic B1 domain of protein G (GB1 domain) (Daugherty et al., 2008 and 2010; Jayaraman et al., 2014; 2014 Booth et al., 2014). Consequently, the HIV-1 Rev protein sequence was fused to an N-terminal His tag followed by the GB1 domain and a TEV protease cleavage site. To obtain such a construct, the Rev sequence (Cochrane et al., 1989) was cloned into a pET28(a^+) vector previously modified to contain a His-GB1-TEV site sequence followed by a multi-cloning site (MCS), that was kindly provided by the group of Martin Blackledge (IBS). The Rev DNA sequence was PCR amplified using primers designed to add *Nde*I and *Xho*I restriction sites upstream and downstream of the sequence respectively, then the PCR fragments were sub-cloned into a pJET cloning vector (*CloneJET™ PCR Cloning Kit, ThermoScientific*). The pJET-insert plasmid was purified at high yield and purity (*NucleoSpin® Plasmid, MACHEREY*

NAGEL) and digested with *Nde*I/*Xho*I enzymes (*New England Biolabs*). The restriction fragment containing the desired coding sequence was then ligated into a pET28(a^+) vector (linearized with the same enzymes) using T4 ligase (*Fermentas*) following the manufacturers' instructions. The ligation products were used to transform *E. coli* DH5α competent cells (*New England Biolabs*) and positive colonies were selected through appropriate antibiotic resistance. Positive plasmids containing the Rev insert were selected by restriction digestion with *Nde*I/*Xho*I and verified by DNA sequencing. The presence of an *NdeI* restriction site upstream of Rev introduced a His residue between the end of the TEV cleavage site and the first Met of Rev, and so this codon was removed by site directed mutagenesis.

2.1.3 RRE AND STEM IIABC CONSTRUCTS

DNA sequences coding for RRE and stem IIABC (IIABC) RNAs were cloned into a pUC19 vector, downstream of a T7 polymerase promoter for *in vitro* transcription.

The RRE sequence used in this study is 353 nt long and corresponds to nucleotides 7713 – 8066 of the viral genome from the ARV-2/SF2 strain (NCBI ref. K02007.1). The RRE gene inserted in a pAM-T vector and flanked by an upstream *Eco*RI and a downstream *Hind*III site was purchased from *Life Technologies*. This vector was first purified at high yield and purity (NucleoSpin® Plasmid, *MACHEREY NAGEL*) and then digested with *Eco*RI/*Hind*III enzymes (*New England Biolabs*). The restriction fragment containing the sequence of interest was then ligated into a pUC19 vector (linearized with the same enzymes) with a T4 ligase (*Fermentas*) following the manufacturers' instructions. The ligation products were used to transform *E. coli* DH5α competent cells (*New England Biolabs*) and positive colonies were selected through appropriate antibiotic resistance. Positive plasmids containing RRE insert were selected by restriction digestion with *Eco*RI/*Hind*III and verified by DNA sequencing.

IIABC RNA is 66 nt long and corresponds to nucleotides 102 - 167 of the RRE sequence. In the pUC19 vector a sequence coding for the hammerhead rybozyme (HHR) is included downstream of IIABC to ensure the production of transcripts with a homogenous 3'-end. A IIABC–HHR construct flanked by *Eco*RI and *Hind*III sites was produced using 6 overlapping oligonucleotides (designed using DNA work software), annealed and amplified by PCR. The purified reaction products were digested with *Eco*RI/*Hind*III (*New*

England Biolabs) and ligated into the pUC19 vector (linearized with the same enzymes) using T4 ligase (*Fermentas*) according to the manufacturers' instructions. The ligation products were used to transform *E. coli* DH5α competent cells (*New England Biolabs*) and positive colonies were selected through appropriate antibiotic resistance. Plasmids containing the IIABC-HHR insert were selected by restriction digestion with *Eco*RI/*Hind*III and verified by DNA by sequencing.

2.2 RECOMBINANT PROTEIN EXPRESSION AND PURIFICATION

2.2.1 COMMON PROCEDURES

All proteins are fused to a His-tag to allow purification by Immobilised Metal Affinity Chromatograpy (IMAC), which was performed using 5 mL His Trap™ FF columns (*GE Healthcare*) pre-packed with Ni-Sepharose on a BioLogic™ Low Pression liquid chromatography system (*BIO-RAD*). TEV protease (N-terminally fused to a His-tag) was recombinantly expressed, purified in house and stored at -80 ºC prior to use. Size exclusion chromatography (SEC) was performed using Superdex colums (*GE Healthcare*) with various fractionation ranges on an AKTA Prime chromatography system (*Amersham Biosciences*). Protein purity at each purification step was evaluated by discontinuous polyacrylamide gel electrophoresis in denaturing and reducing conditions (SDS-PAGE), using different polyacrylamide percentages according to the sample molecular weight. Protein concentration was estimated by a Bradford assay (*BIO-RAD*) in a microtiter plate following the manufacturers' procedures, where protein and standard (bovine serum albumin, *SigmaAldrich*) solutions were assayed in duplicate or triplicate.

2.2.2 CRM1 EXPRESSION AND PURIFICATION

E. coli Tg1 competent cells (*Zymo Research*) transformed with pQE60 CRM1 (wild-type or mutants) were grown at 37 ºC in LB medium containing ampicillin (100 μg/mL) until an OD_{600} of 0.6. Expression was induced with 0.5 mM IPTG and cells incubated at 30 ºC for 5h. Harvested cells were resuspended in lysis buffer (100 mM Tris/HCl pH 7.6, 1M NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, 1 mM PMSF, protease inhibitors) in the presence of lysozyme (1 mg/mL), lysed by sonication at 4 ºC and harvested at 50,000g for

30 min. A 35% ammonium sulfate precipitation step was performed on the cleared lysate. The pellet was resuspended in buffer A (50 mM Hepes pH 7.7, 0.3 M NaCl, 20 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol, 1 mM PMSF) and applied to a His Trap FF column. Immobilized His-CRM1 was washed with buffer A containing 1 M NaCl, eluted with buffer A containing 0.25 M imidazole, concentrated by ultrafiltration and further purified on a Superdex 200 16/60 column in 20 mM Tris/HCl pH 7.6, 0.2 M NaCl, 10 mM DTT, 10% glycerol (Figure 2.1). The fractions of highest purity were pooled and concentrated to 60-80 mg/mL by ultrafiltration.

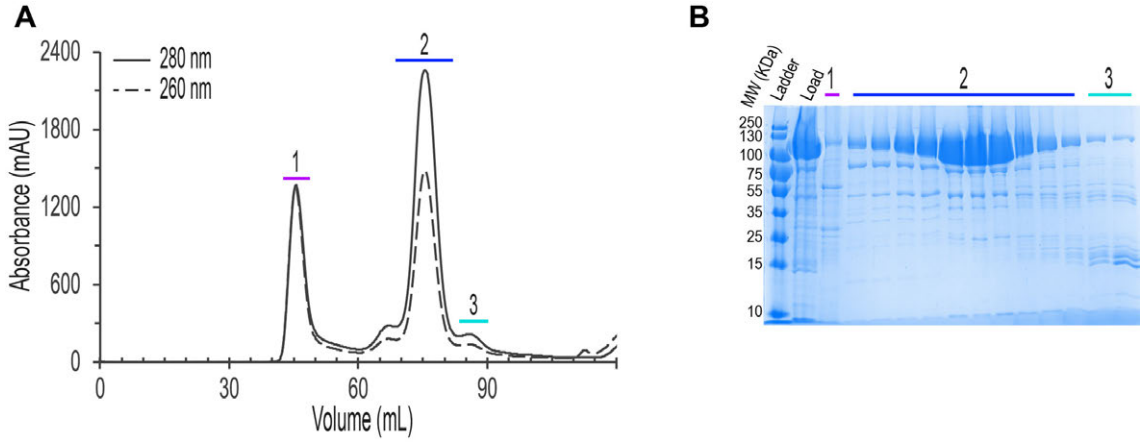


Figure 2.1: SEC profile of His-CRM1. A) His-CRM1 elution profile on the Superdex 200 16/60. **B**) 12% SDS-PAGE of fractions collected: sample injected (Load); different fractions of the elution loaded on gel are identified with coloured lines (purple, blue and cyan).

2.2.3 RAN EXPRESSION AND PURIFICATION

E. coli BL21 DE3 competent cells, transformed with the pROEX Ran^{Q69L} vector, were grown at 37 °C in LB medium containing ampicillin (100 µg/mL) until reaching an OD_{600} of 0.8 and expression was then induced with 0.5 mM IPTG for 4h. Harvested cells were resuspended in buffer A (30 mM Tris/HCl pH 7.6, 180 mM NaCl, 15 mM imidazole, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 1 mM PMSF, protease inhibitors) in the presence of lysozyme (1 mg/mL), lysed by sonication at 4 °C then harvested at 50,000g for 30 min. The cleared lysate was incubated with 1 mM GDP in the presence of 10 mM EDTA, and then 15 mM $MgCl_2$ was added after 30 min incubation to stop the nucleotide exchange reaction. Following application to a His Trap FF column, immobilized His-Ran^{Q69L} was washed with buffer A containing 1 M NaCl and eluted with buffer A containing 1 M NaCl and 0.25 M imidazole. Eluted fractions were pooled, His-TEV protease was added to cleave the His tag while dialysing against buffer A with 10% glycerol for 12-14h at 4 °C. Ran^{Q69L} was separated

from His tag and His-TEV protease by a second Ni^{2+} IMAC and concentrated by ultrafiltration. A 300-fold molar excess of GTP was added to the sample in the presence of 10 mM EDTA and incubated for 12-14h, followed by the addition of 15 mM $MgCl_2$ to stop the nucleotide exchange reaction. Ran^{Q69L} was further purified on a Superdex 75 16/60 column in 10 mM Tris/HCl pH 7.5, 70 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT (Figure 2.2 A, B). Fractions of highest purity were pooled and concentrated to 20-30 mg/mL by ultrafiltration. A native gel indicated that approximately half of the protein was in the GTP-bound state and the rest was bound to GDP (Figure 2.2 C).

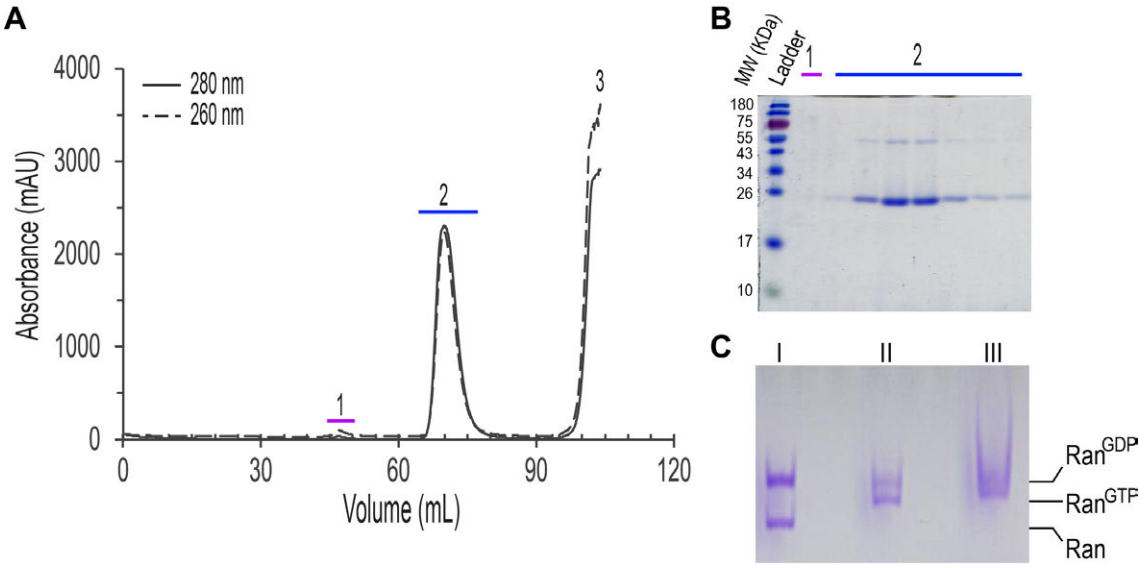


Figure 2.2: SEC profile of GTP-bound Ran^{Q69L}. A) Elution profile on the Superdex 75 16/60 column. The high 280/260 ratio of peak 2, eluting at 69.7 mL, shows that GTP is bound to the protein; the excess of GTP elutes at peak 3. B) SDS-PAGE of fractions collected; different fractions loaded on gel are identified with coloured lines (purple and blue). C) Native gel (6% acrylamide-TBE) showing the nucleotide-bound state of Ran at different steps of the purification: I. Ran before GTP nucleotide exchange; II: Ran after GTP nucleotide exchange and before SEC; III: Ran after SEC.

2.2.4 Rev expression and purification

E. coli BL21 DE3 competent cells transformed with pET28(a^+) His-GB1-Rev (wild-type or mutants) were grown at 37 ºC in LB medium containing kanamicin (50 µg/mL) until an OD_{600} of 0.6, expression was induced with 0.5 mM IPTG and cells incubated at 30 ºC for 4h. Harvested cells were resuspended in lysis buffer (25 mM HEPES pH 7.5, 200 mM NaCl, 0.1% Tween-20, 2 mM β-mercaptoethanol, 1 mM PMSF, protease inhibitors) in the

presence of lysozyme (1 mg/mL) and DNAseI (10 μg/mL), lysed by sonication at 4 °C and then harvested at 50,000g for 30 min.

For a description of the different forms of Rev used and an outline of the early purification steps, see §2.1.1.2 and Figure. 2.1 B. The cleared lysate was incubated for 30 min-45 min with RNAse A (50 μg/mL) and RNAse T_1 (50 U/mL) in the presence of 1 M NaCl at room temperature and subsequently applied to a His Trap FF column. Immobilized His-GB1-Rev was washed first with buffer A (50 mM Tris/HCl pH 8, 2 M NaCl, 0.1% Tween-20, 10mM imidazole, 2 mM β-mercaptoethanol, 1 mM PMSF) then with buffer A' (Buffer A containing 0.25 M NaCl and no detergent). A successive wash step was performed with buffer A' mixed with 10% of buffer B (A' containing 0.25 M imidazole) and finally the protein was eluted using 100% of buffer B. Eluted fractions were pooled and His-TEV protease was added (in a ratio His-GB1-Rev : TEV of 40:1) while dialysing against buffer C (40 mM Tris/HCl pH 8, 0.2 M NaCl, 0.1 M Na_2SO_4, 0.4 M $(NH_4)_2SO_4$, 2mM β-mercaptoethanol) for 12-14h at 4 °C. Rev was separated from His-GB1 tag and His-TEV protease by a second Ni^{2+} IMAC step and concentrated by ultrafiltration.

Rev was further purified by size exclusion chromatography (SEC). Oligomerization defective mutants were purified using a Superdex 75 16/60 in buffer D (40 mM Tris/HCl pH 8, 0.2 M NaCl, 1mM DTT); Rev^{WT} was purified using a Superdex 200 16/60 in buffer C (with 1 mM DTT instead of β-mercaptoethanol). The fractions of highest purity were pooled and concentrated to 10-30 mg/mL by ultrafiltration.

2.3 RNA *IN VITRO* TRANSCRIPTION AND PURIFICATION

Stem IIB (IIB) was purchassed from *IDT* as a lyophilized 34 nt oligonucleotide with a TAMRA fluorophore covalently bound at the 3' end. RNA was resuspended in water to a final concentration of 10 μM, divided in aliquots, snap frozen and stored at -80°C.

RRE and IIABC RNAs were prepared by *in vitro* transcription from a linearized plasmid, using T7 RNA polymerase expressed and purified in house. Linear pUC19 was generated by digestion with *Hin*dIII and purified by phenol/chloroform extraction followed by ethanol precipitation. The pellet was washed with 70% ethanol and the DNA template was resuspended in milliQ water prior to use. Template quality and linearization were evaluated at each step by TAE agarose gel electrophoresis. The transcription reaction

was initially performed on a small scale (30 µL) to optimize critical parameters such as buffer composition, $MgCl_2$ and template concentration and to identify conditions giving the highest yield of RNA per mL of transcription reaction. RNA yield and purity during different purification steps were evaluated by 10% denaturing urea polyacrylamide gel electrophoresis (urea-PAGE), followed by staining with gel red (*Biotium*). RNA concentration was calculated by UV-VIS absorption using the theoretical epsilon (ϵ_{th}) calculated from nucleotide composition, according to the equation:

$$C = \frac{Abs_{260} - Abs_{230}}{\epsilon_{th}}$$

For IIABC and RRE the values of ϵ_{th} are 746,666 and 4,163,251.3 $mol^{-1} \cdot L.cm^{-1}$, respectively.

2.3.1 IIABC

IIABC was produced by starting 5 mL of transcription reaction, incubated for 4 h at 37°C using 0.25 µg/µL of linear DNA template in the presence of 40 mM Tris/HCl pH 8, 24 mM $MgCl_2$, 80 mg/mL PEG 8000, 0.01% Triton X-100, 1 mM spermidine, 34 ng/µL BSA, 5 mM DTT, 4 U/mL PPase, 4 mM of each rNTP and 0.1 µg/µL of T7 polymerase. DNAseI was added to stop the reaction, incubating for 20 min at 37°C. The sample was diluted 5-fold in buffer A (8 M urea, 20 mM Tris/HCL pH 8, 5 mM EDTA), loaded on a Q-sepharose column (70 ml packed in house), washed with 45% buffer B (8 M urea, 1 M NaCl, 20 mM Tris/HCL pH 8, 5 mM EDTA) and eluted using a 45-65% gradient of buffer B (Burdisso et al., 2012). The highest purity fractions containing IIABC were pooled and dialysed against water until the urea and salt concentration were negligible. After quantification IIABC was lyophilized in aliquots and stored at -80°C. This protocol yielded, on average, 0.5 mg of pure IIABC per mL of transcription reaction.

2.3.2 RRE

RRE was produced by starting 1-2 mL of transcription reaction, incubated for 2 h at 37°C using 20 ng/µL of linear DNA template in the presence of 40 mM Tris/HCl pH 8, 22

mM $MgCl_2$, 10 mM NaCl, 0.01% Triton X-100, 2 mM spermidine, 10 mM DTT, 4 mM of each rNTP and 0.1 μg/μL of T7 polymerase. Some other transcription buffers have also been tested as listed in Table 2.2. The DNA template was digested adding 0.5 U of TURBO DNAse (*Life Technologies*) per μg of template incubating at 37ºC for 30 min and RNA polymerase was digested by adding 1.5 μg/mL of proteinase K and incubating at 37ºC for a further 45 min. In order to filter out enzymatic degradation products, the sample was buffer exchanged using an ultrafiltration device (100 KDa cut-off) in 8 mM MOPS pH 6.5, 100 mM KCl, 0.1 mM EDTA pH 8. RRE was further purified by size exclusion chromatography using a Sephacryl 400 10/300 (25 ml packed in house) run on an AKTA purifier FPLC system. The main peak fractions were analyzed by native or denaturing gel electrophoresis and either lyophilized or precipitated with Na Acetate in the presence of 100% EtOH to avoid degradation during storage. For certain experiments, RRE was then purified from abortive transcripts by a 10% denaturing polyacrylamide urea-PAGE run at 80 W for 8h. The band corresponding to RRE was identified by UV_{254} shadowing and excised from the gel using a sterile blade. The excised band was cut into slices and placed in one or more electroelution chambers. The RRE was recovered after 12-16 h continuous run at 100 V in TBE 0.5X buffer, and subsequently precipitated with 100% EtOH in the presence of Na Acetate. The pellet was washed with 70% EtOH and resuspended prior to use. This protocol yielded, on average, 0.5 mg of pure RRE per mL of transcription.

BUFFER (X10)	**1.2**	**1.3**	**2.2**	**3.2**
MgAc (mM)	220	100	220	100
NH_4Ac (mM)	400	600	400	600
NaAc (mM)	-	-	100	100
Spermidine (mM)	20	20	20	20
DTT(mM)	100	20	-	100
Triton X-100 (%)	0.1	0.1	0.1	0.1

Table 2.2: Acetate containing buffers used for the *in vitro* transcription of the RRE. Tris, NaCl and $MgCl_2$ have been removed and replace by NH_4Ac, NaAc, MgAc, respectively.

2.4 ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA)

During this study, EMSAs were used to evaluate complex assembly. 6% acrylamide-TBE gels were hand casted as 1 mm thick mini-PROTEAN gels (*BIO-RAD*) (10 or 15 wells) without a stacking gel. Gels were pre-run at 150 V for at least 30 min before loading samples. Complexes were incubated at 22°C for 20-30 min in standard binding buffer (40 mM Tris/HCl pH 8, 100 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT, 10% glycerol) or in 200 mM AmAc (with 10% glycerol), in a final volume varying between 5 to 15 µL.

Depending on RNA length and on the experiment, different amounts of nucleic acid were used to assemble the RNP complexes. In experiments performed with IIB, a final concentration of 0.1 µM was used, gels were migrated at 150 V for 30 min then excited at 530 nm and revealed at 580 nm using a transilluminator (Chemidoc Bio-RAD). In experiments performed with IIABC, a final concentration of 0.15 to 1 µM was used, gels were migrated at 150 V for 50 min. RRE was used at a final concentration of 0.1-0.2 µM and gels were run at 150 V for 3 h. In both cases, gels were stained using GelRed (*Biotium*) or SYBR gold (*Invitrogen*) and revealed using the appropriate channel under the transilluminator. In all cases, gels were stained using Coomassie brilliant blue or silver staining to detect protein and protein complexes.

2.5 NATIVE MASS SPECTROMETRY

2.5.1 BASIC PRINCIPLES OF ELECTROSPRAY IONIZATION-MASS SPECTROMETRY

Mass spectrometry (MS) is an analytical technique used to determine the mass of a molecule or a mix of molecules with high accuracy, sensitivity and rapidity. A mass spectrometer consists of three main parts: an ion source, a mass analyser and a detector (de Hoffmann and Stroobant, 2007). The ionisation source generates gas phase ions from solid or liquid species. The mass analyser separates ions according to their mass-to-charge ratio (*m/z*). An important analyser for biomolecules (i.e. peptides, proteins, DNA, RNA, macromolecular complexes) is the time-of-flight (TOF). The detector counts ions and measures their relative abundance. Signals from the detector are transmitted to a computer to generate mass spectra where *m/z* values of ions are plotted versus their intensities.

For biological applications the most commonly used ion source is electrospray ionization (ESI) (Fenn et al., 1989). During ESI, the sample, dissolved in a volatile solution, is loaded into a small conductive capillary and a voltage is applied to the capillary (Figure 2.3 A). The resulting electric field induces the accumulation of charges at the surface of the liquid present at the capillary tip. The charged liquid is forced to hold more and more electrical charges. When it cannot hold additional charge, it blows apart into many highly charged droplets. Droplets desolvate as they travel toward the entrance of the mass spectrometer through pressure and potential differences. At the end of the process each droplet contains a molecule of analyte.

In a MS spectrum of intact proteins and macromolecular assemblies, a Gaussian distribution of peaks corresponds to ions having different *m/z* values but the same *m* (Figure 2.3 B). The different peaks have distinct *z* values varying from that of their immediate neighbour by only one charge. To determine the *m* of a species present in a spectrum, two neighboring *m/z* values are determined experimentally (x and y) and two equations are written (m/z_1=x and m/z_2=y). Since $z_1=z_2$-1, the equations are solved to determine *m*, z_1 and z_2.

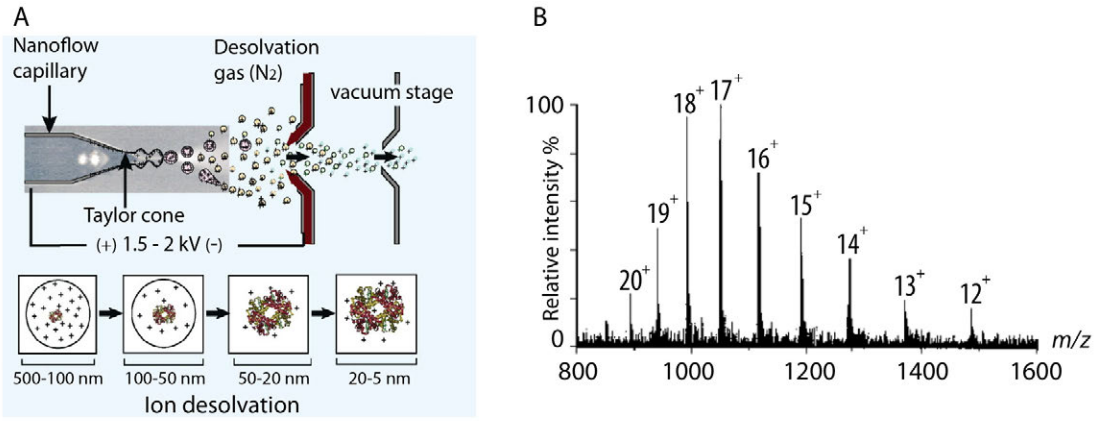


Figure 2.3: ESI source and acquired mass spectrum. A) Schematic representation of sample ionization by nano-ESI (*adapted from Sharon and Robinson, 2007*). B) A mass spectrum of phage λ lysozyme (calculated mass 17,825.2 Da) (*adapted from de Hoffmann and Stroobant, 2007*). During ionization a generated population of ions have the same mass, but different charge states.

2.5.2 Native MS as a tool to characterize macromolecular complexes

By working in conditions that preserve non-covalent interactions, native MS is used to determine the mass of macromolecular assemblies (i.e. protein complexes, DNA- and

RNA-protein complexes) with great accuracy. Thus, native MS is one of the most useful types of MS in the field of structural biology, emerging as a key technique for investigating the stoichiometry, architecture and dynamics of macromolecular complexes. For instance, native MS can be used to map the interactions between subunits and to deduce the assembly order of subunits forming a complex (Boeri Erba and Petosa, 2015).

Native MS has several advantages compared to other biochemical and biophysical techniques. It is highly sensitive, requiring usually only a few microliters of sample (5-20 µl) at low concentration (µM). It can probe a wide range of complexes varying in composition, mass, symmetry and degree of flexibility. A polydisperse sample, containing either multiple complexes or oligomeric states, can be analyzed, providing resolved information about the different species in solution. Finally, the technique does not require samples to be labeled or crosslinked.

To preserve non-covalent interactions native MS uses nano-ESI sources (Wilm and Mann, 1996). Typically, the diameter of nano-ESI capillaries ranges between 1 and 10 µM. Thus, the flow rate is slow (20-30 nL/min) and the required sample amount is in the range of picomoles. Moreover, the small size of the generated droplets allows the use of aqueous buffers without any heating.

Native MS experiments can be carried out only in a volatile buffer such as ammonium acetate (AmAc), and thus the sample must be buffer exchanged prior to analysis (Hernández and Robinson, 2007). This can be considered a limitation because some complexes may precipitate or disassemble in AmAc. The buffer exchange is a critical step because the presence of residual non-volatile salts, commonly used for biochemical preparations, hampers an accurate determination of the mass. In the case of nucleic acids, bound non-volatile salts are particularly challenging to remove because the negatively charged backbone of DNA and RNA binds cations (i.e. Na^+, K^+, Mg^{2+}) with high affinity. A concentration of AmAc ranging from 100 mM to 1 M can be used for the buffer exchange step, but these conditions must be optimized depending on the analysed complex.

The technique has other limitations. Heterogeneous complexes are challenging as different species can have very similar *m/z* values that hamper their unique assignment. Importantly, as the mass of the complex increases, many processes become less efficient, such as desolvation, transmission or detection. It should also be noted that some non-covalent interactions may have different strength in the gas phase compared to solution

(Hernández and Robinson, 2007). In particular, electrostatic interactions may be stronger in the gas phase, whereas hydrophobic interactions may be weaker. The abundance of the different complexes detected by MS may deviate from that in solution since the ionization process, transmission and detection are not equal for all the species.

The MS analysis of macromolecular complexes requires the use of modified instruments (Figure 2.4), because such biomolecules typically ionize with a *m/z* higher than 3000 that is normally detected by standard instruments (Sobott et al., 2002; van den Heuvel et al., 2006). Important modifications are the increment of the pressure (i) in the transfer region between the ionisation source and analyzer and (ii) inside the collision cell. These pressure increments improve the focusing of the ions and can be obtained by reducing the efficiency of some of the mass spectrometer pumps. Moreover inside the collision cell the voltage used to accelerate the ions (collision energy) can be modified.

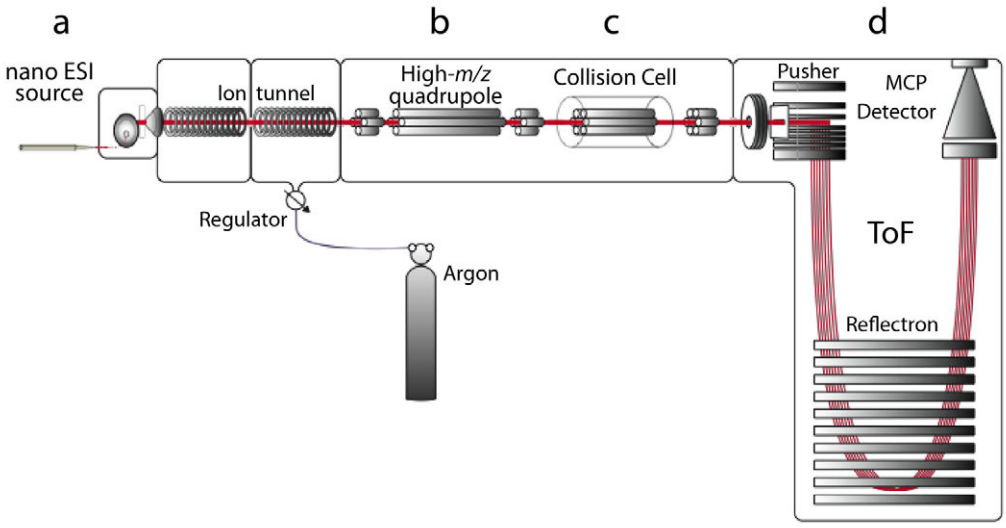


Figure 2.4: Schematic representation of a nano-ESI Q-TOF mass spectrometer. The sample is ionized in the nano-ESI ion source (a). The ions are driven into the mass spectrometer through a series of pressure and potential gradients until they reach the detector (red line). A quadrupole (b) is an analyser and consists of four parallel rods, in which the ions describe a complicated trajectory according to their *m/z*. If needed, dissociation of macromolecular assemblies can take place in the collision cell (c). When ions are the time-of-flight (TOF) analyser, (d) they are separated according to their *m/z*.

Normally, native MS experiments start with the acquisition of spectra in MS mode whereby all the ions reach the detector. This generates spectra representing an overview of

all the species present in the sample (a.k.a. spectra acquired in MS mode). By analysing such spectra, it is possible to identify relevant species which should be further processed. Thus, it is possible to select these species and acquire spectra in tandem MS mode (a.k.a. MS/MS mode). In these experiments, only a specific ion population is transmitted to the collision cell where it collides with molecules of an inert gas (e.g., Argon) (Figure 2.4). Collision induced dissociation (CID) causes the dissociation of a specific complex by ejecting monomers and forming the so-called "stripped complexes" (Kondrat et al., 2015). This allows the confirmation of the stoichiometry and the identification of the subunits at the core and at the periphery of a macromolecular assembly (Boeri Erba and Petosa, 2015).

2.5.3 NANO ESI Q-TOF

Protein, RNA and RNP ions were generated using a nanoflow electrospray (nano-ESI) source. Nanoflow platinum-coated borosilicate electrospray capillaries were bought from Thermo Electron SAS (Courtaboeuf, France). MS analyses were carried out on a quadrupole time-of-flight mass spectrometer (Q-TOF Ultima, Waters Corporation, Manchester, U.K.). The instrument was modified for the detection of high masses (Sobott et al., 2002; van den Heuvel et al., 2006). The following instrumental parameters were used: capillary voltage = 1.2-1.3 kV, cone potential = 40 V, RF lens-1 potential = 40 V, RF lens-2 potential = 1 V, aperture-1 potential = 0 V, collision energy = 30-140 V, and microchannel plate (MCP) = 1900 V. All mass spectra were calibrated externally using a solution of caesium iodide (6 mg/mL in 50% isopropanol) and were processed with the Masslynx 4.0 software (Waters Corporation, Manchester, U.K.) and with Massign software package (Morgner et al., 2012).

2.5.4 PROTEIN PREPARATION FOR MASS SPECTROMETRY

Before native MS experiments, all proteins were buffer exchanged in AmAc by size-exclusion chromatography (SEC). The AmAc concentration used for each protein was calculated to be higher than the total salt concentration used under standard conditions (Table 2.3). Following buffer exchange, protein fractions were pooled and concentrated by ultrafiltration, divided into single-use aliquots, snap frozen in liquid nitrogen and stored at-80 ^{o}C prior to use. Protein stability after storage was evaluated first by SDS-PAGE and

then confirmed by ESI-TOF-MS. The measured masses of proteins in denaturing conditions were used to predict the masses of complexes analysed by native MS.

PROTEIN	STANDARD BUFFER	NATIVE MS BUFFER	COLUMN
CRM1	20 mM Tris/HCl pH 7.6 0.2 M NaCl 10 mM DTT 10% glycerol	250 mM AmAc 10 mM DTT	S200 16/60
Ran	10 mM Tris/HCl pH 7.5 70 mM NaCl 5 mM $MgCl_2$ 5 mM DTT	150 mM AmAc 5 mM Mg Acetate 5mM DTT	S75 10/300
Rev*	40 mM Tris/HCl pH 8 0.2 M NaCl 1mM DTT	300 mM AmAc	S75 10/300
Rev^{V16D}	40 mM Tris/HCl pH 8 0.2 M NaCl 1mM DTT	800 mM AmAc	S200 10/300
Rev^{WT}	40 mM Tris/HCl pH 8 0.2 M NaCl 0.1 M Na_2SO_4 0.4 M $(NH_4)_2SO_4$ 1mM DTT	800 mM AmAc 1mM DTT	S200 10/300

Table 2.3: Native MS buffer composition and columns used for exchange.

2.5.5 RNA PREPARATION FOR NATIVE MS

The RRE and IIABC RNA molecules were resuspended in 1M AmAc, incubated for 1 h at 22°C and dialyzed against water. RNA was quantified, split in aliquots and precipitated adding 0.5*total sample volume of 7.5 M AmAc and then 2.5*total volume of 100% EtOH for 12-16h at 4°C. The pellet was then washed 3 to 5 times with 70% EtOH and stored at 4°C prior to use.

2.6 DENATURING MASS SPECTROMETRY

2.6.1 MALDI-TOF

Matrix Assisted Laser Ionization Time of flight Mass Spectrometry (MALDI-TOF MS) spectra of RNAs (IIABC, RRE) were acquired on an Autoflex mass spectrometer (*Bruker Daltonics, Bremen, Germany*) operated in linear negative ion mode.

External mass calibration of the instrument was obtained using either protein calibration standard II from *Bruker Daltonics* (containing trypsinogen: 23.9 kDa, protein A: 44.6 kDa and BSA: 66.5 kDa) or standard BSA alone (*Sigma*, purity ≥98%) using the peaks of the protein monomer and dimer at 66.5 and 133 kDa, respectively. The calibration provided mass accuracy <0.1 % (i.e., 1000 ppm) in the 10-200 kDa *m/z* mass range. This means an error of 0.2 kDa for a 200 kDa mass. Mass spectra data were processed with Flex Analysis software (v.3.0, *Bruker Daltonics*).

The MALDI matrix was prepared as follows: 3-hydroxypicolinic acid (3-HPA, *Fluka*, purity ≥99%, 50 mg/ml solution in ultrapure water/acetonitrile 50/50) wax mixed in a ratio 9:1 (*v*/*v*) with diammonium hydrogen citrate (DAC, Sigma, purity 99.9%, 100mg/ml solution in ultrapure water).

Before mass spectrometry analysis, the RNA samples (solutions in ultrapure water, with concentrations of 100 μM for IIABC and of 50 μM for RRE) were prepared following either the dried-droplet method (the samples were directly diluted in the matrix and 1-2 μl of the mixture were deposited on the target and allowed to air dry) or the double layer method (1-2 μl of matrix were applied on the target and allowed to dry, then 1 μl of sample was added and allowed to dry once more).

2.6.2 LC/ESI-TOF

Liquid Chromatography Electrospray Ionization Mass Spectrometry (LC/ESI-MS) was performed on a 6210 LC/ESI-TOF mass spectrometer interfaced with an HPLC binary pump system (*Agilent Technologies*). MS data acquisition was carried out in the negative ion mode with spectra in the profile mode. The mass spectrometer was operated with the following experimental settings: ESI source temperature was set at 350 °C; nitrogen was

used as drying gas (9 l/min) and as nebulizer gas (35 psig); the capillary needle voltage was set at 3500 V. Spectra acquisition rate was 1 spectrum/s.

The mass spectrometer was calibrated in the *m/z* 300-3200 range in the negative ion mode with standard calibrants (ESI-L, Low concentration tuning mix, *Agilent Technologies*) before each series of measurements. Mass spectra were recorded in the 450-3000 *m/z* range and the calibration provided mass accuracy <50 ppm in this *m/z* range. The MS data were processed with MassHunter software (data acquisition v.B.04.00, qualitative analysis with Bioconfirm v.B.07.00, *Agilent Technologies*).

All solvents used were HPLC-grade water (LC-MS Chromasolv quality, *Sigma-Aldrich*); methanol (*Carlo Erba*). Ammonium acetate (AmAc) 7.5 M solution was from *Sigma*. Solvent A was 50 mM AmAc pH 7.1 in water, solvent B was 100% methanol containing 10 mM AmAc.

Just before analysis, the RNA samples were diluted in 100 mM AmAc to a final concentration of 5 or 10 μM. Samples were thermostatted at 10°C in the autosampler and the analysis was run by injecting 4 or 8 μL of each sample. They were first trapped and desalted on a reverse phase-C18 cartridge (Zorbax 300SB-C18, 5μm, 300μm ID×5mm, Agilent Technologies) for 3 min at a flow rate of 50 μl/min with 100% solvent A and then eluted with 70% solvent B at flow rate of 50 μl/min for MS detection. The RP-C18 cartridge was then re-equilibrated for 4 min with 100% solvent A at a flow rate of 50 μl/min.

2.7 FLUORESCENCE POLARIZATION

Fluorescence Polarization (FP) is a technique that can be used to determine the binding affinity between two macromolecules. The technique is based on the observation that when a fluorescently labeled macromolecule is excited with polarized light, the degree of polarization of the emitted light varies with the size of the macromolecule. Large macromolecules tumble slowly in solution, and so their orientation at the moment of emission closely resembles that at the moment of excitation, resulting in strong polarization of the emitted light. In contrast, small macromolecules tumble more quickly, and so their orientation at emission is randomized relative to that at excitation, resulting in a low degree of polarization. In a system of two interacting proteins in which the smaller protein (ligand) is fluorescently labeled, the degree of polarization observed for the isolated

ligand will increase upon addition of the binding partner, as this causes the tumbling rate of the fluorescent probe to decrease (Figure 2.5 A). Consequently, the degree of polarization is low for the unbound state and high for the bound state.

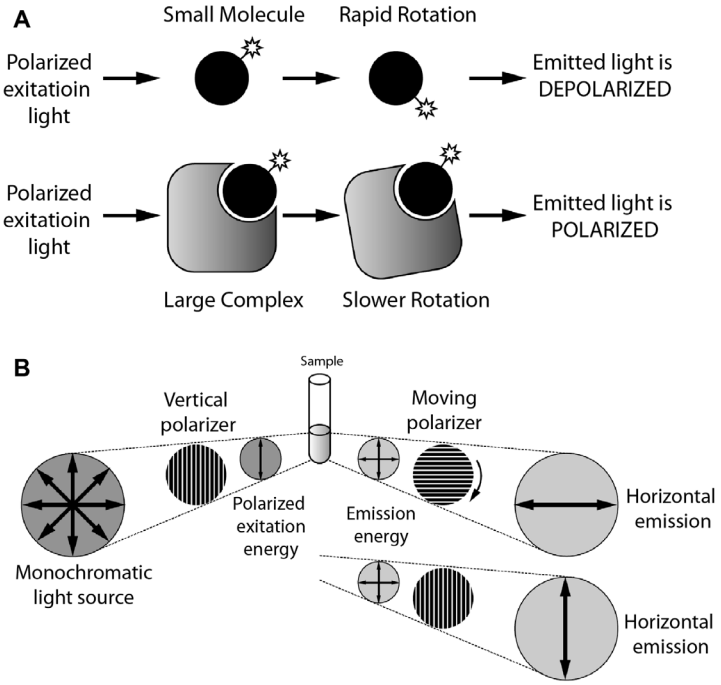


Figure 2.5 Schematic representation of FP principle and detection. A) Small fluorescent molecules rotate rapidly during the excited state, resulting in low polarization values. Large fluorescent molecules, caused by the binding of a second molecule, rotate slower and yield high polarization values. **B**) Monochromatic light passes through a polarizing filter and excites fluorescent molecules in the sample tube. Only molecules with vertical orientation can absorb light, get excited and emit light. Emitted light is measured both in the horizontal and vertical planes. Figure taken from XXXX.

Fluorescent polarization detection is schematically described in (Figure 2.5 (panel B). The fluorescent probe is excited by a monochromatic wavelength passing through a vertical polarization filter. Only probes oriented in the vertical plane can absorb light, get excited and subsequently emit light. Emitted light is measured both in the vertical ($P_{\parallel}$) and horizontal ($P_{\perp}$) plane. The value of anisotropy (A) is calculated as in equation (1).

$$A = \frac{P_{\parallel} - P_{\perp}}{P_{\parallel} + P_{\perp}} \quad (1)$$

When the ligand is not bound the value of A will be low, whereas when the ligand is bound A will increase until reaching a maximal value when 100% of the ligand is bound. Anisotropy can be correlated to the protein-ligand binding curve and used to estimate an apparent dissociation constant of the complex (K_D). The protein-ligand interaction at equilibrium is expressed as in equation (2):

$$\frac{(PL)}{P_0} = \frac{L_0 - PL}{K_D + (L_0 - PL)} \quad (2)$$

where PL is the concentration of protein-ligand complex, P_0 is the total protein concentration, L_0 is the total ligand concentration and K_D is the dissociation constant.

Anisotropy is related to the concentration of bound ligand by equation (3):

$$\frac{(PL)}{L_0} = \frac{A - A_0}{A_1 - A_0} \quad (3)$$

where A_0 and A_1 are the anisotropy values when all the ligand is unbound and bound, respectively. Combining (2) and (3) yields equation (4):

$$A = A_0 + (A_1 + A_0)\frac{\left(K_D + L_0 + P_0 - \sqrt{(K_D + L_0 + P_0)^2 - 4L_0P_0}\right)}{2L_0} \quad (4)$$

Thus, by measuring the value of A at different values of total protein concentration and a fixed ligand concentration, one can use the above equation to estimate K_D.

During this study FP was used to calculate the dissociation constant between CRM1 (WT or mutants) and Rev-NES or PKI-NES peptides, which were chemically synthesized with a carboxytetramethylrhodamine (TAMRA) moiety at the N-terminus (Peptide Specialty Laboratories). Peptides at 10 nM concentration were incubated for 1 h with varying concentrations of WT or mutant CRM1 in 20 mM Tris-HCl pH 7.5, 130 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA and 0.05 % β-octylglucoside. Anisotropy was measured at 25°C in a Tecan Safire2 fluorimeter. Excitation and emission wavelengths were 530 nm and 580 nm, respectively, and the slit width was 5 nm in both cases. Binding curves were fit using equation (4).

2.8 CELL CULTURE ASSAY AND DATA ANALYSIS

HeLa cells were grown on Dulbecco's modified Eagle's medium (*GIBCO, Life technologies*) supplemented with 10% fetal bovine serum (*Hyclone*) on poly-D-lysine (*Sigma*) coated 12-mm diameter glass coverslips for at least 24 h before transfection. Cells were transfected with 7 μg of the plasmid pSG5-CRM1-GFP (containing CRM1 wild-type or point mutation/deletion forms listed in Table 2.1) using Lipofectamine 2000 reagent (*InVitrogen*) according to the manufacturer's protocols. Cells were fixed 48h post-transfection with 2% paraformaldehyde in PBS for 20 min at 37°C and permeabilized with 0.2% Triton X-100 in PBS for 3 min before being mounted with VECTASHIELD Mounting Medium (*Vector laboratories*). Images were acquired with an inverted Olympus IX81 epifluorescence motorized microscope equipped with a motorized piezo stage (*Ludl Electronic Products, USA*) and a Retiga-SRV CCD camera (*QImaging*) driven by VOLOCITY software (*Improvision*) with a binning of 1, using a PlanApo 60x NA1.42 objective (*Olympus*). The nuclear compartment, nuclear rim and cytoplasmic compartment were defined and the corresponding fluorescence intensity was quantified using VOLOCITY image analysis software on a total of 30 cells for each construct.

3. RESULTS PART I

INVESTIGATING THE AUTOINHIBITORY FUNCTION OF THE ACIDIC LOOP OF CRM1

ABSTRACT

This chapter summarizes my contribution to a larger project in which I was partially involved during the PhD. This work aimed to confirm the role of the acidic loop of CRM1 in the allosteric regulation for the cargo binding. When the cargo is bound to CRM1 the loop adopts an extended conformation whereas in the unbound form it is in a retracted conformation. In order to confirm the auto inhibitory role of the loop a series of CRM1 mutants were designed *in silico*, predicted to enhance or reduce the affinity of CRM1 for the cargo by stabilizing either the extended or retracted conformation of the loop. The affinity of these CRM1 mutants for the cargo was tested both *in vitro*, using fluorescence polarization (FP) assay, as well as *in vivo*, by cellular localization of GFP-CRM1 in HeLa cells. The results demonstrate that mutants predicted to stabilize the extended form of the loop have increased affinity for the cargo *in vitro* and are sequestered in the cytosol. On the contrary CRM1 mutants that stabilize the retracted conformation of the loop show a reduced affinity for the cargo *in vivo* and are localized in the nucleus.

RÉSUMÉ

Ce chapitre résume ma contribution à un projet plus vaste dans lequel j'ai participé partiellement pendant mon doctorat. Ce travail visait à confirmer le rôle de la boucle acide de CRM1 dans la régulation allostérique de l'interaction avec le cargo. Lorsque le cargo est lié à CRM1, la boucle adopte une conformation étendue alors que sous la forme non liée, elle se trouve dans une conformation rétractée. Afin de confirmer le rôle d'auto-inhibition de la boucle, une série de mutants de CRM1 ont été conçus *en silico*, prédits pour augmenter ou réduire l'affinité de CRM1 pour le cargo en stabilisant soit la conformation étendue ou rétractée de la boucle. L'affinité de ces mutants pour le cargo a été testée à la fois *in vitro*, en utilisant un test de polarisation de fluorescence (FP), ainsi que *in vivo*, par localisation cellulaire de GFP-CRM1 dans des cellules HeLa. Les résultats démontrent que les mutants de CRM1 conçus pour stabiliser la forme étendue de la boucle ont une affinité accrue pour le cargo *in vitro* et sont séquestrés dans le cytosol des cellules HeLa. Au contraire, les mutants de CRM1 qui stabilisent la conformation rétractée de la boucle montrent une affinité réduite pour le cargo *in vivo* et sont localisés dans le noyau des cellules HeLa.

3.1 BACKGROUND

This thesis primarily reports the work which was the focus of the majority of my PhD studies, namely, the investigation of Rev and export proteins bound to RRE species by EMSA and native MS analysis. However, before describing these results I wish to dedicate this "mini-chapter" to a secondary project I worked on during my PhD. This project was the extension of a study I was involved in during my Master's thesis (performed in the same lab), which investigated the function of the C-terminal helix in CRM1 and gave rise to a publication in the journal *Structure* (Dian et al., 2013). The new project investigated the function of the acidic loop of CRM1 and was a collaborative effort with three other members of the lab, Florent Bernaudat, Cyril Dian and Fabienne Hans. I will give only a brief outline of this project here, and focus primarily on the part to which I contributed most significantly, namely, the *in vitro* binding affinity and cellular localization studies.

As mentioned in the Introduction, CRM1 is composed of 21 HEAT repeats that define a ring-shaped solenoid. HEAT repeat 9 contains an acidic loop which has been shown to act as a Ran-responsive allosteric regulator of the NES-binding groove (although at the time we initiated this project its function was not entirely clear). In unbound CRM1, the loop adopts a retracted conformation that interacts with the B helices of HEAT repeats 10-12, located behind the A helices which form the NES-binding groove (Figure 3.1 A). These interactions stabilize the groove in a constricted conformation which is incompatible with NES binding. The binding of Ran^{GTP} to CRM1 induces the loop to adopt a more extended conformation, stabilized by the interactions with Ran^{GTP} and by contacts with the B helices of repeats 12-15. This conformational change causes the NES-binding groove to widen, increasing its affinity for the NES peptide. Thus, the acidic loop inhibits cargo binding by stabilizing repeats 11 and 12 in a conformation antagonistic to NES binding, and Ran relieves this inhibition by pulling away the loop.

3.2 MUTATIONS OF CRM1 ENGINEERED TO MODULATE NES-BINDING AFFINITY

Our project aimed to confirm that the acidic loop acted as an "allosteric switch" that regulates NES-binding affinity. Such an allosteric switch had been previously hypothesized based on how RanBP1 promotes export complex disassembly (Koyama and Matsuura, 2010). Notably, the authors of that study made point mutations in three acidic loop residues (^{430}VLV^{432}) that were found to enhance NES binding activity in the absence of Ran. To confirm the role of the acidic loop we designed additional CRM1 mutants that were expected to stabilize the loop either in the retracted or extended conformation and therefore modulate the affinity of CRM1 for the cargo. If the loop is auto-inhibitory, then mutations in the B helices of repeats 10-12 which destabilize the retracted loop conformation should behave like the VLV mutation and enhance NES binding. To this end, we introduced either a double point mutation (M583A/K590A) on helix 12B or an arginine substitution (N495R) on helix 10B (Figure 3.1 B, C). By the same reasoning, deleting the entire loop should enhance NES binding even more than the above point mutations. We therefore constructed a loop-truncation mutant (ΔLoop) in which 23 loop residues were replaced by a 10-residue flexible linker (Figure 3.1 B). Finally, if the extended and retracted loop conformations co-exist in dynamic equilibrium, then destabilizing the former should reduce NES binding. We therefore introduced two charge reversal mutations on helices 14B and 15B (K693E/R753E) designed to destabilize the extended loop (Figure 3.1 D).

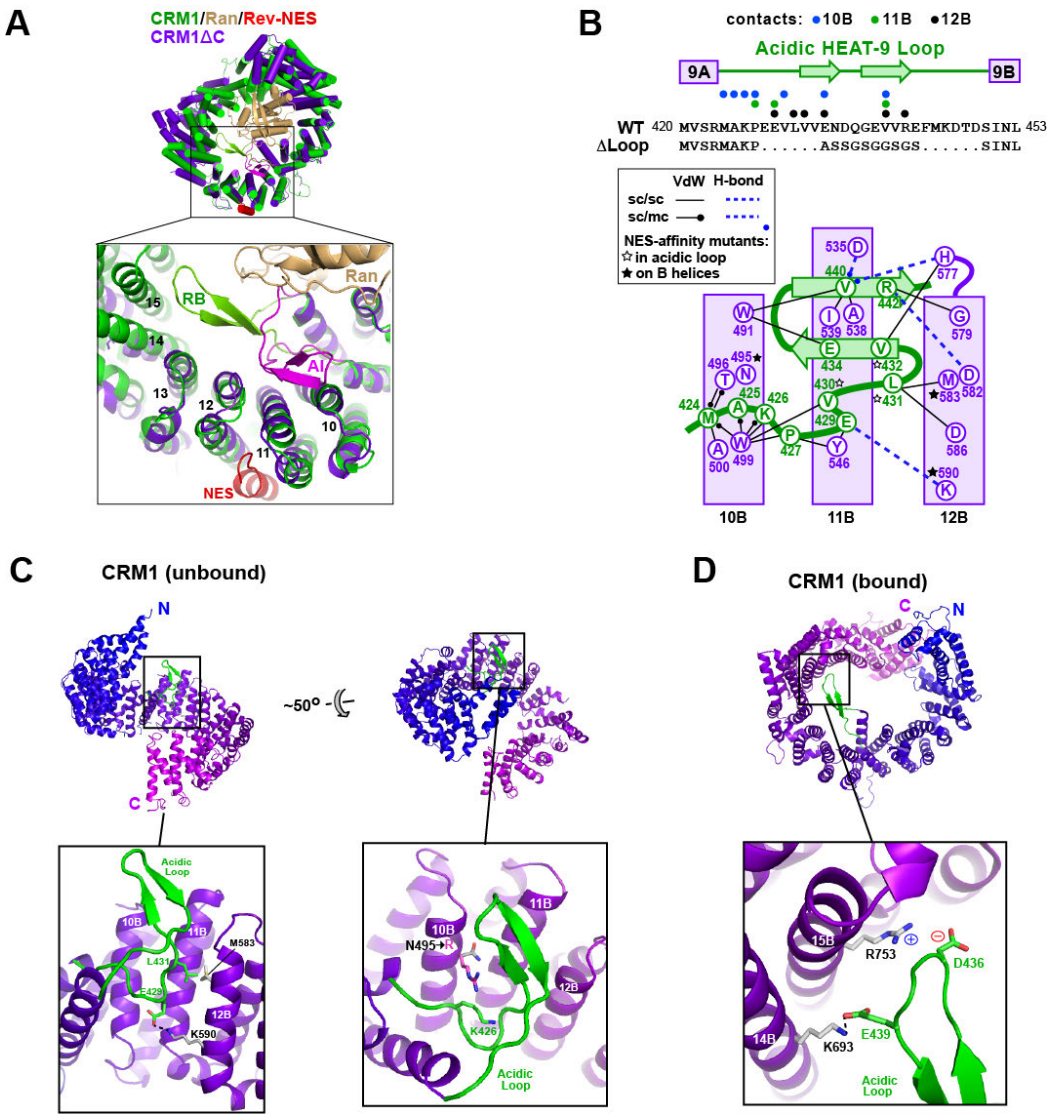
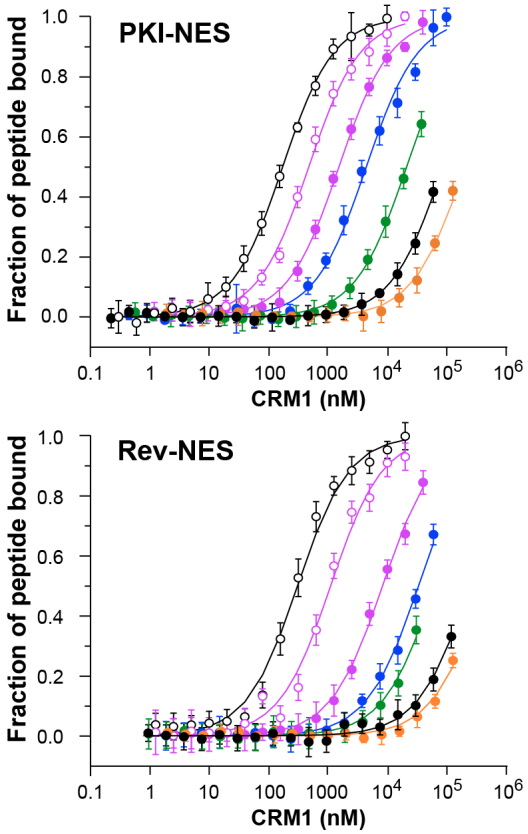


Figure 3.1: Conformation of the acidic loop and design of mutations. A) Structural alignment of unbound CRM1 (PDB ID 4BSM) with the CRM1/Ran/Rev-NES complex (PDB ID 3NBZ). The extended Ran-binding (RB) and retracted auto-inhibitory (AI) conformations of the acidic loop are shown. B) Summary of loop-mediated contacts with HEAT repeats that define the NES-binding groove. Solid and dashed lines indicate van der Waals contacts and H bonds, respectively; lines with or without a dot indicate main chain (mc) and side chain (sc) interactions. Asterisks indicate mutated residues. The sequence of linker residues used in the ΔLoop mutant is indicated. C) Mutations which destabilize the retracted loop conformation. *Left,* Residues Lys590 and Met583 on helix 12B interact with loop residues Glu429 and Leu431 through a salt bridge and hydrophobic contact, respectively. Alanine substitution of these residues is predicted to eliminate these interactions. *Right*, Replacement of Asn495 on helix 10B by an arginine is predicted to introduce steric repulsion with the base of the loop and electrostatic repulsion with Lys426. D) Mutations which destabilize the extended loop conformation. In the CRM1/Ran/NES complex, residues Lys693 and Arg753 on helices 14B and 15B mediate electrostatic interactions with loop residues Glu439 and Asp436. Glutamate replacement of the two basic residues are predicted to abolish these interactions.

3.3 *IN VITRO* BINDING AFFINITY OF CRM1 MUTATIONS

To assess the effect of the above mutations on NES-binding affinity, we used a fluorescence polarization (FP) assay in which the ability of wild-type (WT) or mutant CRM1 to bind a fluorescently labelled PKI-NES or Rev-NES peptide was determined. As expected, WT CRM1 bound these peptides with low affinity (K_d values of ~90 and ~250 μM for PKI- and Rev-NES, respectively; Figure 3.2, black curves with filled circles) and with greatly enhanced affinity upon the addition of Ran^{GTP} (K_d of 0.15 and 0.3 μM; Figure 3.2, black curves with open circles). In the absence of Ran^{GTP}, the M583A/K590A mutant bound both NES peptides approximately 4 times better than WT, while the N495R mutant bound 7 (Rev-NES) or 18 (PKI-NES) times better than WT (Figure 3.2, green curve), confirming our prediction that these mutations would enhance NES-binding affinity by destabilizing the retracted loop conformation. Also consistent with our predictions, the ΔLoop mutant bound the two NES motifs 30 or 60 times better than WT and significantly better than the N495R and M583A/K590A mutants (Figure 3.2, magenta curve with filled circles). Interestingly, the addition of Ran^{GTP} was less effective in enhancing the NES-binding affinity for the ΔLoop mutant than for WT CRM1 (Figure 3.2, compare magenta and black curves with open circles). This makes sense, because deletion of the loop is expected to considerably decrease the affinity of CRM1 for Ran^{GTP}. Finally, the K693E/R753E mutant designed to destabilize the extended loop conformation led to a two-fold decrease in NES-binding affinity (Figure 3.2, orange curves), supporting the notion that Ran-induced conversion to the extended conformation relieves loop-mediated inhibition. Taken together, these results strongly support the hypothesis that the retracted and extended conformations of the acidic loop correspond to low- and high-affinity NES-binding states of CRM1.



NES sequence

Rev	DAVPLQLPPLERLTLD
PKI	SGNSNELALKLAGLDINKTE

	Kd (μM)	
	PKI-NES	**Rev-NES**
WT+Ran	0.15 ± 0.03	0.30 ± 0.05
ΔLoop+Ran	0.48 ± 0.11	1.1 ± 0.3
ΔLoop	1.5 ± 0.3	8.3 ± 2.1
N495R	4.9 ± 0.5	34 ± 5
M583A/K590A	21.4 ± 2.2	64 ± 9
WT	87 ± 6	250 ± 35
K693E/R753E	190 ± 40	400 ± 60

Figure 3.2: Fluoresence polarization results. *Left,* Binding curves determined for WT and mutant proteins incubated with fluorescent Rev- or PKI-NES peptide with or without 20 μM RanGTP. *Right,* Peptide sequences and summary of K_d values. The mean and standard deviations from three independent experiments are shown. Destabilizing the retracted acidic loop conformation (green and blue curves) or deleting the loop (magenta) enhanced binding, whereas destabilizing the elongated conformation (orange) reduced affinity.

3.3. EFFECT OF CRM1 MUTATIONS ON CELLULAR LOCALIZATION

Previous work by our collaborator Maarten Fornerod (Erasmus MC, Rotterdam) led to the discovery of so-called Supra-physiological NES (supraNES) motifs – unusually strong NESs that bind stably to CRM1 (K_d in the low μM range) without the need for RanGTP (Engelsma et al., 2004). Interestingly, it was shown that cargos bearing such supraNES motifs can sequester CRM1 in the cytosol (Engelsma et al., 2004). Because some of our CRM1 point mutants also bind NES peptides in the low μM range, we wondered whether they might also have an effect on the cellular localization of CRM1.

To verify this hypothesis we transfected HeLa cells with GFP-tagged WT or mutant CRM1 proteins and quantified the GFP signal in the nucleus, cytosol and on the nuclear rim. The detailed results are shown for mutant N495R in Figures 3 A and B, and results for all mutants are summarized in Figure 3.3 C. (The ΔLoop mutant was cytotoxic and hence was excluded from our analysis). Compared to WT, the N495R mutant showed similar nuclear rim localization relative to total cellular signal (Figure 3 B, top panel; similar results were observed for the other mutants). In contrast, the N495R mutant was visibly depleted from the nucleus and gave a small but measurable increase in cytosolic localization relative to WT (Figure 3 B, middle and bottom panels). This yielded a significantly reduced nuclear/cytoplasmic ratio (from 1.01 for WT to 0.63 for the mutant; Figure 3.3 C). A similar effect was observed for the ^{430}VLV^{432} mutant previously shown to enhance Ran-independent binding (Koyama and Matsuura, 2010). The observed effect presumably reflects RanGTP-independent binding of the mutant to cytosolic cargos, providing indirect evidence that the allosteric mechanism by which the acidic loop regulates NES binding is operative *in vivo*. In support of this hypothesis, the M583A/K590A mutant also showed a reduced nuclear/cytoplasmic ratio; however, the reduction was less significant compared to the N495R mutant, consistent with the higher NES-binding affinity of the N495R mutant observed in our FP assays (Figure 3.2).

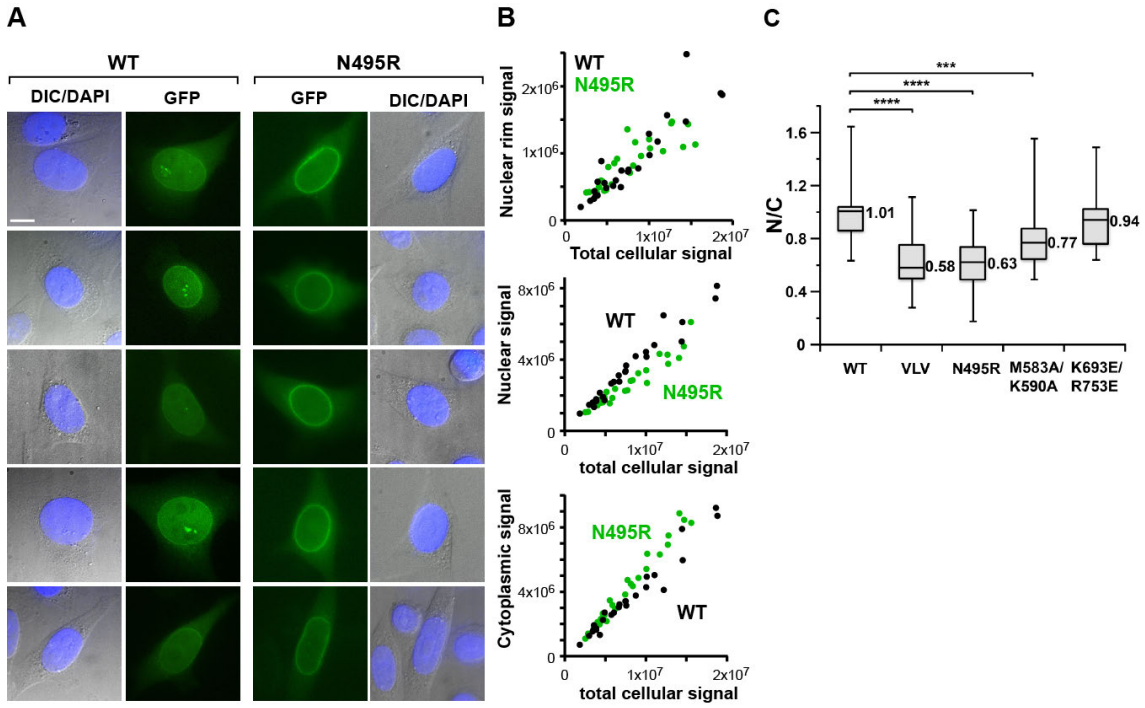


Figure 3.3: A mutation that enhances NES binding affinity retains CRM1 in the cytosol. A) Image gallery of HeLa cells transfected with GFP-tagged WT CRM1 or the N495R mutant and imaged by differential interference contrast (DIC) and fluorescence microscopy. DIC and DAPI-fluorescence images are overlayed. All images are at the same magnification. Scale bar, 10 µm (top left image). B) Total nuclear (left) and (c) cytoplasmic (right) GFP signals from individual cells plotted against total cellular GFP signal reveals decreased nuclear and increased cytoplasmic localization of the N495R mutant. Vertical scale is in arbitrary intensity units. C) Box plots showing nuclear/cytoplasmic ratio for different mutants. Median values are indicated. (n=30 per experiment; ***, $p < 10^{-4}$; ****, $p < 10^{-5}$ in a Mann-Whitney test).

In conclusion, the above *in vitro* and cell-based studies confirmed that the acidic loop of CRM1 is an allosteric regulator of NES-binding affinity that operates by switching between two conformations mediating distinct interactions with the inner surface of the HEAT-repeat solenoid.

4. RESULTS PART II
INVESTIGATING THE STOICHIOMETRY OF THE RRE-REV EXPORT COMPLEX

ABSTRACT

This study aims to investigate the architecture of the RRE/Rev/CRM1/Ran complex by using native mass spectrometry (MS), a powerful method to determine the stoichiometry of macromolecular complexes. I set up protocols for the large-scale preparation of the full length RRE and a 66-nucleotide RRE fragment (IIABC) and adapted these for compatibility with native MS analysis. Using a mutant form of Rev (Rev*) I set up and optimize the experimental conditions to analyze IIABC/Rev* and IIABC/Rev*/CRM1/Ran RNPs by native mass spectrometry and obtain high quality data. Adapting the experimental conditions I also worked with the wild type form of Rev, obtaining high quality spectra showing that IIABC binds 3 copies of Rev, as well as the ability of the protein to bridge distant molecules of IIABC. The analysis of the IIABC/Rev/CRM1/Ran complex suggested that at least 3 copies of Rev are needed to recruit 2 molecules of CRM1 on the IIABC. Additional efforts at analyzing the entire RRE and complexes with wild-type Rev have also yielded informative spectra, while analysis of the intact RRE/Rev/CRM1/Ran complex has had more limited success.

RÉSUMÉ

Cette étude vise à étudier l'architecture du complexe RRE/Rev/CRM1/Ran en utilisant la spectrométrie de masse (MS) native, une méthode puissante pour déterminer la stœchiométrie des complexes macromoléculaires. J'ai mi en place des protocoles pour la préparation à grande échelle du RRE entier et d'un fragment de 66 nucleotides (IIABC) et je les ai adaptés pour être compatible avec l'analyse de MS native. En utilisant une forme mutante de Rev (Rev*), j'ai mis en place et optimisé les conditions expérimentales pour analyser des RNPs (IIABC/Rev* et IIABC/Rev*/CRM1/Ran) par MS native et j'ai obtenu des données de haute qualité. En adaptant les conditions expérimentales, j'ai travaillé avec la forme de type sauvage de Rev. J'ai obtenu des spectres de haute qualité, montrant que IIABC lie 3 copies de Rev, ainsi que la capacité de la protéine à interagir avec plusieurs molécules de IIABC. L'analyse du complexe IIABC/Rev/CRM1/Ran a suggéré qu'au moins 3 copies de Rev sont nécessaires pour recruter sur le IIABC 2 molécules de CRM1. L'analyse du RRE entier et des complexes avec Rev ont également générés des spectres informatifs, tandis que l'analyse du complexe RRE/Rev/CRM1/Ran entier a connu un succès plus limité.

4.1 OVERALL STRATEGY

As mentioned in the Introduction, we realized that the goal of determining the stoichiometry of the Rev-RRE-CRM1-Ran^{GTP} complex – referred below to as the "full RNP export complex" – was technically formidable. To overcome the anticipated obstacles, we decided on a "divide-and-conquer" strategy, which involved dissecting the full RNP export complex into smaller complexes that would be easier to assemble and characterize (Figure 4.1). To this end we used three different RRE molecules: a 34-nt fragment corresponding to IIB; a 66 nt fragment corresponding to IIA, IIB, IIC (designated IIABC); and the full length RRE (350 nt). Furthermore, knowing that Rev was difficult to work with because of its tendency to aggregate in solution, we used three different Rev proteins: a highly soluble, monomeric mutant; a less soluble, dimeric mutant; and the wild-type protein. All proteins were overexpressed in *E. coli*, whereas the RNAs (except for IIB which was commercially synthesized) were *in vitro* transcribed and then extensively purified.

Electrophoretic mobility shift assays (EMSAs) were used to investigate and optimize the *in vitro* assembly of the different RNPs used in this study. Complexes were then assembled in a manner compatible with native MS analysis. As shown in Figure 4.1 we worked at different levels of complexity:

- First level of complexity: different RNA/Rev complexes were investigated. The IIB is expected to accommodate only one molecule of Rev and was used as a control for the RNA binding activity of Rev. IIABC and RRE are expected to accommodate more than one Rev molecule and therefore we first aimed to determine the stoichiometry of these complexes.

- Second level of complexity: The three different RNA/Rev complexes were used to assemble export RNPs by the addition of CRM1 and Ran^{GTP}. The **minimal RNP complex** contains IIB, anticipated to accommodate only one Rev molecule and therefore to recruit only one CRM1/Ran^{GTP} complex. This minimal complex was useful for optimizing conditions for complex assembly and as a quality control for protein functionality. The **small RNP export complex** and **full RNP export complex** contain IIABC and the full RRE, respectively. Depending on the number of Rev molecules assembled on IIABC or RRE, we aimed to determine the stoichiometry of CRM1/Ran^{GTP} recruited by Rev.

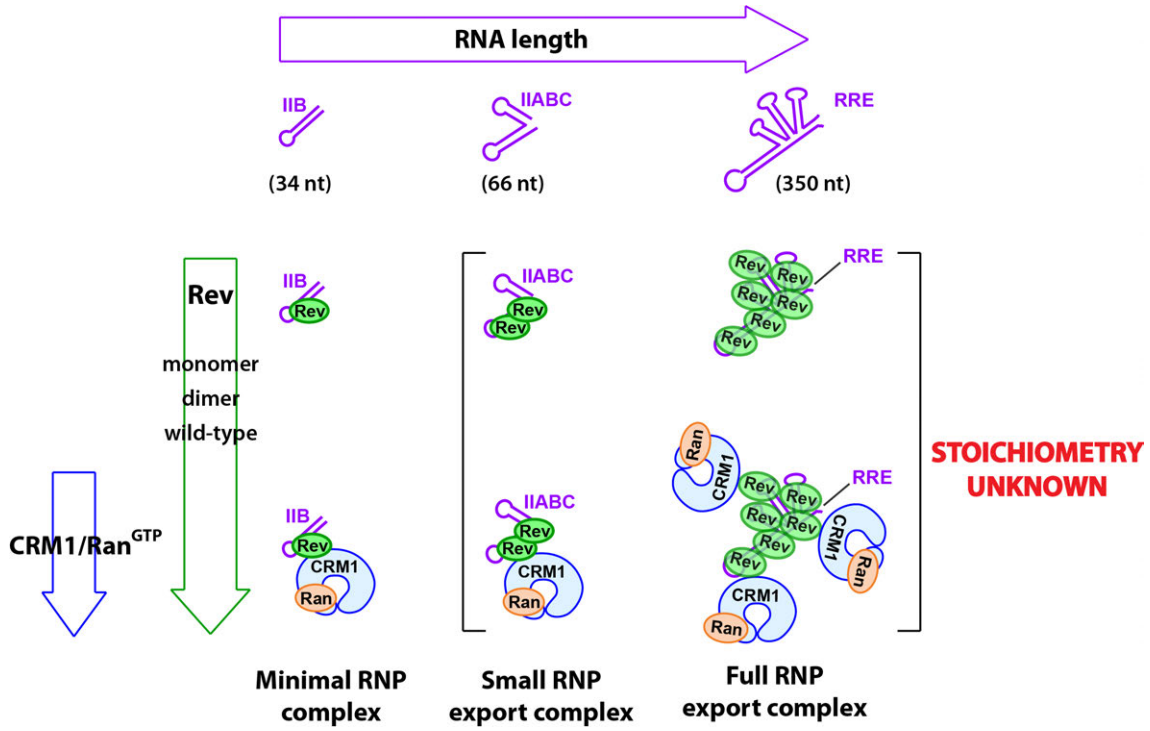


Figure 4.1. Strategy used to study the stoichiometry of the HIV-1 RNP export complex. Three different RRE fragments (in magenta) were used to assemble HIV-1 RNPs in a stepwise manner. Initially RNA/Rev complexes were assembled using three different forms of Rev (wild-type and monomeric and dimeric mutants) and investigated by native mass spectrometry. These complexes were then used to assemble CRM1/Ran^{GTP}-containing RNPs (the small RNP export complex and full RNP export complex) to determine their stoichiometry.

4.2 *IN VITRO* ASSEMBLY OF THE HIV-1 RNP EXPORT COMPLEX

As explained above, the full HIV-1 RNP export complex was decomposed into smaller complexes to facilitate its study. In this section the experimental conditions for the *in vitro* assembly of these complexes will be presented.

As CRM1 and Ran production and purification were already established in the laboratory (See chapter 2, § 2.2.2 and § 2.2.3), my main tasks consisted of setting up the production and purification of Rev protein, the *in vitro* transcription of IIABC and the RRE, and the assembly of RNP complexes. Throughout this study electrophoretic mobility shift assays (EMSAs) were used extensively to evaluate the *in vitro* assembly of different complexes. Nevertheless, the analysis of such complexes was not an easy task. Rev multimerization on the RNA resulted in the formation of many different species differing

from one another by the presence of a single Rev molecule that were often not sufficiently resolved on the gel. The complexity increased even further when CRM1 and Ran^{GTP} were added to the mix. This was a source of considerable uncertainty for several experiments, which in many cases required me to proceed by a trial-and-error approach.

4.2.1 Production and purification of Rev

4.2.1.1 General purification strategy for Rev

Rev is a difficult protein to work with because of its high affinity for nucleic acids and its tendency to multimerize and form large aggregates. The approach I used to circumvent these problems was adapted from the strategy used to solve one of the first crystal structures of Rev (Daugherty et al., 2010a) as well as a more recent structure of the protein in complex with an RRE fragment (Jayaraman et al., 2014). In this approach the solubility and monodispersity of Rev is enhanced by its fusion to the B1 domain of a streptococcal protein G (GB1). The acidic domain of GB1 is thought to interact with the basic ARM of Rev, mimicking the presence of RNA and thus reducing its association with nucleic acids and with other copies of itself. Therefore, for this study the Rev sequence was N-terminally fused to a hexahistidine tag followed by the GB1 sequence (His-GB1 tag) and a TEV protease cleavage site (Figure 4.2 A).

In addition to the wild-type protein (Rev^{WT}), we also produced two oligomerization defective mutants, based on functional and structural data (Jain and Belasco, 2001; Thomas et al., 1998; Daugherty et al., 2010b; DiMattia et al., 2010). One contains four point mutations (L12S/L18Q/I59D/L60D, hereafter denoted **Rev***) that should impair both 'H/H' and 'T/T' interfaces, favoring a monomeric form of Rev. The second mutant contains a single substitution (V16D; hereafter denoted $\mathbf{Rev^{V16D}}$) that should favor a dimeric form of Rev by destabilizing only the 'H/H' interface (Figure 4.2 A).

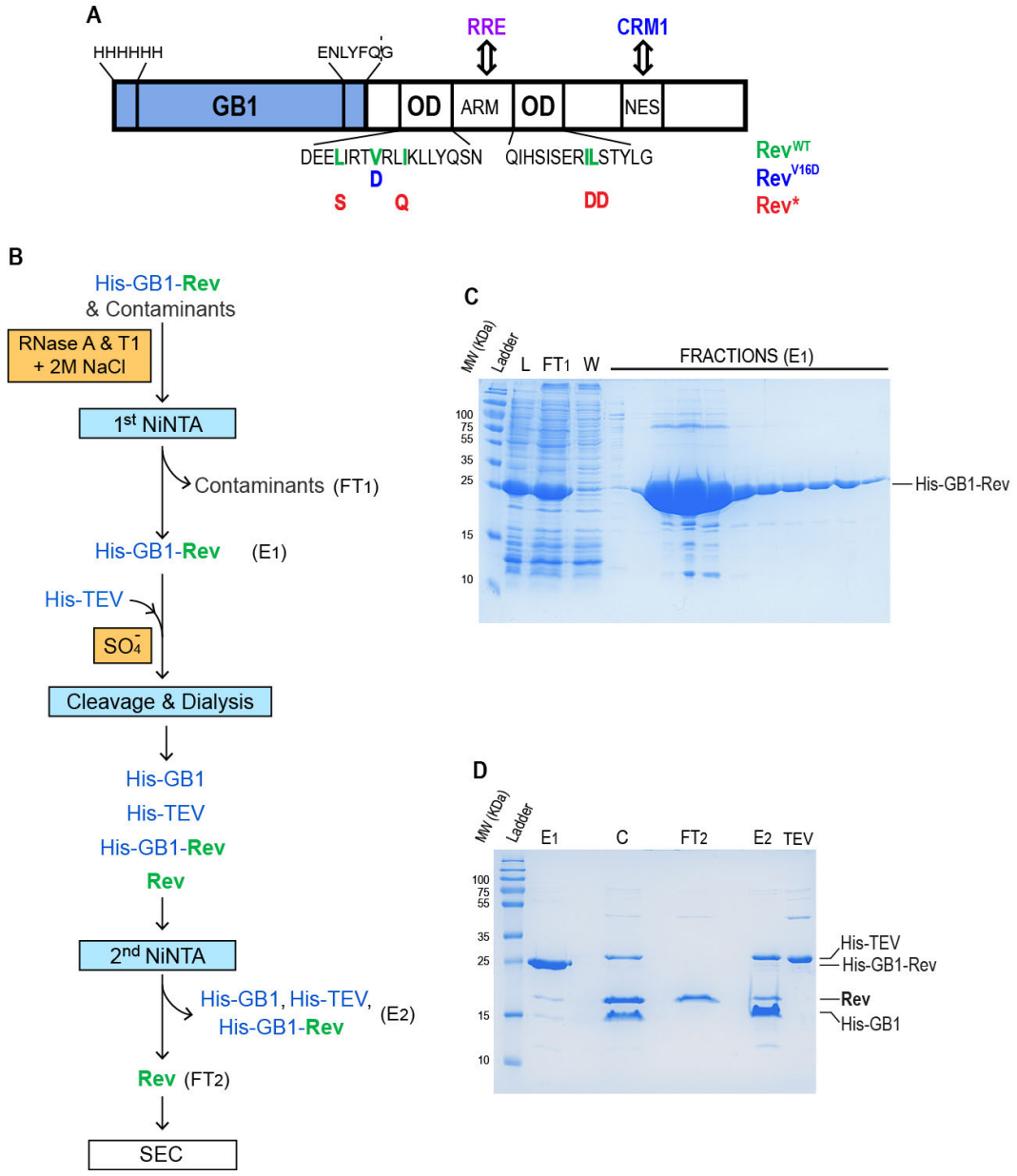


Figure 4.2: General purification strategy for Rev. A) His-GB1-Rev construct. The His tag and TEV cleavage site sequences are shown above the box. Sequences within the "oligomerization domain" (OD) that differ between WT and mutant forms of Rev are indicated below the box, with point mutations indicated in blue (Rev^{V16D}) or red (Rev*). B) Protocol used to purify Rev. C) Denaturing SDS-PAGE (14% polyacrylamide) analysis of the first NiNTA affinity step showing the initial sample loaded onto the NiNTA column (L), the flow through (FT_1), the sample recovered during the wash step (10% buffer B) (W) and the eluted fractions (E_1). D) Denaturing SDS-PAGE analysis of the TEV cleavage and second NiNTA affinity steps. Samples analysed were as follows: E_1, the eluted fractions pooled after the first NiNTA step; **C**, the cleaved and dialysed sample loaded onto the second NiNTA-affinity column; FT_2, the flow through of the second NiNTA step; E_2, elution of the second NiNTA step; TEV protease used for cleavage. Both gels correspond to early purification steps of Rev^{WT}.

The protocol used to purify the different forms of Rev is explained in detail in chapter 2 section 2.2.4 and summarized in Figure 4.2 B. Briefly, the cleared lysate was incubated with RNAse A and T1 in the presence of 2 M NaCl to remove nucleic acid contamination. Wild-type or mutant His-GB1-Rev was retained on a Ni-NTA column whereas contaminants and nucleic acids were removed (Figure 4.2 C). TEV protease was added to cleave the His-GB1 tag, while dialysing against a buffer containing a high concentration of sulfate ions to keep Rev soluble. His-GB1-Rev, His-GB1 and the his-tagged TEV protease were removed with a second Ni-NTA affinity column, while the cleaved and untagged Rev was collected in the flow through (Figure 4.2 D). Rev was then purified by size exclusion chromatography (SEC) at a high degree of purity.

4.2.1.2 Purification of Rev oligomerization defective mutants

Rev* and Rev^{V16D} were predicted to be primarily monomeric and dimeric in solution, respectively. Therefore, the SEC purification step for these mutants was performed using a buffer with a reduced amount of salts, in order to remove the sulphates that would hamper subsequent biochemical characterization (i.e. EMSA). As expected, the two mutants eluted at different volumes (Figure 4.3).

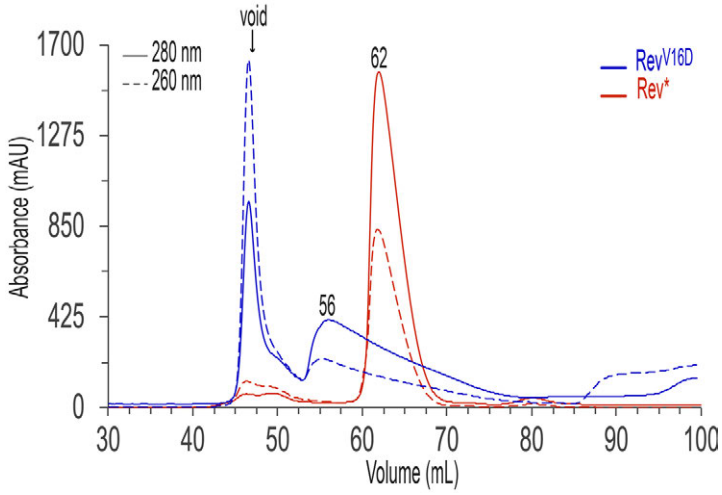


Figure 4.3: SEC elution profile of Rev* and Rev^{V16D} on a Superdex75 16/60. SEC was performed in a 40 mM Tris/HCl pH 8, 0.2 M NaCl, 1 mM DTT. The arrow indicates the void volume.

The SEC profile of Rev* was predominated by a sharp peak centred at 62 mL, with only minor nucleic acid contamination (detected as small peaks at or near the void volume, in which the absorbance at 260 nm exceeded that at 280 nm), indicating that the sample was highly homogeneous and monodisperse (Figure 4.3). The Rev^{V16D} profile was characterised by a much larger peak at the void volume, corresponding to nucleic acid contaminants, and a second broad and asymmetric peak with its maximum height at 56 mL, which contains Rev^{V16D} (Figure 4.3). The strong presence of residual nucleic acids suggests that the RNAse treatment step was less efficient for Rev^{V16D} than for Rev*. The shape and position of the second peak suggest that the oligomeric state of Rev^{V16D} is higher than that of Rev*. I was able to subsequently concentrate both mutants to high concentration without the formation of any visible precipitate. In conclusion, I was able to purify both the Rev* and Rev^{V16D} mutants at the high yield and homogeneity required for subsequent experiments.

4.2.1.3 Purification of Rev wild-type

In initial experiments, Rev^{WT} was purified following the general purification protocol described above in section 4.2.1.1. Interestingly, during the TEV cleavage step we were unable to obtain 100% cleavage of Rev^{WT} and in the subsequent NiNTA step we were unable to completely separate the cleaved and uncleaved forms (data not shown). Therefore, we loaded a mixture of both forms of Rev^{WT} onto the SEC column. The SEC profile was characterized by a sharp peak at the void volume, corresponding to nucleic-acid-bound aggregates, and a very broad peak spanning from 50 to ~70 mL (Figure 4.4), suggesting the presence of multiple oligomeric states of Rev. Eluted fractions contained both His-GB1-Rev^{WT} and Rev^{WT}, as expected.

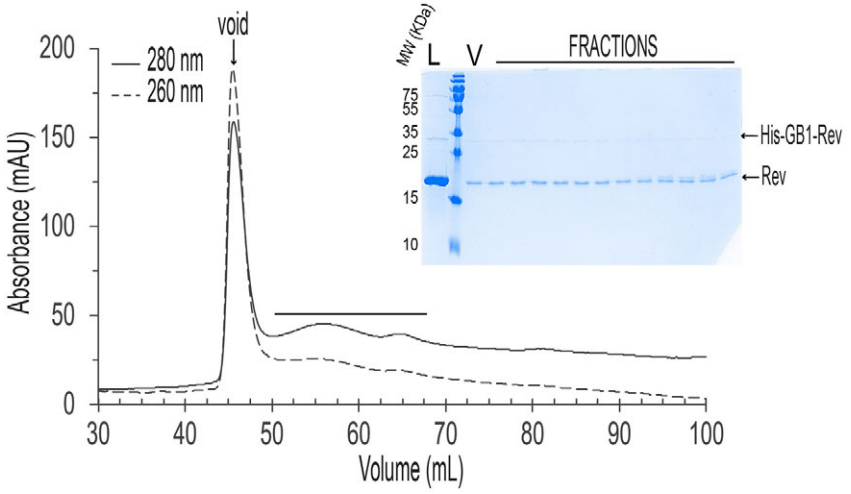


Figure 4.4: SEC elution profile of Rev^{WT} : first attempt of purification. SEC was performed on a Superdex 75 16/60 column in 40 mM Tris/HCl pH 8, 0.2 M NaCl, 1 mM DTT. The inset shows 14% SDS-PAGE analysis of: the loaded sample (L), a representative fraction at the void volume (V), and fractions eluted between 50 and 68 mL (indicated on the chromatogram by a black line).

To obtain a more homogenous Rev^{WT} sample, two changes were made to the protocol: 1) the ratio of TEV to His-GB1-Rev^{WT} was increased to enhance cleavage efficiency; and 2) a sulphate-containing buffer was used during SEC to prevent the self-association of Rev^{WT} in solution. These changes yielded a great improvement: the SEC profile now exhibited a well-defined (though still relatively broad) Rev^{WT} peak at 51.5 mL, suggesting that Rev^{WT} was present in solution as a larger oligomer than Rev^{V16D} (Figure 4.5).

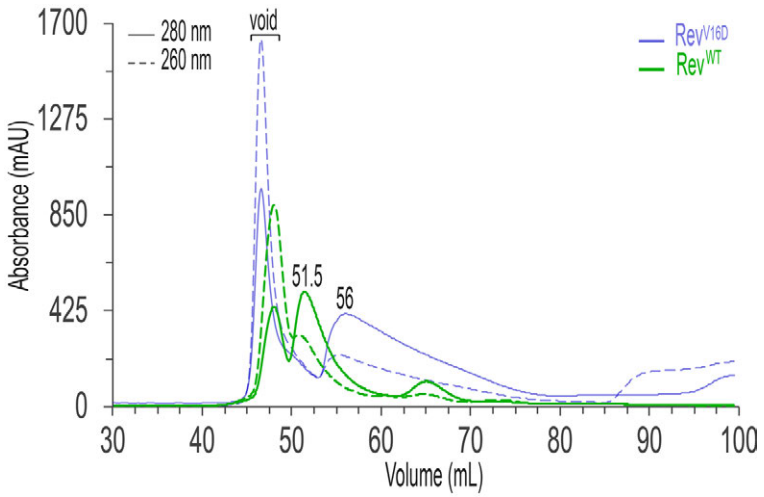


Figure 4.5: Comparison of SEC profiles for Rev^{WT} and Rev^{V16D} on a Superdex75 16/60. SEC was performed in 40 mM Tris/HCl pH 8, 0.2 M NaCl, 0.1 M Na_2SO_4, 0.4 M $(NH_4)_2SO_4$, 1 mM DTT for Rev^{WT} and in 40 mM Tris/HCl pH 8, 0.2 M NaCl, 1 mM DTT for Rev^{V16D}.

Nevertheless, a large amount of aggregates containing nucleic acids was present and partially overlapped with the nucleic-acid-free Rev^{WT} elution peak (Figure 4.5, void). We supposed that residual nucleic acids in the sample favored Rev^{WT} aggregation. Therefore, two other steps of the purification protocol were refined: 1) RNAse treatment was extended by incubating the cleared lysate for 90' at room temperature; 2) SEC was performed on a Superdex 200 (rather than Superdex 75) column to improve the separation of Rev^{WT} from aggregates. These modifications yielded a greatly improved Rev^{WT} sample: the SEC elution profile was characterized by a single major peak at 63.8 mL, albeit asymmetric and relatively broad (Figure 4.6). We decided that the homogeneity of this sample was sufficient for subsequent biochemical and biophysical studies.

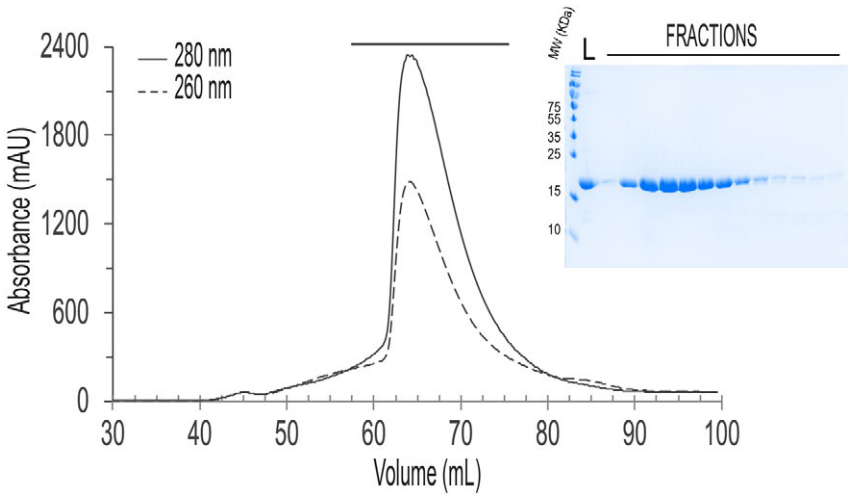


Figure 4.6: SEC profile of Rev^{WT} obtained with the optimized protocol. SEC was performed in 40 mM Tris/HCl pH 8, 0.2 M NaCl, 0.1 M Na_2SO_4, 0.4 M $(NH_4)_2SO_4$, 1 mM DTT on a Superdex 200 16/60 column. The inset shows 14% SDS-PAGE analysis of: the loaded sample (L) and fractions eluted between 58 and 80 mL (indicated on the chromatogram by a black line).

4.2.2 Production and purification of IIABC and RRE

In order to reconstitute viral RNP complexes *in vitro*, I used two different RNA molecules: the entire RRE (350 nt) sequence recognized by Rev and a 66 nt fragment (IIABC) corresponding to RRE stem structures IIA, IIB and IIC (Figure 4.7 A). These RNAs were transcribed *in vitro* and purified using distinct protocols, as summarized in Figure 4.7 and described in detail in chapter 2, § 2.3.1 and 2.3.2.

IIABC was purified under denaturing conditions using a Q-sepharose ion-exchange resin. To obtain a homogenous 3' end, a sequence coding for the Hammer Head Rybozyme (HHR) was inserted downstream of IIABC, resulting in cleavage of the full-length transcript between the two RNA moieties. The transcription reaction was loaded on the resin and the HHR and non-cleaved transcript were separated from IIABC using a salt gradient (Figure 4.7 B and C).

RRE was purified under native conditions (Chillon et al., 2015), including a step of purification by size exclusion chromatography (Figure 4.7 D). The first attempts of purification, starting from medium transcription reaction volumes (30 to 200 µL), yielded a homogeneous sample. However, during subsequent RRE preparations, we noticed variations in the degree of purity (Figure 4.7 E, lane I), which appeared to depend on the quality and freshness of the DNA template and RNA polymerase. We therefore resorted to

a classic technique: separation on denaturing gel followed by electroelution. As seen in Figure 4.7 E, this yielded pure and homogenous RRE.

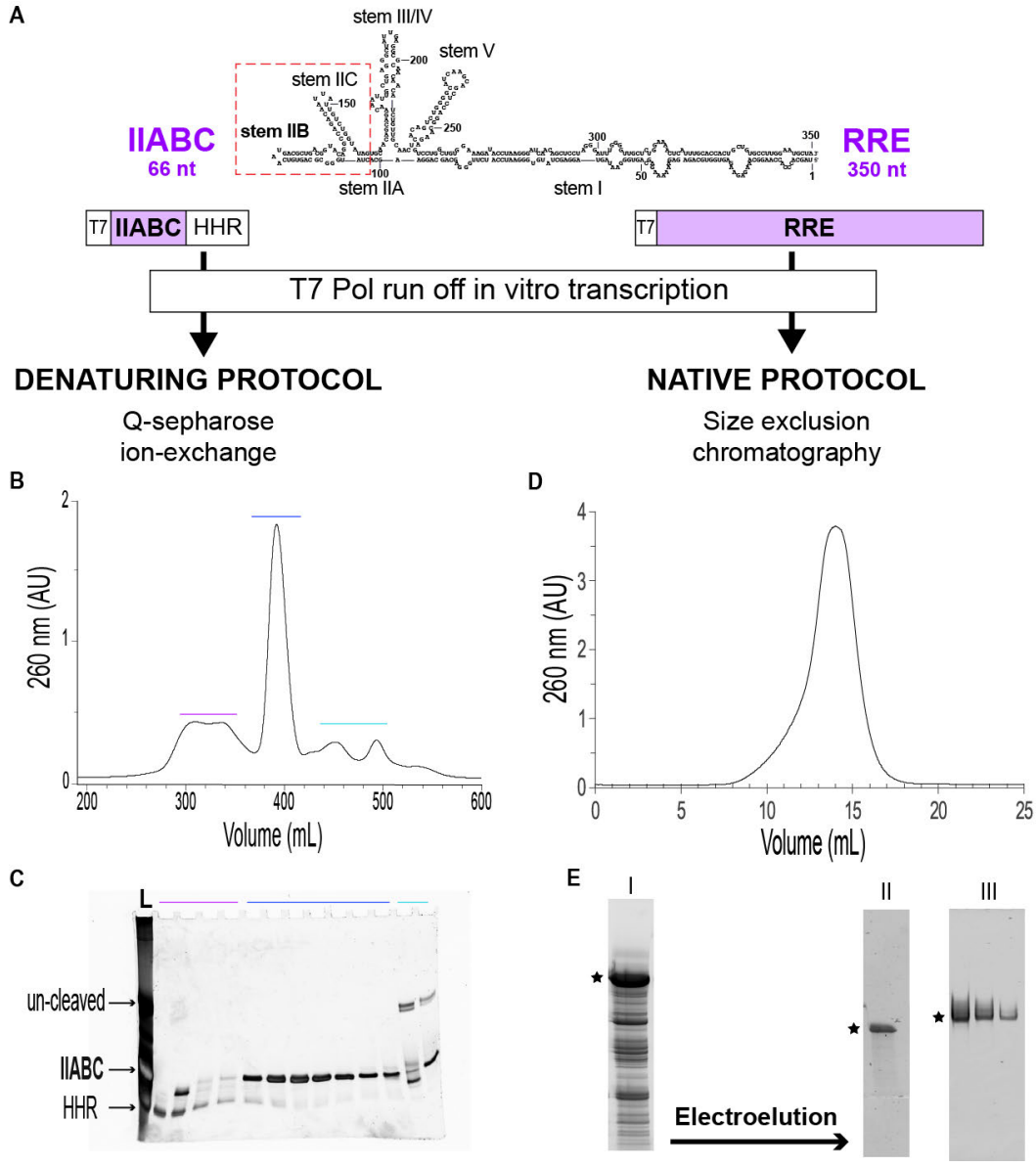


Figure 4.7: Overview of IIABC and RRE production and purification. A) Secondary structure of the RRE. The IIABC fragment is boxed in red. Schematic views of the linearized DNA highlight the position of the T7 promotor and HHR. B) Elution profile of IIABC purified by Q-sepharose ion-exchange. C) 10% Urea-PAGE analysis of the sample loaded onto the Q-sepharose column (L) and of eluted fractions. Magenta, blue and cyan lines correspond to the three peaks similarly indicated in B. D) SEC elution profile of *in vitro* transcribed RRE, performed on a Sephacryl 400 10/300 column. E) Fractions pooled from the SEC were loaded on a 10% urea-PAGE gel (I). After electroelution, the purity and homogeneity of the RRE were evaluated on 10% urea (II) and 6% TBE (different amounts loaded) (III) acrylamide gels. All gels were stained with GelRed.

4.2.3 Assembly of the minimal HIV-1 RNP quaternary complex

In initial studies we used IIB,the highest affinity-binding site of Rev on the RRE, to assemble a minimal HIV RNP quaternary complex. We first tested WT and mutant Rev proteins for their ability to recognize stem IIB. We then used the stem IIB/Rev^{WT} complex to build up the quaternary complex containing CRM1 and Ran^{GTP}.

4.2.3.1 Rev recognition of IIB

We tested the three Rev proteins (Rev^{WT}, Rev^{V16D} and Rev*) for the ability to recognize IIB. Increasing amounts of each protein were added to a constant amount of RNA and analyzed by EMSA. As shown in Figure 4.8, the migration of IIB was retarded upon addition of all three forms of Rev. For Rev* and Rev^{V16D}, we observed only a single shifted species, consistent with the formation of a 1:1 complex. In contrast, additional higher-migrating species were observed at high concentrations of Rev^{WT} (Figure 4.8 C, lanes 7 to 9), presumably due to protein aggregation in the low ionic strength buffer used for this assay.

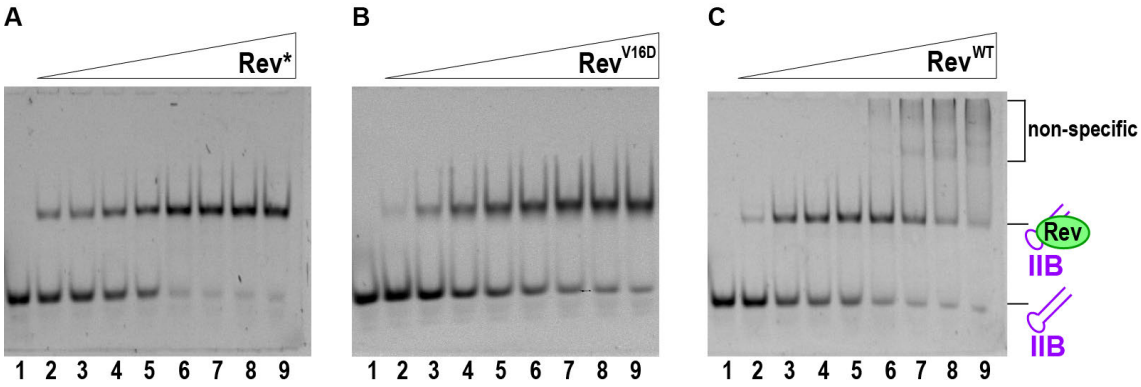


Figure 4.8: Binding of WT and mutant Rev to RRE stem IIB. IIB at 100 nM was incubated with increasing amounts of Rev* (A), Rev^{V16D} (B) or Rev^{WT} (C) in standard binding buffer. 5 µL of sample was loaded in each well: lane 1 contains unbound IIB, for lanes 2 to 9 the molar ratio of RNA to protein was 1:1, 1:2, 1:3, 1:4, 1:6, 1:8, 1:10 and 1:14 respectively.

4.2.3.2 CRM1 binds to a minimal Rev/IIB complex in presence of Ran^{GTP}

Using stem IIB and Rev^{WT} the formation of a minimal CRM1/Ran/Rev/RNA complex was analyzed by EMSA. The complex was assembled in a stepwise manner: stem IIB was first incubated with Rev^{WT} then CRM1 and Ran were added. As shown in Figure 4.9, IIB

was shifted as a consequence of complex formation. As observed before (Figure 4.8), addition of Rev^{WT} yielded the formation of a first complex (Figure 4.9, lane 2) that remained un-shifted upon the addition of CRM1 in the absence of Ran (Figure 4.9, lane 3). In the presence of Ran, the IIB/Rev^{WT} complex was further retarded, yielding a band corresponding to the quaternary complex (Figure 4.9, lanes 4-5). Furthermore, the quaternary complex disassembled upon addition of leptomycin B (Figure 4.9, LMB, lanes 6-7), a specific inhibitor of CRM1 that competes with the cargo by covalently reacting with the NES-binding groove.

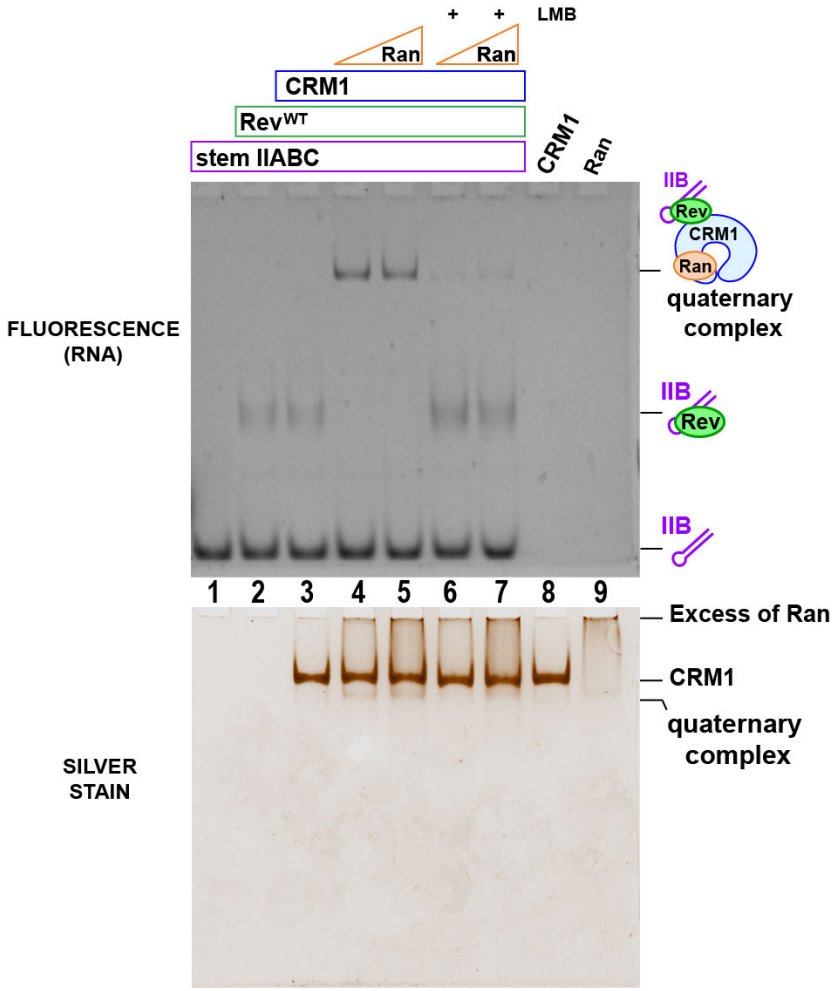


Figure 4.9: Assembly of a minimal HIV RNP complex. Migrating species were detected by fluorescence excitation of the TAMRA fluorophore, covalently bound to IIB (top), and by silver staining (bottom). Samples were prepared in standard binding buffer. IIB at 100 nM (lane 1) was incubated with Rev^{WT} in a ratio of 1:4 (lanes 2 to 7) for 30 min. Then CRM1 was added (in a Rev^{WT}:CRM1 ratio of 2:1) in lanes 3 to 7 together with Ran and samples were incubated for an additional 30 min. The CRM1:Ran ratio was 1:1 (lanes 4 and 6) or 5:1 (lanes 5 and 7). Leptomycin B (LMB) was added and samples were incubated for a further 20 min (lanes 6-7). Lanes 8 and 9 contain CRM1 and Ran, respectively.

A quaternary complex containing IIB, Rev^{WT}, CRM1 and Ran^{GTP} was successfully assembled and analyzed by EMSA, demonstrating that Rev^{WT} activity is fully retained *in vitro*. As expected, the interaction of Rev^{WT} with CRM1 was Ran^{GTP}-dependent and sensitive to LMB, confirming that the interaction occurred through the NES binding groove.

4.2.4 Assembly of the small RNP export complex

Our ability to assemble the minimal HIV RNP export complex *in vitro* confirmed that our Rev, CRM1 and Ran proteins were functionally active. We therefore used the 66 nt RRE fragment IIABC to build up and analyze a larger HIV RNP export complex. This RNA was expected to accommodate at least two molecules of Rev, as it contains the IIB sequence as well as a second known binding site for Rev in the adjacent region. We first analyzed the interaction of our three Rev proteins with IIABC and then tested some of these complexes for their ability to recruit CRM1 and Ran^{GTP}.

4.2.4.1 Assembly of Rev/IIABC complexes

To study the Rev-IIABC interaction, we performed EMSA experiments in which we incubated IIABC with increasing amounts of Rev*, Rev^{V16D} or Rev^{WT}. We observed complex formation in all cases, as indicated by discrete bands of reduced electrophoretic mobility (Figure 4.10). Interestingly, the number of complexes observed was protein-dependent: four discrete bands were detected for Rev* (Figure 4.10 A), whereas only three were observed for Rev^{V16D} and Rev^{WT} (Figure 4.10 B and C). Note that the relative amounts of RNA and protein were optimized to best visualize the full set of complexes on a single gel. We assume that the first band shift (complex I) corresponds to the binding of a single Rev molecule to IIABC and each subsequent shift (complexes II-IV) to that of an additional Rev molecule. This is the simplest working hypothesis. However, we had doubts regarding complex II of Rev^{WT} and Rev^{V16D}, as this band did not appear when experiments were performed using higher Rev and IIABC concentrations; for example, see Figure 4.11 B and C where this band is missing. Thus, we cannot exclude that complexes II and III of Rev^{WT} and Rev^{V16D} might represent alternate conformations of the same species.

In all cases, complex I became detectable at the lowest Rev concentration tested. In contrast, the ratio of Rev to IIABC required to form higher complexes increased with the

number of Rev point mutations. For Rev^{WT}, complexes II and III were clearly detected at a [Rev]/[IIABC] ratio of 3.4 (Figure 4.10 C, lane 6), whereas a similar intensity for the corresponding Rev^{V16D} complexes was only observed at a ratio of 8.1 (Figure 4.10 B, lane 9). For Rev*, complexes II, III and IV were detected at ratios of 9.1, 12.7 and 17.9, respectively (Figure 4.10 A, lanes 10-12). These findings are consistent with the partial or complete loss of binding cooperativity expected for Rev^{V16D} and Rev* due to the destabilization of one or both dimer interfaces, respectively. We also observed additional bands for Rev^{WT} at high protein concentrations (Figure 4.10 C, lanes 13-14). We attribute these to the binding of additional copies of Rev through Rev-Rev (as opposed to Rev-RNA) interactions, consistent with the tendency of Rev^{WT} to self-associate in solution.

Complex I migrated with approximately the same relative mobility for all three proteins (the mobility relative to unbound IIABC was 0.7), consistent with the first Rev monomer binding to the same site (IIB) in all three cases. Strikingly, the mobility of the higher complexes varied among Rev proteins: for both Rev^{WT} and Rev^{V16D} the relative mobility measured for complexes II and III was approximately 0.62 and 0.58, respectively, whereas for Rev* these values were 0.52 and 0.39 (i.e., the Rev* complexes migrated more slowly). This difference in mobility is not due to a difference in electrostatic charge, because Rev* is more acidic than Rev^{WT} and hence should migrate comparatively faster, not slower. Instead, the difference in mobility suggests a difference in conformation, whereby Rev* complexes II and III adopt a more open conformation that hampers migration, while the Rev^{WT} and Rev^{V16D} complexes are more compact and migrate faster. A difference in conformation would also explain the different maximal number of binding sites observed: the more open conformation allows IIABC to accommodate four Rev* molecules, whereas the more compact conformation only accommodates three molecules of Rev^{WT} or Rev^{V16D} (or only two molecules if complexes II and III are alternate conformations of the same species).

In conclusion, the above EMSA experiment was highly informative: it confirmed the degree of binding cooperativity expected for our three Rev proteins, revealed the number of Rev binding sites on IIABC, and suggested that at least two conformations (open and closed) are possible for Rev/IIABC complexes.

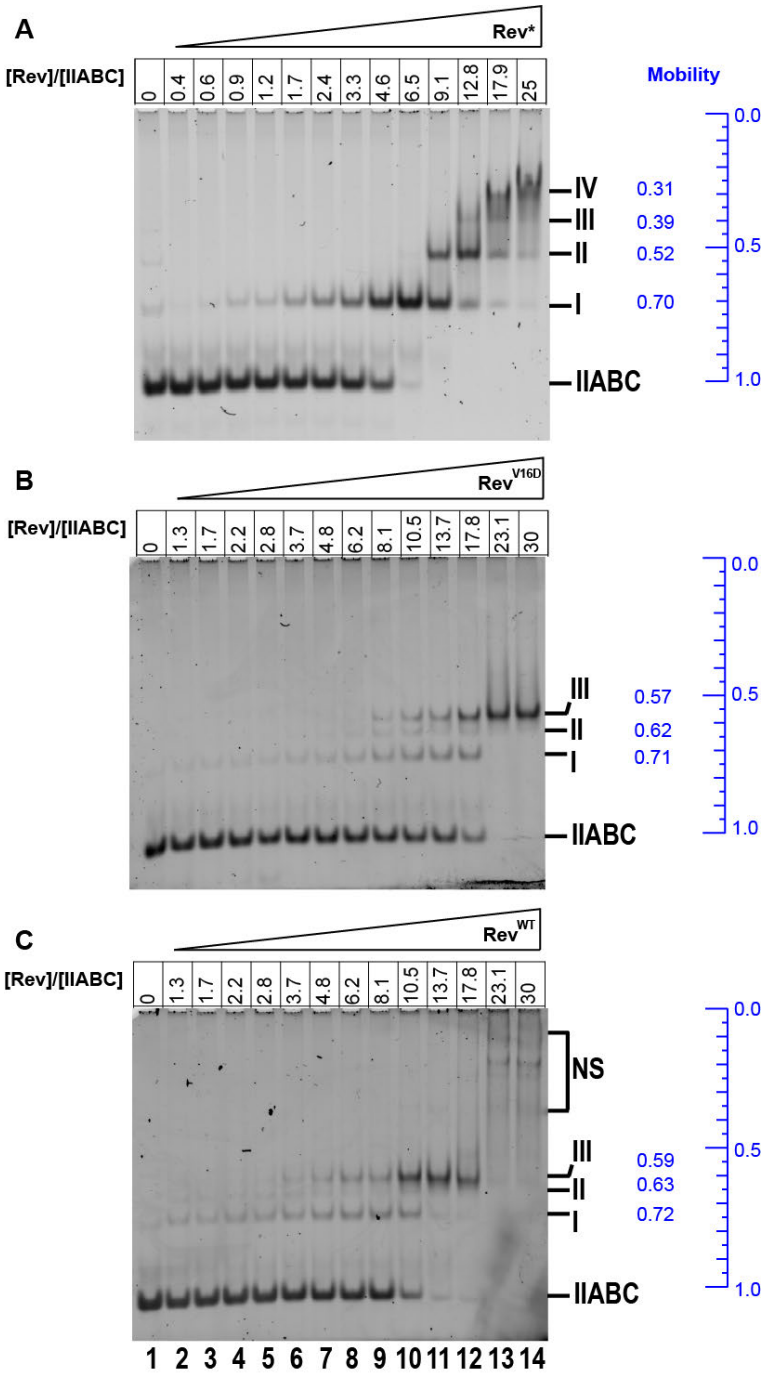


Figure 4.10: Assembly of IIABC/Rev complexes. EMSAs of IIABC in complex with Rev* (A), Rev^{V16D} (B) and Rev^{WT} (C). IIABC (0.3 µM in A, 0.15 µM in B and C) was incubated with increasing concentrations of Rev protein, previously serially diluted. The [Rev]/[IIABC] ratio is indicated above each lane. The maximal protein concentration (lane 14) was 7.5 µM for Rev* (dilution factor of 1.4) and 4.5 µM for both Rev^{V16D} and Rev^{WT} (dilution factor of 1.3).

4.2.4.2 CRM1 binds to the IIABC/Rev complexes in presence of Ran^{GTP}

As shown above IIABC can bind more than one copy of each Rev form. All these complexes can potentially recruit at least one molecule of CRM1 in the presence of RanGTP. With the aim of determining the stoichiometry of a small HIV RNP export complex by native MS, we first verified *in vitro* complex assembly by EMSA (in standard binding buffer). Complexes were assembled in a stepwise manner: first IIABC was incubated with Rev and then CRM1 and Ran were added.

At first we used the Rev* mutant to assemble a quaternary complex (Figure 4.11 A). Using IIABC at 1 μM (Figure 4.11 A lane 1), a 3-fold molar excess of Rev* was sufficient to assemble a 1:1 complex, with only a minor amount of the IIABC/$(Rev^*)_2$ complex (Figure 4.11 A, lane 2). The addition of CRM1 and Ran caused a supershift, yielding bands that corresponded to the quaternary complex (Figure 4.11 A, lanes 4-6). As samples were assembled with Rev* in molar excess of CRM1 throughout the experiment, the IIABC/Rev* complex was not completely supershifted. As expected, an excess of Ran relative to CRM1 was required to reduce the amount of free IIABC/Rev* and stabilize the quaternary complex, since the Ran protein was not fully in the GTP-bound form.

Small HIV RNP export complexes were then assembled using Rev^{V16D} and Rev^{WT}. First IIABC was incubated with Rev^{V16D} or Rev^{WT} in order to form the different species (Figure 4.11 B and C, lanes 2) and then CRM1 and Ran were added. The two IIABC/Rev^{V16D} complexes were supershifted into two complexes of reduced mobility, each one putatively containing one molecule of CRM1 and one of Ran^{GTP} (Figure 4.11 B, lane 3). Further addition of CRM1 and Ran caused the complete supershift of the IIABC/Rev^{V16D} bands and free CRM1 was detected (Figure 4.11 B, lanes 4-5). Similarly, IIABC/Rev^{WT} complexes were also supershifted by the addition of CRM1 and Ran: interestingly at least three different species of reduced mobility were detected, an intense band and two fainter bands (Figure 4.11, lane 4).

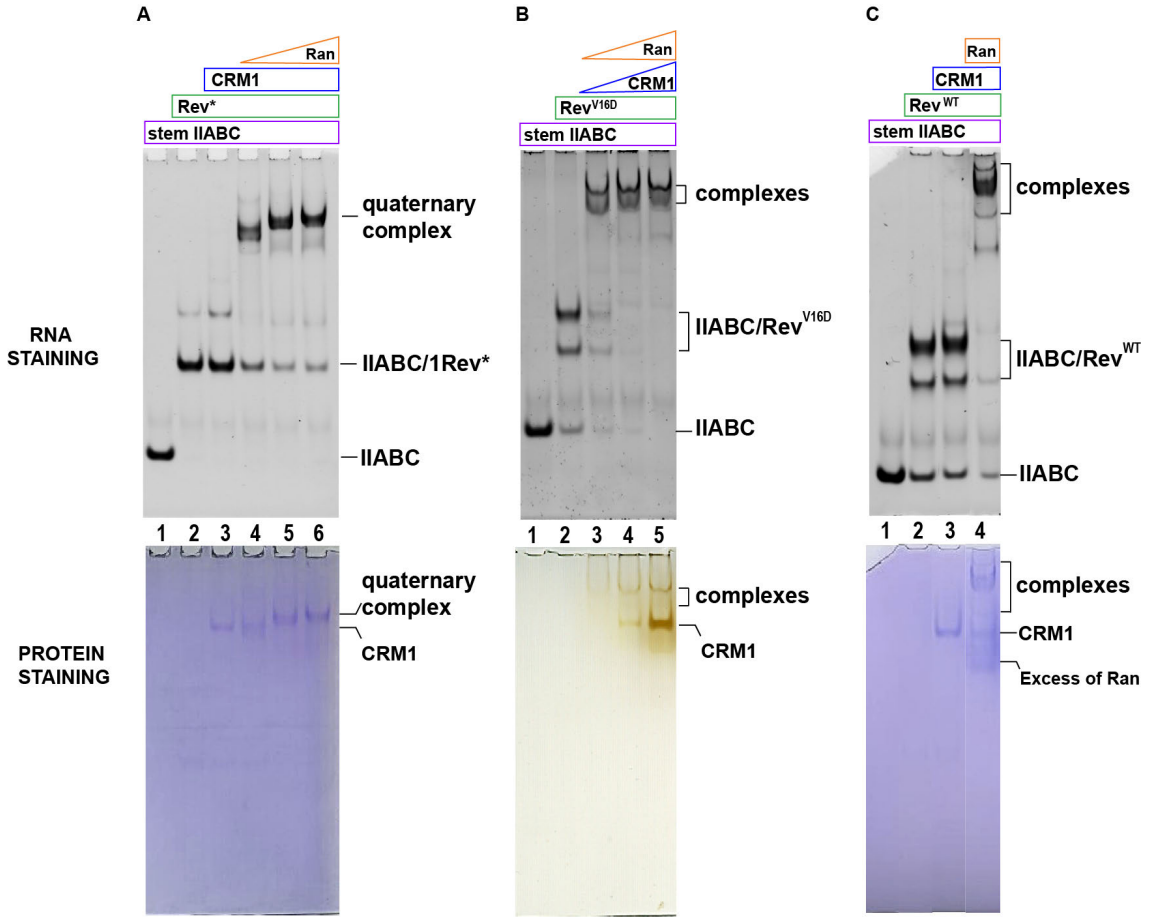


Figure 4.11: Small HIV RNP export complexes *in vitro* assembly. All samples were incubated in standard binding buffer in a stepwise manner: IIABC was first incubated with Rev for 30 min, then CRM1 and Ran were added and samples were incubated for 30 min more. Gels in A and C were stained with GelRed (top) and Coomassie blue (bottom); the gel in B was stained with SYBR gold (top) and silver (bottom). A) IIABC at 1 µM was incubated with Rev* in a ratio of 1:3. In lanes 3 to 6 the molar ratio of CRM1 to Rev* is 0.5. Ran was in excess of CRM1 by 5-, 10- and 15-fold in lanes 4, 5 and 6, respectively. B) IIABC at 0.15 µM was incubated with Rev^{V16D} in a ratio of 1:18. In lanes 3, 4 and 5 the molar ratio of CRM1 to Rev^{V16D} was 0.25, 0.5, and 1.0, while the CRM1:Ran molar ratio was 1:15. C) IIABC at 1 µM was incubated with Rev^{WT} in a molar ratio of 1:9. In lanes 3 and 4 the molar ratio of CRM1 to Rev^{WT} was 0.5, while that of Ran to CRM1 in lane 4 was 15.

These results show that all three IIABC/Rev complexes, assembled using different forms of Rev, were able to recruit CRM1 in the presence of Ran^{GTP}. Using Rev* the sample composition could be more easily controlled and a homogenous quaternary complex could be assembled. With Rev^{V16D} and Rev^{WT}, as expected, a mixture of at least 2 complex species was obtained.

4.3 MS ANALYSES OF THE SMALL HIV-1 RNP EXPORT COMPLEX

As shown above (§ 4.2.4), we can successfully assemble HIV-1 RNP complexes *in vitro* using a standard binding buffer. EMSA experiments were critical for optimizing assembly conditions, especially for determining the best relative component ratios to use to minimize heterogeneity of the desired complex. However the use of this assay to determine the stoichiometry of such complexes was severely limited by the resolution of the polyacrylamide gel. With the long-term goal of determining the stoichiometry of the fully assembled HIV-1 RNP export complex with high accuracy, we decided to turn towards native MS.

As underlined in chapter 2, native MS has advantages and limitations (§ 2.5). One advantage that we exploited during this study is the ability to analyse polydisperse samples. The ability of Rev to oligomerize on the RRE give rises to a number of complexes that may differ in mass by a single copy of this small protein (~13 kDa relative to a total mass for the complex of at least ~300 kDa). Such complexes cannot easily be discriminated or separated by biochemical techniques.

In contrast, a major limitation of native MS is that the sample must be prepared in a volatile buffer (such as ammonium acetate, AmAc). As previously shown, the different components of the HIV-1 RNP export complex were purified using common biochemical buffers that are incompatible with native MS (§ 2.2.2; 2.2.3; 4.2.1). One of the main technical challenges of my thesis work has therefore been to assemble complexes in a manner compatible with subsequent native MS analysis. The difficulty in overcoming this challenge primarily concerned the preparation of RNA. Very few studies of RNP complexes analyzed by native MS have been reported in the literature (Fadouloglou et al., 2015; Saliou et al., 2015), as the high affinity of RNA for non-volatile cations (such as Na^+, K^+, Mg^{2+}) makes analysis very challenging.

In a typical native MS experiment, the complex of interest is first assembled using standard buffers and then buffer exchanged using desalting columns or ultrafiltration devices into AmAc (up to 1 M) to remove non-volatile salts (Hernández & Robinson, 2007). However, this approach proved problematic for our complexes. Instead, we chose a different approach: the components of the HIV-1 RNP complex were independently buffer exchanged into AmAc and then assembled prior to native MS analysis. By following this

strategy we could first check the quality of each individual component and, if needed, optimize sample preparation. The complexes were directly assembled in a solution of AmAc whose concentration was chosen to match the ionic strength of the standard binding buffer used for EMSA, avoiding high concentrations that might destabilize the complex.

To study the full HIV-1 RNP export complex by native MS we followed the same strategy presented in section 4.2.4: first we analyzed RNA/Rev complexes and then we added CRM1 and Ran to build up higher-order RNP complexes.

This part of the experimental work was done in close collaboration with Dr. Elisabetta Boeri Erba, who was in charge of the acquisition, analysis and interpretation of native MS spectra.

4.3.1 Mass Spectrometry of proteins and protein complexes

4.3.1.1 Analysis of individual proteins by MS

Proteins to be used in native MS experiments were prepared as described in Experimental Procedures (§ 2.5.4). Buffer exchange into AmAc was performed by SEC, allowing us to remove any possible aggregates and to check sample monodispersity in the native MS buffer. This was critical especially for Rev^{WT}, whose oligomeric state and tendency to aggregate is sensitive to salt concentration. For convenience, each protein was buffer exchanged and stored in single-use aliquots that allowed us to more easily deal with multiple components when preparing samples for native MS analysis. The use of single-use aliquots had several advantages: it avoided a time-consuming buffer exchange procedure before each native MS experiment; it let us more quickly perform small tests to optimize sample preparation; and it allowed us to use the same batch of protein across many experiments.

Proteins were individually analyzed by denaturing MS in order to determine their intact masses and assess sample quality, as well as by native MS to determine their oligomeric state (Figure 4.12 A).

A

Protein	Theoretical Mass	Measured Mass ESI-TOF	Measured Mass nano-ESI-TOF
Rev*	13,168.6	13,167.9	13,167 ± 1
Rev^{V16D}	13,191.8	13,191.6	*Not acquired*
Rev^{WT}	13,175.9	13,175.7	13,175 ± 1
Ran	24,593.4	24,593.0	-
Ran^{GTP}	25,140.5	-	25,140 ± 1
Ran^{GDP}	25,060.5	-	25,061 ± 1
CRM1	125,028.0	124,897.6 (2-1085 aa) 124,566.4 (6-1085 aa)	124,565 ± 1

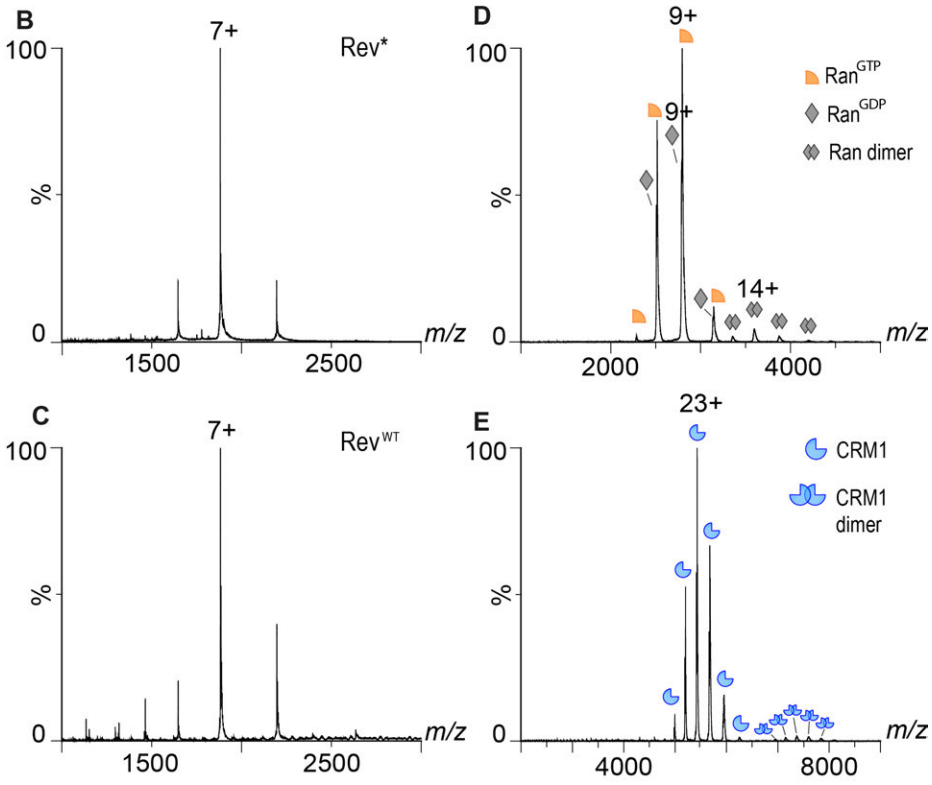


Figure 4.12: Masses of the individual proteins and their native MS spectra. All proteins were analysed in 200 mM AmAc, except CRM1 to which 1 mM DTT was added. A) Summary of masses expressed in Da. The theoretical mass was calculated from the primary sequence of each protein. The experimental masses measured by ESI-TOF in denaturing conditions and nano-ESI-TOF in native conditions are reported. For Ran the theoretical mass corresponding to the apo-protein and the GTP bound form are reported. For CRM1 two N-terminally truncated forms were detected. The one lacking the first residue (res. 2-1085) was the most abundant (75%). Spectra acquired for (B) Rev* and (C) Rev^{WT} at protein concentrations of 10 and 7 µM, respectively. D) Spectrum acquired for Ran at 5 µM concentration. Charge-states corresponding to Ran^{GTP}, Ran^{GDP} and a Ran dimer are labeled as indicated in the legend. E) Spectrum acquired for CRM1 at 15 µM concentration. Only the most intense charge-state is labeled in each spectrum.

Rev*: Native MS analysis showed that, as expected, the protein was detected as a monomer (Figure 4.12 B).

Rev^{WT}: This was the most difficult protein to analyze because of repeated clogging of the needle, probably due to the protein's tendency to aggregate. Nevertheless, it was detected as a monomer when acquired at 7 μM in 200 mM AmAc (Figure 4.12 C). This protein concentration is close to that estimated at which Rev^{WT} starts to oligomerize (Wingfield et al., 1991) which may explain why Rev was detected as a monomer and not an oligomer, as hydrophobic interactions are weaker in the gas phase than in solution.

Ran: Native MS analysis of Ran showed three charge-state distributions (Figure 4.12 D). These corresponded to Ran^{GDP} (25034.37 Da), Ran^{GTP} (25113.42 Da) and a Ran dimer (50070.09 Da). These data agree with native gel analysis (Figure 2.2 C), which also revealed three forms of Ran. We estimated that ~25-30% of Ran was bound to GTP and therefore active for the assembly of the HIV RNP export complexes. This explained why a ~4-5 fold excess of Ran was required to assemble and stabilize complexes in our EMSA experiments (§ 4.2.4.2).

CRM1: This protein was difficult to analyze because of frequent needle clogging (possibly because of protein unfolding and oxidation during native MS acquisition, as CRM1 is highly conformational flexible and lacks a significant hydrophobic core). We could improve the situation by adding DTT (~1 mM) to the CRM1 buffer during acquisition. CRM1 was detected as a monomer (Figure 4.12 E) present as a mixture of two truncated forms, as previously observed by MS ESI-TOF analysis (Figure 4.12 A). Less intense charge-state distributions corresponding to unfolded protein (2000-4000 m/z) (data not shown) and a CRM1 dimer (centred at ~7500 m/z) were also detected.

In conclusion, buffer exchange by SEC is ideal for preparing proteins for native MS analysis. For all proteins native MS analysis indicated the presence of monomers. Acquiring native MS spectra for individual proteins proved to be crucial for interpreting more complicated spectra obtained from RNP complexes, as shown below.

4.3.1.2 Rev/CRM1/Ran^{GTP} complex analysis by native MS

We next decided to analyze the Rev/CRM1/Ran^{GTP} complex in the absence of RNA, reasoning that information gained about the assembly of this complex and its behavior by native MS would help us for later analysis of the entire RNP export complex.

Given that the analysis of Rev^{WT} by native MS was difficult, we decided to first work with Rev*, which is monomeric in solution. The four point mutations of Rev* are located far from the NES motif and hence should not compromise binding to CRM1. Rev* was incubated with CRM1 in the presence of an excess of Ran^{GTP} and a native MS spectrum was acquired. A charge-state distribution corresponding to the intact Rev*/CRM1/Ran^{GTP} complex was detected (Figure 4.13 A, B). Signals corresponding to the binary CRM1/Ran^{GTP} complex, free CRM1 and Ran were also detected, due to the molar excess of these proteins compared to Rev* in the sample.

A

Label	Stoichiometry				Measured	Calculated
	Rev*	Rev^{WT}	CRM1	Ran^{GTP}	Mass ± error	Mass
	0	0	1	1	149,709 ± 2	149,706.9
a	1	0	1	1	162,876 ± 2	162,874.8
b	0	1	1	1	162,883 ± 3	162,822.6

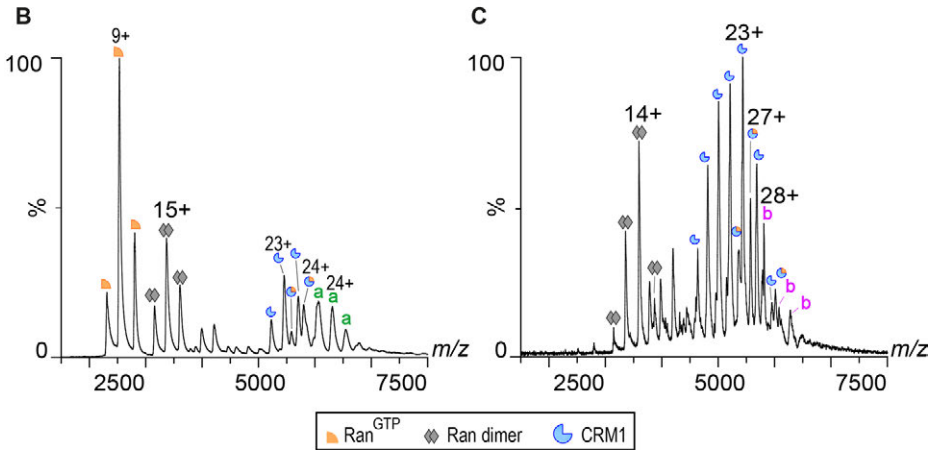


Figure 4.13: Native MS analysis of Rev/CRM1/Ran^{GTP} protein complexes. Complexes were prepared and spectra acquired in 200 mM AmAc, 1 mM Mg Acetate, 1 mM DTT. A) Summary of complexes detected by native MS. the measured masses of intact complexes in native conditions (measured mass) and their stoichiometry are shown. Calculated masses were obtained by summing the individual protein masses determined by ESI-TOF denaturing MS. Labels a and b refer to the corresponding ion series shown in B and C. B,C) Native MS spectra. Ion distributions corresponding to free CRM1, Ran and the CRM1/Ran^{GTP} complex are labeled as in the legend. B) 8 µM of Rev* was incubated with CRM1 and Ran in a molar ratio of 1:2:6. The Rev*/ CRM1/Ran^{GTP} complex is labeled 'a'. C) 10 µM of Rev^{WT} was incubated with CRM1 and Ran in a ratio of 1:1:5. The Rev^{WT} /CRM1/ Ran^{GTP} complex is labeled 'b'.

In similar conditions Rev^{WT} was incubated with CRM1 and Ran and analyzed by native MS: a charge-state distribution corresponding to the binary CRM1/Ran^{GTP} complex was clearly present, as well as signals corresponding to free CRM1 and Ran (Figure 4.13 C). The intact complex corresponding to Rev^{WT}/CRM1/Ran^{GTP} was also detected as a low intensity charge-state distribution (labeled with b in Figure 4.13 C). Other ratios were tested to increase the signal intensity of the trimer but without success For instance, increasing the Rev^{WT} concentration caused sample precipitation.

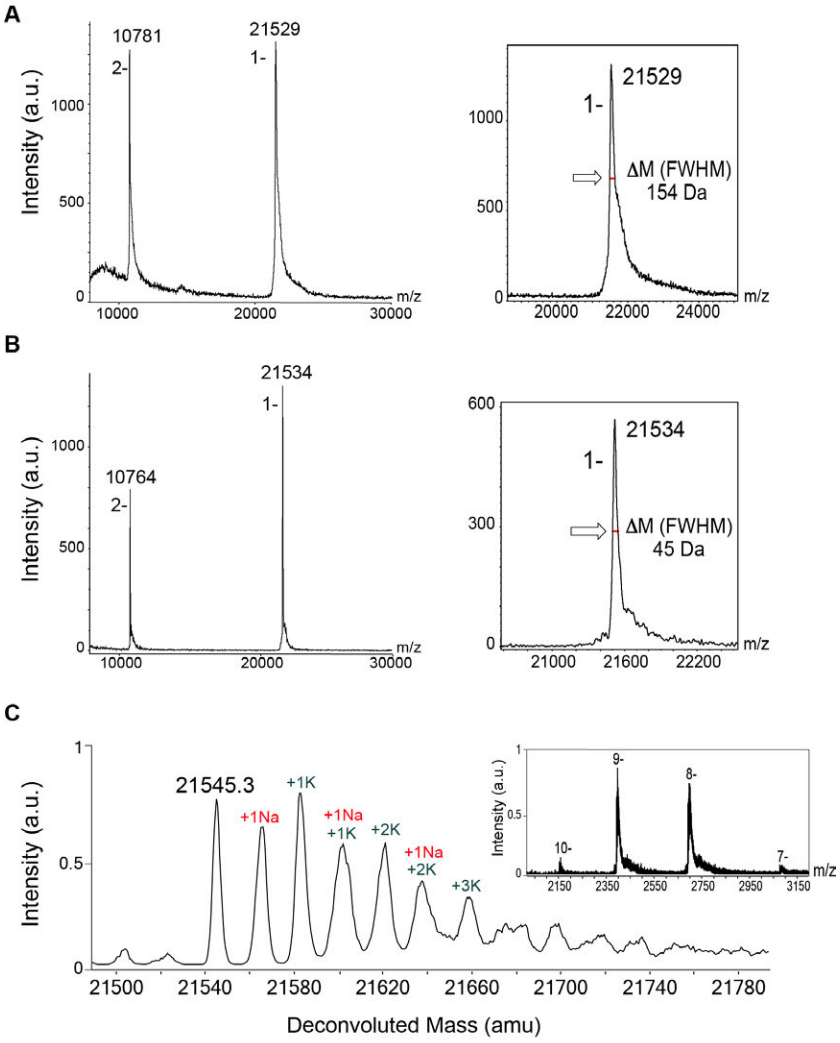
4.3.2 IIABC AND MASS SPECTROMETRY: A CHALLENGING MATCH

One of the greatest challenges in analyzing HIV-1 RNP complexes by native MS was to obtain RNA samples free of non-volatile salts. In order to establish a robust protocol for obtaining RNA samples compatible with native MS analysis, we focused our initial efforts on the 66-nt RRE fragment IIABC.

4.3.2.1 IIABC analysis by MALDI-TOF and LC-ESI-TOF

An efficient and reliable protocol for the analysis of nucleic acids by MS in denaturing conditions has been developed (collaboration with Dr. Luca Signor). This protocol was first tested on synthetic DNA and RNA oligonucleotides (15 nt to 40 nt) (data not shown). Then, using this protocol, samples of IIABC taken at different purification steps were analyzed by MALDI-TOF and LC-ESI-TOF mass spectrometry.

The purification steps of IIABC are summarized in section 4.2.2. The IIABC sample obtained after Q-sepharose chromatography and after dialysis against water was designated sample A. To further remove non-volatile cations, sample A was precipitated with AmAc (sample B). After re-suspension in water samples were analyzed by MALDI-TOF (Figure 4.14 A and B). The masses measured for samples A and B (21,529 and 21,534 Da, respectively) deviate from the theoretical mass by 50 and 55 Da, respectively, which potentially correspond to ~2 Na^+ or K^+ cations still bound to IIABC. The signal for sample B was sharper than for sample A, as shown by the full width at half maximum (FWHM) (Figure 4.14 insets in A and B), suggesting that sample B contains less salt adducts.



D

IIABC sample	Theoretical Mass	Measured Mass MALDI-TOF ± mass deviation	Measured Mass ESI-TOF ± mass deviation
A	21,479.7	21529.0 ± 50	n.d.
B	21,479.7	21534.0 ± 55	21545.3 ± 65

Figure 4.14: IIABC mass spectrometry analysis under denaturing conditions. A) and B) MALDI-TOF MS spectra acquired on sample A (10 pmol deposited on target) and B (5 pmol deposited on target). On the left, a zoom of the singly negatively charged ion peak $[M-H]^-$ with the corresponding value of FWHM. C) ESI-TOF MS deconvoluted mass spectrum of sample B (20 pmol injected). In the inset is reported the non-deconvoluted spectra showing four different charge states (from 10- to 7-). D) Summary of masses, expressed in Da, detected by MALDI-TOF and ESI-TOF MS.

Sample B was further analyzed by ESI-TOF MS (Figure 4.14 C). The measured mass of 21,545.29 Da deviated from the theoretical mass by 65 Da. In addition extra peaks of higher molecular mass were detected, corresponding to IIABC-salt adducts. These results confirmed the higher quality of sample B, identifying the precipitation with AmAc as the key step in improving sample preparation. This allowed us to determine an accurate mass value for IIABC which would subsequently prove useful for the interpretation of native MS data.

4.3.2.2 Preparation of IIABC for native MS

As described in the chapter 2 (§ 2.3.1), after *in vitro* transcription and purification in denaturing conditions, IIABC was aliquoted and lyophilized to minimize degradation during long-term storage. For native MS analysis, IIABC was then processed in different ways (Figure 4.15 A). IIABC was re-suspended in 1.5 M AmAc before analysis by native MS (sample A). The acquired spectrum was characterized by very broad peaks, presumably due to the presence of non-volatile salts such as Na^+ that hampered an accurate mass assignment (Figure 4.15 B). However, we could distinguish two main charge states, at ~3200 and 3800 *m/z*, which encouraged us to further improve the IIABC purification process. To favor the exchange of non-volatile salts with ammonia, after re-suspension in water IIABC was incubated with 1 M AmAc and then dialyzed extensively against water (sample B). As shown in Figure 4.15 C, the acquired spectrum was characterized by very sharp peaks, proving that sample quality was greatly improved compared to sample A. An experimental mass of 21,523 Da was measured, very close to the one measured by ESI-TOF in denaturing conditions (21,545.3 Da, Figure 4.14 D). However each charge-state was characterized by the presence of extra peaks of higher *m/z*, corresponding to IIABC-salt adducts (Figure 4.15 C). To efficiently remove these salts, sample B was further precipitated with AmAc and extensively washed with 70% EtOH (sample C). The sharper and better resolved peaks with fewer salt adducts (Figure 4.14 D) demonstrate that the sample quality was improved even further compared to sample B.

We judged the preparation of IIABC as in sample C to be of sufficient quality for subsequent analysis of RNP complexes by native MS. These results demonstrate that a valid purification procedure has been established to obtain IIABC free from non-volatile salts and compatible with native MS analysis.

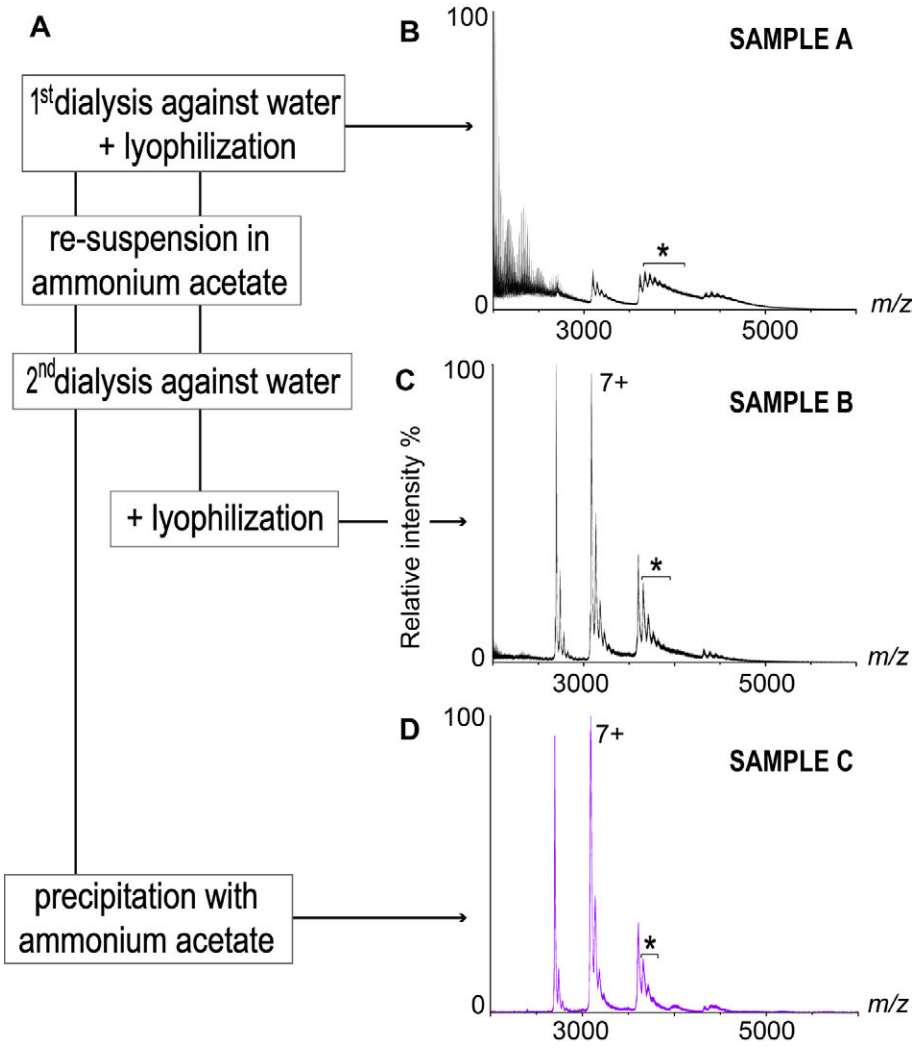


Figure 4.15: IIABC preparation for native MS analysis. A) Protocol used to prepare IIABC for native MS analysis. B-D) Native MS spectra of IIABC samples. All acquisitions were made using IIABC at a final concentration of 5 μM. B) Spectrum acquired for sample A in 0.8 M AmAc. C) Spectrum acquired for sample B in 0.2 M AmAc. D) Spectrum acquired for sample C in 0.2 M AmAc. The IIABC-salt adducts peaks are highlighted by an asterisk only in the 6^+ charge-state of the distribution. Only the most intense charge-state is labelled in each spectrum.

4.3.3 HIV-1 RNP COMPLEXES ASSEMBLY IN NATIVE MS BUFFER

With the aim to analyse HIV-1 RNP complexes by native MS, single components were purified in AmAc as previously described (§ 2.5.4, 2.5.5). The next step was to verify the proper assembly of complexes in AmAc *in vitro*. A preliminary EMSA experiment revealed that IIABC and Rev* mixed in the presence of different AmAc concentrations (10, 50, 100, 200, 600 mM) formed a detectable complex in all concentrations; however, not surprisingly, at 600 mM concentration the affinity of Rev* for IIABC was reduced (data not shown). We therefore decided to perform native MS analysis using 200 mM AmAc.

For native MS experiments the sample concentration commonly recommended is between 1 and 20 μM (Hernandez and Robinson, 2007). Our previous EMSA experiments on IIABC/Rev complexes (§ 4.2.4.1) were performed using concentrations 6-fold lower than the minimal recommendation. Therefore, IIABC/Rev complexes were assembled in 200 mM AmAc by incubating 1 μM of IIABC with increasing amounts of Rev*, Rev^{V16D} and Rev^{WT} and analyzed by EMSA (Figure 4.16). As shown in Figure 4.16 complexes of reduced electrophoretic mobility were detected, indicating that IIABC/Rev complex assembly occurred efficiently in 200 mM AmAc. Interestingly, gel profiles differed from those obtained in standard buffer conditions (Figure 4.10). As expected, since IIABC and Rev were incubated at a 6-fold higher concentration, formation of the same complexes occurred at lower ratios. For example, working at this concentration, a ratio of IIABC:Rev* of 1:4 was sufficient to completely shift all of IIABC (Figure 4.16 A, lane 4), whereas working at the lower concentration a ratio of 1:9 was required (Figure 4.10 A, lane 11). Surprisingly, for Rev* and Rev^{V16D} 5 distinct complexes were detected whereas, in standard conditions, only 4 were observed for Rev* and 3 for Rev^{V16D} (compare Figures 4.16 A, B with 4.10 A, B). Whether the higher number of complexes observed in AmAc at these higher IIABC and Rev concentrations is due to an additional Rev-RNA binding event or to non-specific Rev-Rev interactions is unclear. As expected, the tendency of Rev^{WT} to self-associate led to the formation of non-specific complexes which became increasingly evident as the protein concentration increased (Figure 4.16 C, lanes 5 to 8).

Taken together, the above results confirmed that IIABC/Rev complexes could be successfully assembled in AmAc and were therefore suitable for native MS analyses.

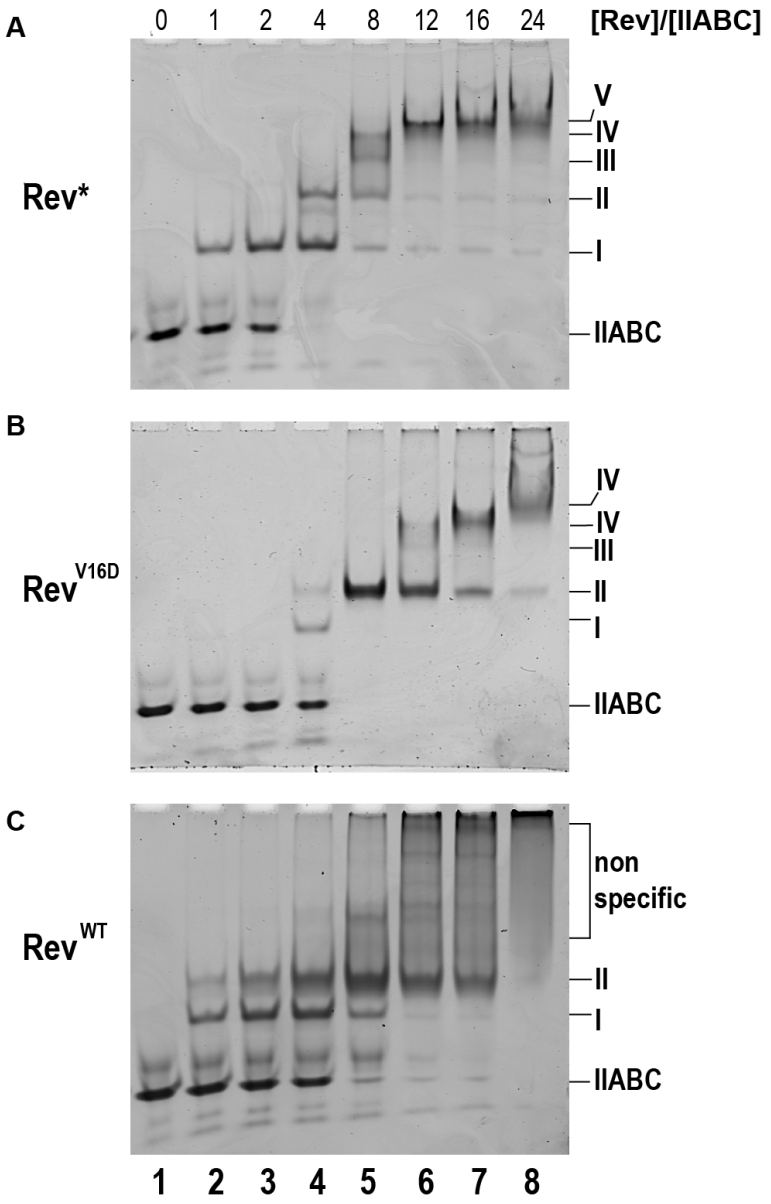


Figure 4.16: EMSA of IIABC/Rev complexes assembled in 200 mM AmAc. 1 μM IIABC was mixed with increasing amounts of (A) Rev*, (B) Rev^{V16D} and(C) Rev^{WT} and complexes were analyzed by EMSA. For each gel the [Rev]/[IIABC] ratio in lanes 1-8 was 0, 1, 2, 4, 8, 12, 16 and 24, respectively. Samples were incubated for 30 min at 22 °C and 5 μL were loaded in each well. Gels were stained using GelRed.

4.3.4 Native MS analysis of IIABC/Rev complexes

Even though our IIABC preparation was compatible with native MS analysis, the RNA still contained residual non-volatile salts (Figure 4.15 D) which might reduce the resolution of signals corresponding to IIABC/Rev complexes by broadening the peaks. To optimize conditions for native MS analysis, we worked first with the IIABC/Rev* complex, and then with complexes containing Rev^{V16D} and Rev^{WT}.

4.3.4.1 Native MS analysis of IIABC/Rev*

Our EMSA results revealed that at least 5 distinct IIABC/Rev* complexes could be formed (Figure 4.16 A). Using this Rev mutant we could remark how the assembly of complexes for native MS analysis was strongly dependent on the initial concentrations of IIABC and Rev* used for sample preparation.

Spectra acquired from samples in which IIABC (5 μM) was mixed with two or four molar equivalents of Rev* (i.e., 10 or 20 μM) only exhibited signals corresponding to unbound IIABC; increasing to eight molar equivalents of Rev* (i.e., 40 μM) yielded well-resolved signals corresponding to a IIABC/Rev* complex with 1:1 stoichiometry (data not shown). These results were surprising since by EMSA, a first complex was detected at a IIABC:Rev* ratio of 1:1 using a 5-fold lower overall concentration of components (Figure 4.16 A, lane 2). To increase the amount of complex, we therefore increased the overall IIABC and Rev* concentrations by 3-fold. Samples prepared with a IIABC:Rev* ratio of 1:2 and 1:4 yielded ion-distributions corresponding to IIABC/Rev* complexes with a stoichiometry of 1:1 and 1:2 (Figure 4.17 B and C). These results agree with the number of species detected by EMSA using the same molar ratios (Figure 4.16 A, lanes 3-4). Other signals were detected in the *m/z* range between 4500-6000, suggesting that higher stoichiometry complexes were present but poorly resolved (Figure 4.17 C).

A

Label	Stoichiometry		Measured	Calculated
	IIABC	Rev*	Mass ± error	Mass
a	1	1	34,700 ± 1	34,696.9
b	1	2	47,868 ± 5	47,864.8
c	1	3	61,032 ± 7	61,032.7
d	1	4	74,206 ± 7	74,200.6
e	1	5	87,373 ± 5	87,368.5
f	1	6	100,533 ± 5	100,536.4

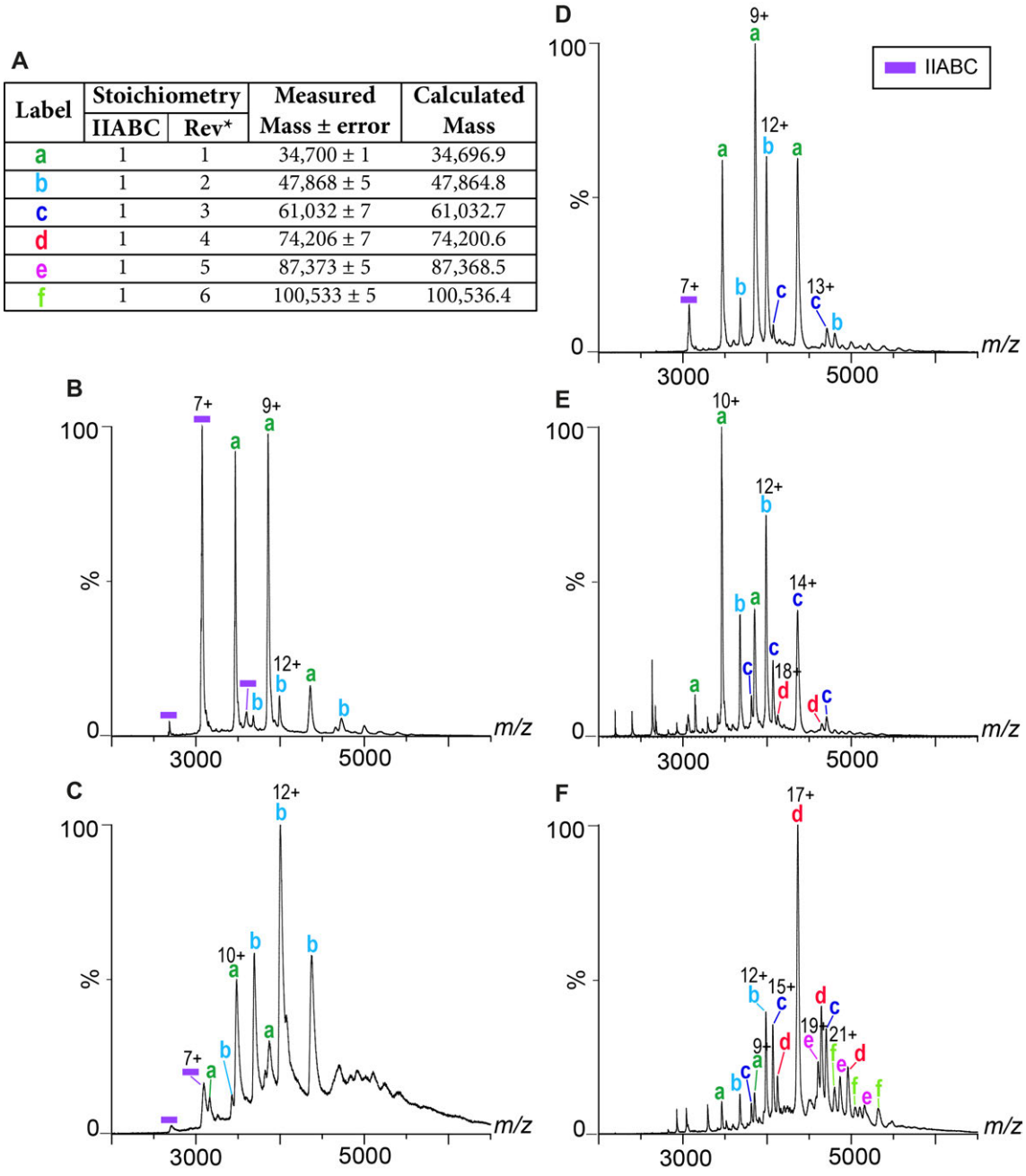


Figure 4.17: Stoichiometry of IIABC/Rev* complexes determined by native MS. All samples were incubated and acquired in 200 mM AmAc. A) Summary of complexes detected. Labels refer to the corresponding ion series in the spectra B to F. B-F) Native MS spectra. B and C) Spectra acquired using 15 µM IIABC and B) 2-fold and C) 4-fold excess of Rev*. Samples were diluted 3-fold immediately before injection. IIABC/Rev* complexes with a stoichiometry of 1:1 (a) and 1:2 (b) were detected. D) Spectrum acquired with 10 µM IIABC and a [Rev*]/[IIABC] ratio of 8. E) Spectrum of IIABC at 5 µM and a [Rev*]/[IIABC] ratio of 12. Spectra in D) and E) identified IIABC/Rev* complexes with stoichiometries ranging from 1:1 to 1:4 (a to d). F) Spectrum acquired with IIABC at 5 µM and a [Rev*]:[IIABC] ratio of 24, revealing two additional IIABC:Rev complexes with stoichiometries of 1:5 (e) and 1:6 (f).

To improve spectral quality, we decided to reduce the IIABC and Rev* concentrations mixed when working at ratios higher than 1:4. At a ratio of 1:8 IIABC/Rev* complexes corresponding to stoichiometries ranging from 1:1 to 1:3 were clearly identified (Figure 4.17 D), consistent with the number of species detected by EMSA at the same ratio (Figure 4.16 A, lane 5). An additional IIABC/Rev* stoichiometry of 1:4 was detected at a ratio of 1:12 (Figure 4.17 E). At the highest ratio tested (1:24), IIABC/Rev* complexes with stoichiometries ranging from 1:1 to 1:6 (Figure 4.17 F) were detected.

Using native MS we could identify species not easily observed by EMSA, due to the limited resolving power of the polyacrylamide gel. At the higher ratios tested only one main band could be detected by EMSA (Figure 4.16 A, lanes 6-8), whereas native MS clearly resolved complexes containing 5 copies of Rev* from those containing 6. Another interesting difference is that at the highest [Rev]/[IIABC] ratios tested, the EMSA profile showed that the majority of IIABC was bound within the highest stoichiometry complex, with only faint bands for the lower-order complexes. In contrast, by native MS the lower-order complexes remained significantly populated, suggesting that the relative stability of complexes may differ from that in solution.

In conclusion, the above experiments show that IIABC can accommodate at least 6 copies of Rev*. More importantly, they establish the feasibility of using native MS to study viral RNP complexes. The resolving power of this technique is illustrated by the fact that several distinct IIABC/Rev* complexes were readily identified with accuracy even in very polydisperse samples.

4.3.4.2 Native MS analysis of IIABC/Rev^{V16D}

Next we analysed the IIABC/Rev^{V16D} complex. Guided by our EMSA data on this complex (Figure 4.16 B) and by our native MS results on IIABC/Rev*, we assembled the IIABC/Rev^{V16D} complex in AmAc at a [Rev]/[IIABC] ratio of 20. Native MS spectra revealed IIABC/Rev^{V16D} complexes with stoichiometries ranging from 1:1 to 1:4 (Figure 4.18), consistent with the four distinct complex species detected by EMSA. Increasing the Rev^{V16D} concentration relative to IIABC did not result in the appearance of additional signals for higher-order complexes. Thus, we conclude that IIABC can accommodate a maximum of four molecules of Rev^{V16D}.

A

Label	Stoichiometry		Measured Mass ± error	Calculated Mass
	IIABC	Rev^{V16D}		
a	1	1	34,572 ± 4	34,720.6
b	1	2	56,252 ± 4	56,249.6
c	1	3	77,779 ± 2	77,778.6
d	1	4	99,310 ± 7	99,307.6

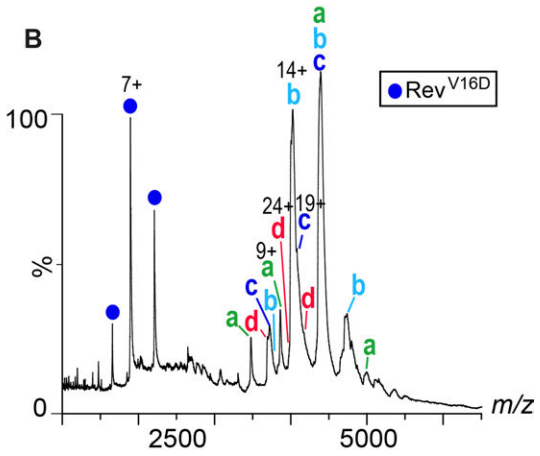


Figure 4.18: Native MS analysis of IIABC/Rev^{V16D} complexes. All samples were prepared in 200 mM AmAc. A) Summary of complexes detected. Labels refer to the corresponding ion series in the spectrum shown in B). B) Native MS spectrum acquired with IIABC at 2.5 µM and a [Rev^{V16D}]/[IIABC] ratio of 20. IIABC/Rev^{V16D} complexes with stoichiometries ranging from 1:1 to 1:4 (labeled a to d) were detected as well as free Rev^{V16D}.

4.3.4.3 Native MS analysis of IIABC/Rev^{WT}

We next studied the IIABC/Rev^{WT} complexes. The analysis was hampered by the tendency of Rev^{WT} to aggregate, but after extensive optimization of initial component concentrations and ratios mixed we succeeded in acquiring informative native MS spectra. We were able to reliably identify seven distinct complexes (Figure 4.19 A), which we classified as either 'specific' or 'non-specific'. Specific complexes contain only one molecule of IIABC bound to 1, 2 or 3 copies of Rev^{WT}. The relative proportion of these three species shifts to higher stoichiometry as the ratio of Rev to IIABC is increased. These results agree with previous EMSA analysis where 2 or 3 different complex bands were observed (Figures 4.10 C and 4.16 C). Complexes containing more than one molecule of IIABC were considered non-specific. These include complexes with 2 molecules of IIABC bound to 2, 3, 4 or 6 copies of Rev^{WT}. These complexes presumably arise due to homotypic interactions between Rev monomers bound to distinct IIABC molecules (see below) and may correspond to some of the well-defined bands that we observe among the non-specific signals in our EMSA experiments at [Rev]/[IIABC] ratios of 8 or higher (Figure 4.16 C lanes 5 to 8).

A

Label	Stoichiometry		Measured Mass ± error	Calculated Mass
	IIABC	Rev^{WT}		
a	1	1	34,704 ± 1	34,704.7
b	1	2	47,880 ± 0	47,880.4
c	1	3	61,058 ± 2	61,056.1
d	2	2	69,007 ± 3	69,409.4
e	2	3	83,063 ± 6	82,585.1
f	2	4	96,312 ± 9	95,760.8
g	2	6	123,042 ± 4	122,112.2

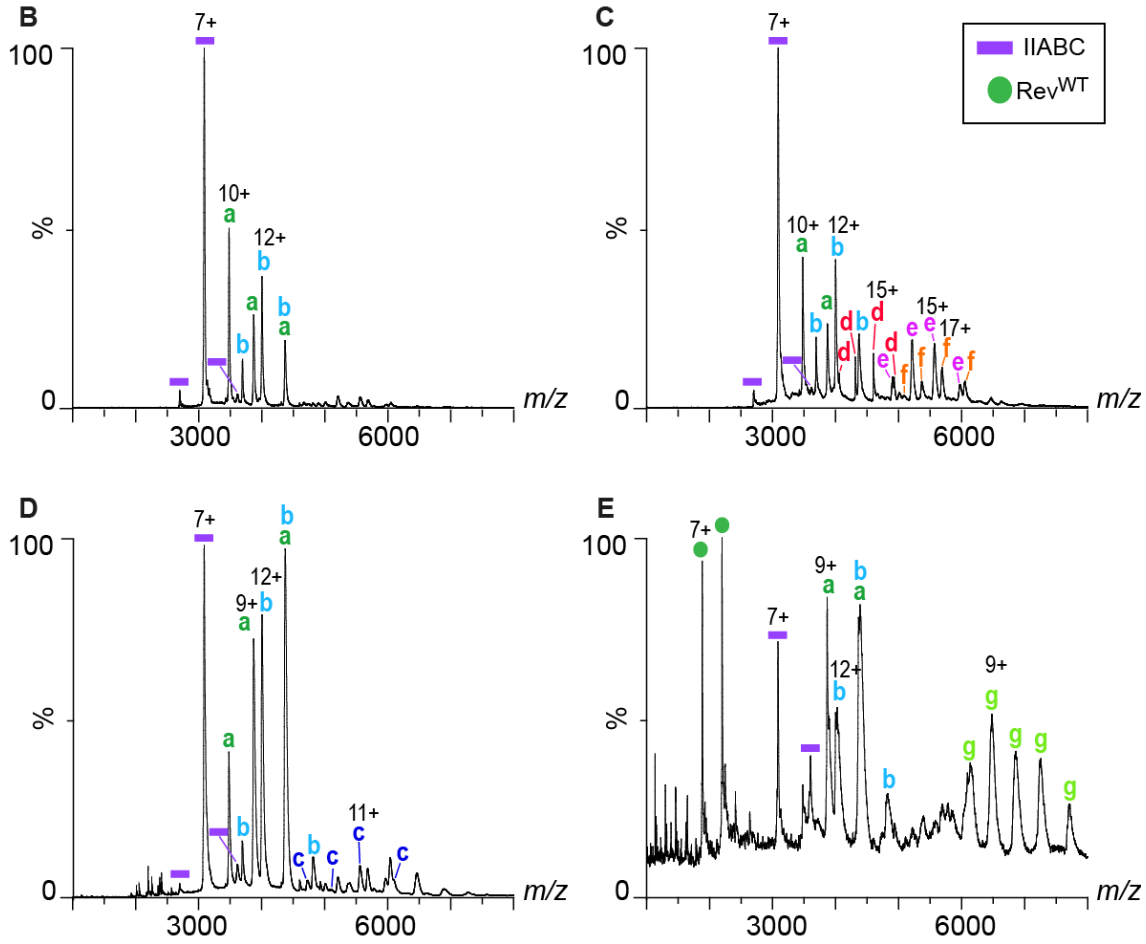


Figure 4.19: Native MS analysis of IIABC/ Rev^{WT} complexes. All samples were acquired in 200 mM AmAc. A) Summary of complexes detected. All the ion series are labelled. Free IIABC and free Rev^{WT} are labeled as indicated in the legend. B) and C) Spectra acquired with IIABC at 15 µM and a [Rev^{WT}]/[IIABC] ratio of 0.7 and 1.5, respectively. Specific IIABC/Rev^{WT} complexes with a stoichiometry of 1:1 (a) and 1:2 (b) and non-specific complexes with a stoichiometry of 2:2 (d), 2:3 (e) and 2:4 (f) were detected. D) Spectrum of IIABC at 5 µM and a [Rev^{WT}]/[IIABC] a ratio of 1:7. Specific IIABC/Rev^{WT} complexes with a stoichiometry of 1:1 (a), 1:2 (b) and 1:3 (c) were detected. E) Spectrum of IIABC at 2.5 µM and a [Rev^{WT}]/[IIABC] ratio of 15. Specific complexes as in D and a non-specific IIABC/Rev^{WT} complex with a stoichiometry of 2:6 (g) were detected.

The above data confirm that IIABC can accommodate up to 3 copies of Rev^{WT}, which is one fewer than Rev^{V16D} and three fewer than Rev*. We note that complexes containing 1 or 2 Rev^{WT} monomers bound to IIABC were more abundant than the complex containing 3 Rev^{WT} monomers, suggesting that the latter was less stable. The stoichiometry deduced for the non-specific complexes detected can be explained by summing the stoichiometries of two smaller complexes (2:2 = 1:1 + 1:1; 2:3 = 1:1 + 1:2; 2:4 = 1:1 + 1:3 or 1:2 + 1:2; 2:6 = 2:3 + 2:3), suggesting that their formation reflects the ability of Rev^{WT} to bridge two distinct RNA molecules. Such species were not observed with Rev* and Rev^{V16D}, even at high relative Rev concentrations, suggesting that the ability of Rev^{WT} to bridge two RNA molecules requires the integrity of both the H-H and T-T oligomeric interfaces.

4.3.5 Native MS analysis of IIABC/Rev/CRM1/Ran^{GTP} complexes

An important aim of this study was to determine the Rev:CRM1 stoichiometry in the full HIV-1 RNP export complex. As native MS of this viral RNP complex is very challenging, we first focused on the analysis of a smaller HIV-1 RNP export complex containing IIABC as the RNA scaffold. We assembled complexes containing CRM1 and Ran using a stepwise approach. First, informed by our native MS data (§ 4.3.4), we pre-assembled IIABC/Rev complexes in such a way that the complex with the desired stoichiometry would form the most abundant species (when possible). Then CRM1 and Ran were added to form a quaternary complex. As a general rule, we tried to reduce the sample polydispersity (number of different complexes in solution) as much as possible. Based on EMSA experiments (§ 4.2.4.2) CRM1 was added in a molar ratio that was at least half that of Rev, while Ran was added in a 4- or 5-fold molar ratio relative to CRM1, to ensure that CRM1 was fully bound to Ran and to the IIABC/Rev complex. The precise mixing order, incubation time and buffer composition proved to be crucial for obtaining a sample that gave a stable signal over time without blocking the injection needle. As expected, the most difficult samples to analyze were those containing Rev^{WT}.

4.3.5.1 Native MS analysis of IIABC/Rev*/CRM1/Ran^{GTP}

EMSA and native MS analyses of IIABC/Rev* complexes showed that at a [Rev*]/[IIABC] molar ratio of 2 or 4, the predominant complex formed contained either 1 or 2 Rev* molecules bound to IIABC, respectively (§ 4.3.3, 4.3.4.1). These complexes were pre-assembled and then incubated with CRM1 and Ran to build up quaternary complexes that were subsequently analyzed by native MS. When IIABC was incubated with Rev* to generate a 1:1 complex as the most abundant stoichiometry (Figure 4.17 B), the addition of CRM1 and Ran yielded a very intense ion distribution centered at ~7000 *m/z*, corresponding to the intact IIABC/Rev*/CRM1/Ran^{GTP} complex with 1:1:1:1 stoichiometry (Figure 4.20 A, B).

A

Label	Stoichiometry				Measured	Calculated
	IIABC	Rev*	CRM1	Ran^{GTP}	Mass ± error	Mass
a	0	1	1	1	162,874 ± 3	162,874.8
b	1	1	1	1	184,404 ± 2	184,403.8
c	1	2	1	1	197,572 ± 4	197,571.6

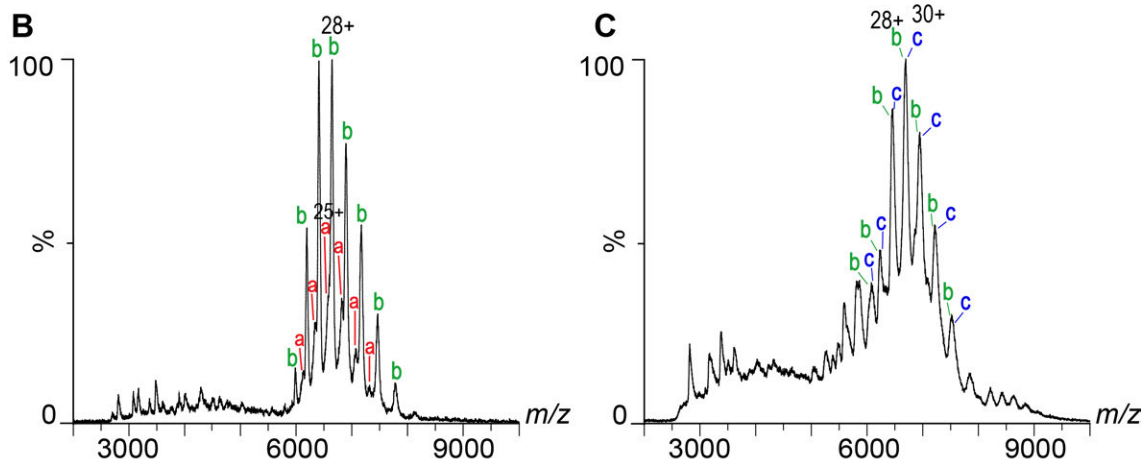


Figure 4.20: Native MS analyses of IIABC/Rev*/CRM1/Ran^{GTP} complexes. All samples were incubated and acquired in 200 mM AmAc, 2.5 mM MgAc and 1 mM DTT. A) Summary of complexes detected. Labels refer to the corresponding ion series in the spectra shown in B) and C). B) Spectrum acquired at lower Rev concentration. 15 µM of IIABC were pre-incubated with Rev* in a ratio of 1:2 in 200 mM AmAc, the sample was diluted 3 times and then 5 µM of CRM1 and 20 µM of Ran were added. Detected signals correspond to the Rev*/CRM1/Ran^{GTP} (a) and the IIABC/Rev*/CRM1/Ran^{GTP} (b) complexes. C) Spectrum acquired at higher Rev concentration. 15 µM of IIABC were incubated with Rev* in a ratio of 1:4 in 200 mM AmAc, the sample was diluted 3 times and then 10 µM of CRM1 and 40 µM of Ran were added. Detected signals corresponding to the IIABC/2Rev*/CRM1/Ran^{GTP} complex are labeled c.

When IIABC was incubated with Rev* to generate a mixture of 1:1 and 1:2 complexes (Figure 4.17 C), the addition of CRM1 and Ran resulted in an intense ion distribution characterized by broader peaks compared to those in the previous experiment. Analysis of the spectrum suggested that these signals resulted from the overlap of two ion distributions, corresponding to IIABC/Rev*/CRM1/Ran^{GTP} complexes with a stoichiometry of 1:1:1:1 and 1:2:1:1 (Figure 4.20 C). The fact that the mass difference between the two complexes is that of a small protein (Rev*) causes the two ion distributions to be very close to each other, explaining the reduced quality of the second spectrum compared to the first.

These results demonstrate that quaternary HIV-1 RNP export complexes can be successfully assembled in AmAc and analyzed by native MS. In particular, these data show that IIABC bound by 1 or 2 copies of Rev* can recruit one molecule of CRM1.

4.3.5.2 Native MS analysis of IIABC/Rev^{V16D}/CRM1/Ran^{GTP}

Our native MS data acquired on IIABC/Rev^{V16D} showed that IIABC could bind up to 4 molecules of Rev^{V16D} (§ 4.3.4.2). We therefore wondered how many molecules of CRM1 could bind to a IIABC/Rev^{V16D} complex containing more then 2 copies of Rev^{V16D}.

IIABC was incubated with Rev^{V16D} in order to achieve stiochiometries ranging from 1:1 to 1:4 (Figure 4.18 B), and then CRM1 and Ran were added. The detected complexes are summarized in Figure 4.21 A. The samples were extremely polydisperse and ion distributions in the high *m/z* range had low intensity, hampering reliable mass assignments (Figure 4.21 B). Thus the acquisition was performed in the *m/z* range 5000-10000 (Figure 4.21 C). In the 5000-7000 *m/z* range signals corresponding to the Rev^{V16D}/CRM1/Ran^{GTP} complex with a stoichiometry of 1:1:1 were clearly detected. Two intense peaks, indicated with a question mark, could not be assigned. In the 8000-10000 *m/z* range two ion distributions were detected and they overlapped strongly in the lower half of the ion population and were better resolved above 9000 *m/z*, suggesting that the difference in mass could correspond to a small component. The most plausible mass deconvolution of these signals corresponds to IIABC/Rev^{V16D}/CRM1/Ran^{GTP} complexes with a stoichiometry of 1:4:2:2 and 1:4:2:1 (Figure 4.21 C). Interestingly, compared with complexes assembled using Rev*, no quaternary complexes containing only 1 or 2 copies of Rev^{V16D} were detected.

Label	Stoichiometry				Measured Mass ± error	Calculated Mass
	IIABC	Rev^{V16D}	CRM1	Ran^{GTP}		
a	0	1	1	1	162,902 ± 2	162,898.4
b	1	4	2	2	373,707 ± 7	373,709.0
c	1	4	2	1	348,567 ± 9	348,568.5
?	-	-	-	-	127,082 ± 12	

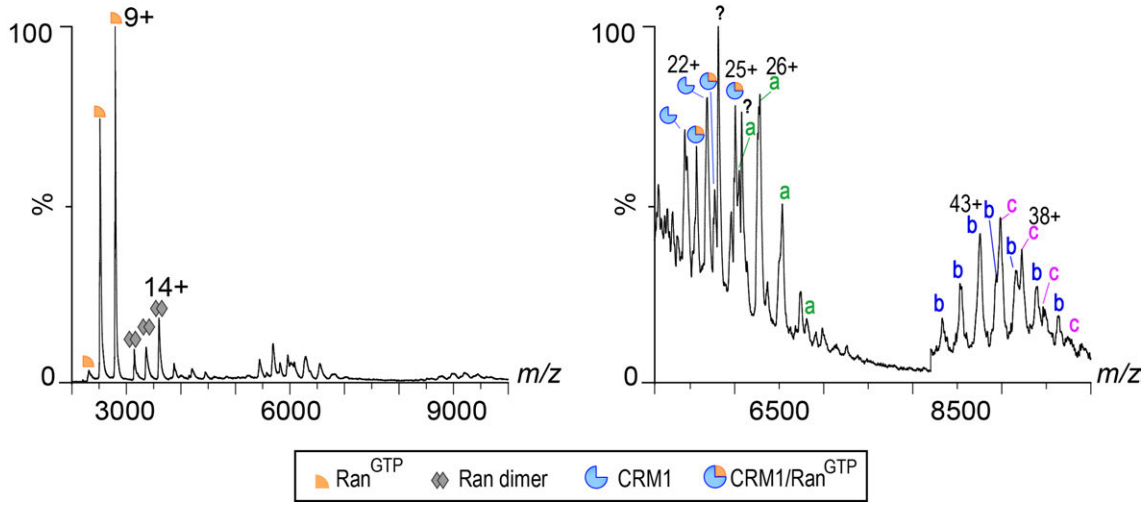


Figure 4.21: Native MS analysis of IIABC/ Rev^{V16D}/CRM1/Ran^{GTP} complexes. A) Summary of complexes detected.. B) Native MS spectrum. 2.5 µM of IIABC were incubated with Rev^{V16D} in a ratio of 1:20 in 200 mM AmAc. Then 10 µM of CRM1 and 100 µM of Ran were added and incubated in the same buffer with addition of 2.5 mM Mg Ac and 1 mM DTT. C) Acquisition in the *m/z* range 5000-10000. In the 5000-7500 *m/z* range the Rev^{V16D}/CRM1/Ran^{GTP} complex (163,201 Da) is labeled with letter a. In the 8000-10500 *m/z* range (magnified 3X) two detected complexes correspond to a IIABC/Rev^{V16D}/CRM1/Ran^{GTP} stoichiometry of 1:4:2:2 (b) and 1:4:2:1 (c).

The above data suggest that a IIABC/Rev^{V16D} complex containing 4 copies of Rev^{V16D} can recruit at least 2 molecules of CRM1. However, these results need to be confirmed, as we were unable to verify the stoichiometry by performing MS-MS analysis due to the low signal intensity.

4.3.5.3 Native MS analysis of IIABC/Rev^{WT}/CRM1/Ran^{GTP}

Based on native gel performed in standard conditions (Figure 4.11 C lane 4), at least 3 different complexes were detected when CRM1 was added, in presence of Ran, to IIABC/Rev^{WT} complexes.

We next decided to tackle the quaternary export complex containing Rev^{WT}. IIABC was incubated with Rev^{WT} to form 1:1 and 1:2 complexes (Figure 4.19 D), and then CRM1 and Ran were added. As expected, the sample was extremely polydisperse (Figure 4.22 A, B). Nevertheless, signals were sufficiently well resolved to allow the ion distributions to be assigned. In the 3000-6000 *m/z* range IIABC/Rev^{WT} complexes, free CRM1 and CRM1/Ran^{GTP} complex, all observed in previous experiments, were detected. Two peaks could not be assigned that interestingly have similar *m/z* value of the non-assigned ones detected during the IIABC/Rev^{V16D}/CRM1/Ran^{GTP} analyses (Figure 4.21 C). A charge-state distribution centered at ~7000 *m/z* corresponded to the intact IIABC/Rev^{WT}/CRM1/Ran^{GTP} complex with a stoichiometry of 1:1:1:1 (Figure 4.22 B).

To assemble complexes with higher stoichiometry, we increased the CRM1 and Ran concentrations used to prepare samples. Up to ~7500 *m/z*, signals detected correspond to complexes or free species seen with the lower CRM1/Ran concentrations (Figure 4.22 C). In the 8500-9500 *m/z* range two new ion distributions were detected. The close proximity of the signals suggested two complexes that differed by a small component (Figure 4.22 D). The data are compatible with two hypothetical pairs of complexes. According to hypothesis 1, the signals are due to a Rev^{WT}/CRM1/Ran^{GTP} complex with 3:2:2 stoichiometry (peaks labelled 'B') and a IIABC/Rev^{WT}/CRM1/Ran^{GTP} complex with 1:3:2:2 stoichiometry (peaks labelled 'C'), i.e., that differ by the presence or absence of IIABC (Figure 4.22 A, D). According to hypothesis 2, the signals are due to two IIABC/Rev^{WT}/CRM1/Ran^{GTP} complexes having a stoichiometry of 2:3:2:1 (peaks labelled 'b') and 2:4:2:1 (peaks labelled 'c'), i.e. that differ by one Rev monomer (Figure 4.22 A, D). The charge state distribution, centered at ~ 10500 *m/z* was identified as a IIABC/Rev^{WT}/CRM1/Ran^{GTP} complex with a stoichiometry of 2:4:3:3 (f) (Figure 4.22 D).

These data suggest that a complex of IIABC bound to either 3 or 4 copies of Rev^{WT} can recruit up to 2 molecules of CRM1. Nevertheless these data must be confirmed, as we could not perform MS-MS analysis due to signal low intensity.

A

Label	Stoichiometry				Measured Mass ± error	Calculated Mass
	IIABC	Rev^{WT}	CRM1	Ran^{GTP}		
a	1	1	1	1	184,411 ± 1	184,411.6
B	0	3	2	2	338,938 ± 8	338,940.9
C	1	3	2	2	360,470 ± 8	360,469.9
b	2	3	2	1	356,867 ± 10	356,858.4
c	2	4	2	1	370,031 ± 4	370,034.1
f	2	4	3	3	544,874 ± 9	544,881.5
g	1	1	0	0	34,704 ± 1	34,704.7
h	1	2	0	0	47,879 ± 0	47,880.4
i	2	2	0	0	69,007 ± 3	69,409.4
?	-	-	-	-	127,082 ± 12	

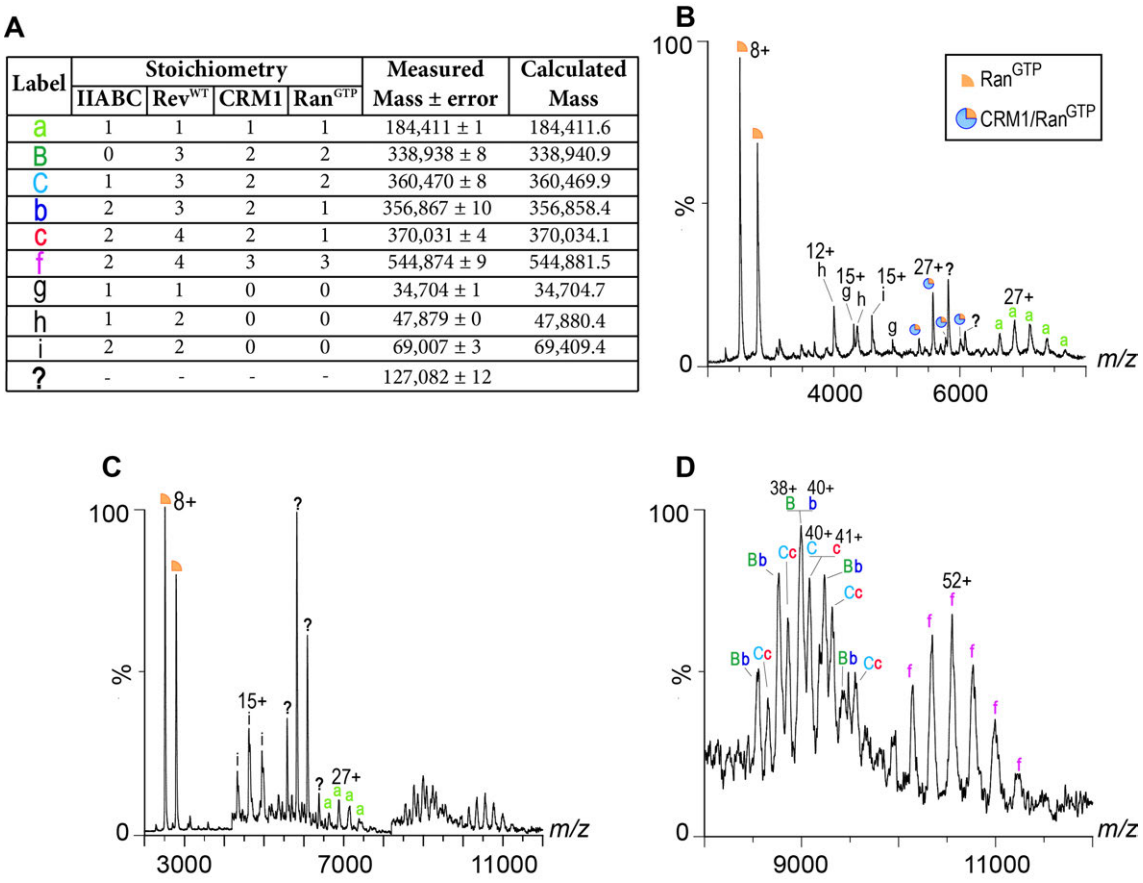


Figure 4.22: Native MS analysis of the IIABC/Rev^{WT}/CRM1/Ran^{GTP} complex. A) Summary of complexes detected. Samples were incubated in 200 mM AmAc, 2.5 mM MgAc and 1 mM DTT. B) Native MS spectrum at low Ran concentration. 5 µM of IIABC was pre-incubated with Rev^{WT} in a ratio of 1:7 and then 10 µM of CRM1 and 20 µM of Ran were added. Free species and complexes previously observed are labeled with symbols (as indicated in the legend) or letters (as indicated in panel A). The IIABC/RevWT/CRM1/Ran^{GTP} complex with a stoichiometry of 1:1:1:1 is labeled with letter a. C) Native MS spectrum at high Ran concentration. 2.5 µM of IIABC was pre-incubated with Rev^{WT} in a ratio of 1:7 then 8.75 µM of CRM1 and 35 µM of Ran were added. Signals below 7000 *m/z* correspond to free species or complexes previously observed. Signals in the *m/z* range 4000-8000 and 8000-12000 are magnified 3-fold and 15-fold respectively. D) Magnified view of the 8000-12000 *m/z* range of the spectrum shown in C. The pair of complexes described by Hypothesis 1 (labeled B and C) or Hypothesis 2 (b and c) are indicated. The ion distribution centered at ~10500 *m/z* corresponds to a 2:4:3:3 IIABC/Rev^{WT}/CRM1/Ran^{GTP} complex.

4.4 MS ANALYSES OF THE FULL HIV-1 RNP EXPORT COMPLEX: PRELIMINARY RESULTS

4.4.1 ASSEMBLY OF THE FULL RNP EXPORT COMPLEX

Next we turned to the more challenging task of investigating the binding of Rev, CRM1 and Ran to the intact RRE. The difficulty of this task is hinted at by the different results obtained by EMSA analysis when the experiment is performed under standard conditions or in AmAc (Figure 4.23). In standard conditions, titrating the RRE with Rev* results in at least 12 distinct bands that become differentially populated as the Rev* concentration is gradually increased over the first 10 lanes, ending in a single thick band that migrates near the top of the gel as the highest concentrations tested (Figure 4.23 A). By contrast, in AmAc a similar titration yields fewer and more poorly resolved bands that end in a more smeared band at the highest Rev* concentrations (Figure 4.23 B). These results would have discouraged us from pursuing analysis of the RRE if we had not performed an EMSA experiment using Rev^{WT}. Surprisingly, the titration with Rev^{WT} in AmAc gave sharp bands on the gel (Figure 4.23 C). More importantly, we observed a sharp transition at a given Rev concentration in which a relatively high-order complex appeared while lower-order complexes remained only weakly populated, suggesting a strongly cooperative effect. This observation encouraged us to put significant energy into analyzing RRE-bound complexes.

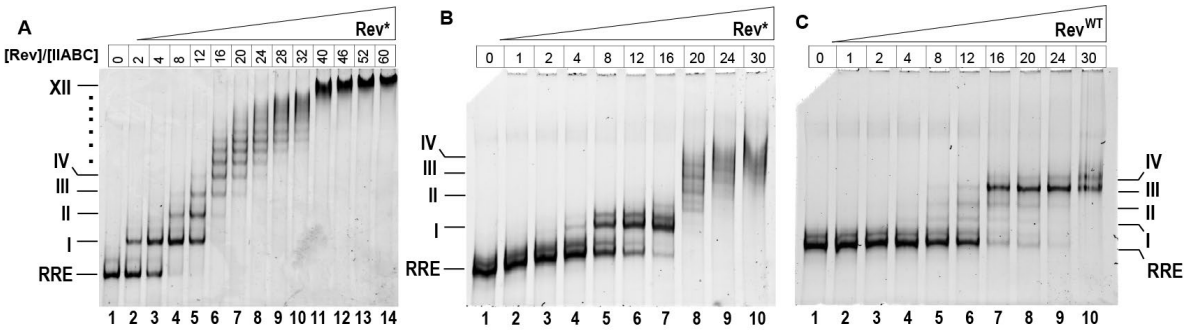


Figure 4.23: EMSA of full-length RRE bound to Rev. A) 0.1 μM RRE was mixed with increasing amounts of Rev* in standard binding buffer. B) and C) 0.2 μM RRE was mixed with increasing amounts of Rev* and Rev^{WT}, respectively, in 200 mM AmAc. The [Rev]/[IIABC] ratio is indicated in the table above each gel. Samples were incubated for 30 min at 22 °C and 5 μL were loaded in each well. Gels were stained using GelRed.

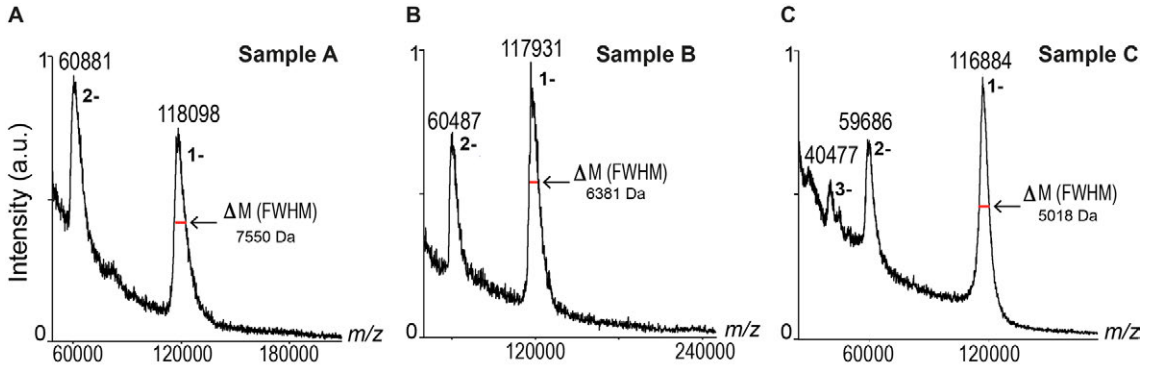
4.4.2 RRE ANALYSIS BY MS: A NEW CHALLENGE

In order to assemble and study the full HIV-1 RNP export complex by native MS, the first challenge was to produce the RRE in a manner compatible with the MS analysis. As described above, we established a robust protocol for producing IIABC binding low amount of non-volatile salts (§ 4.3.2.2). Based on these results, we optimized RRE sample preparation for native MS analysis. As discussed, some residual IIABC-salt adducts were still clearly detected by native MS. As the RRE is 5 times longer, we expected it to interact with at least 5 times more cations then IIABC, making its preparation for native MS analysis even more challenging.

4.4.2.1 RRE analysis by MALDI-TOF

Following the roadmap we had established for IIABC, we analysed the RRE at different stages of the purification (detailed in § 4.2.2) to assess sample quality: after size exclusion chromatography (sample A); after AmAc precipitation (sample B); and following electroelution from the gel and AmAc precipitation (sample C). Samples were analyzed by MALDI-TOF (Figure 4.24). The obtained experimental masses (118,098 Da for sample A; 117,931 Da for sample B; 116,884 Da for sample C) deviated from the theoretical masses by between 2,515 and 1,301 Da (Figure 4.24 D). These data show how the RRE quality improved throughout the different purification steps. Sample C was clearly the best preparation as demonstrated by the smallest mass deviation from the theoretical mass and the sharpest peak (smallest FWHM).

These results confirm that an *in vitro* transcribed RNA molecule as long as 350 nt can be successfully analyzed by MALDI-TOF MS in denaturing conditions. Moreover, spectra are a highly informative indicator of the state of the RNA sample, making this technique suitable for the quality control of nucleic acids.



D

RRE sample	Theoretical Mass	Measured Mass MALDI-TOF	Mass deviation
A	115,585.5	118,098.0	2,514.5
B	115,585.5	117,931.0	2,347.5
C	115,585.5	116,884.0	1,300.5

Figure 4.24: RRE mass spectrometry analysis by MALDI-TOF. Spectra acquired on sample A (4.5 pmol deposited on target), B (5 pmol deposited on target) and C (1.65 pmol deposited on target) are shown in A), B) and C) respectively. The values of FWHM are indicated in correspondence of singly negatively charged ion peak $[M-H]^-$. D) Summary of masses, expressed in Da, detected by MALDI-TOF.

4.4.2.2 RRE preparation for native MS

Many RRE preparations were tested during this study to optimize sample quality for native MS. The key experiments that helped us to establish a protocol for obtaining an RRE sample compatible with native MS studies are presented below (Figure 4.25). Briefly, after *in vitro* transcription (detailed in § 2.3.2 and 4.2.2), the RRE was purified by SEC in a standard buffer. Before native MS analysis a second buffer exchange step, by SEC or ultrafiltration, in up to 2 M AmAc was performed (sample A). The RRE at 3 µM was analyzed by native MS using a collision energy of 30 Vs, but the detected signals were not resolved (data not shown). Afterwards, the same sample was analyzed using a collision energy of 180 Vs and yielded sharp ion peaks (Figure 4.25 B). This result was encouraging, but the sample quality was far from ideal for further analysis of RNP complexes by native MS. Therefore, we explored a different approach. To reduce the non-volatile salt concentration in the RRE samples, we modified the *in vitro* transcription protocol by replacing Tris-HCl, NaCl and MgCl2 with AmAc, sodium acetate and magnesium acetate, respectively (§ 2.3.2, Table 2.2).

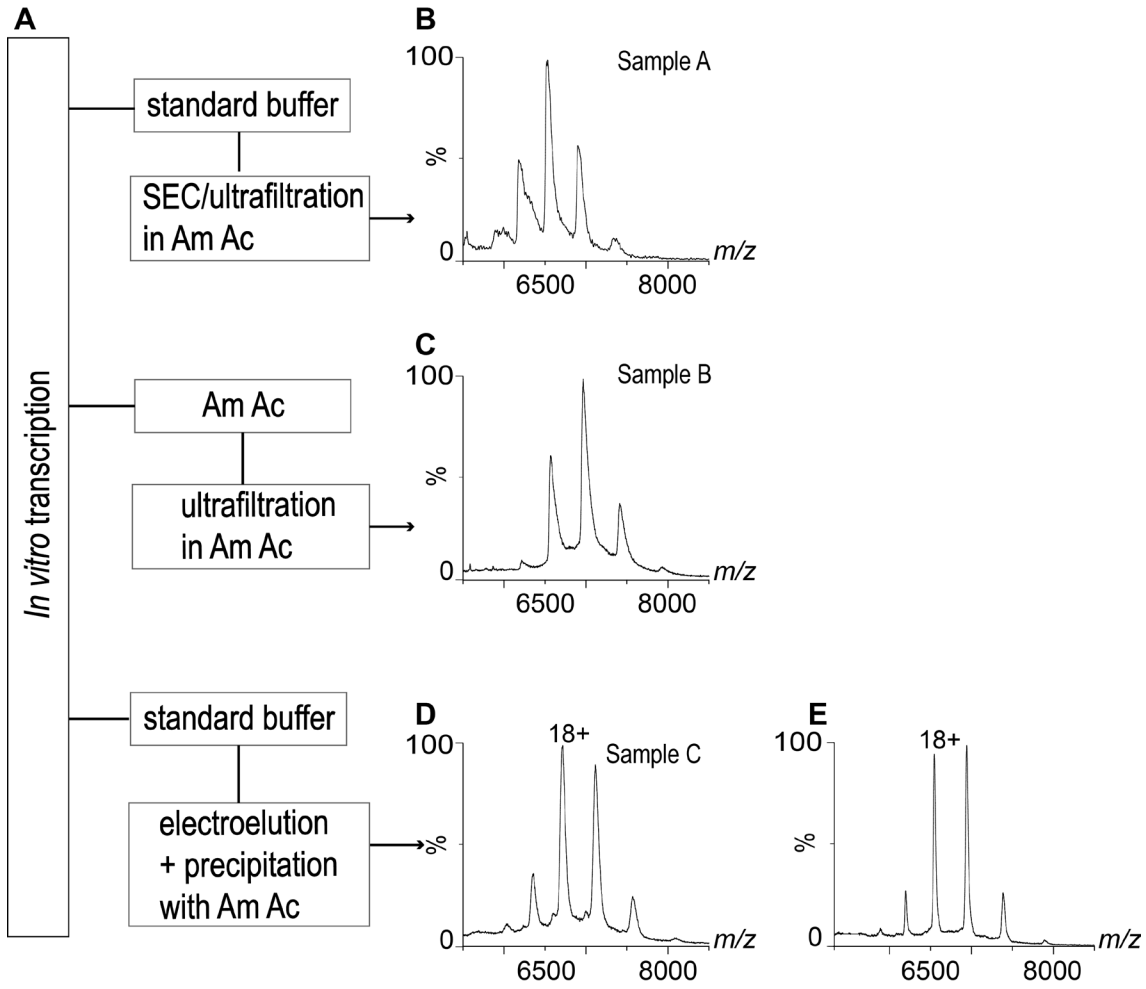


Figure 4.25: RRE preparation for native MS analysis. A) Schema of RRE preparation. All acquisitions were made at a final RRE concentration between 3 and 5 µM in 0.2 M AmAc. B) native MS spectrum of sample A. C) Native MS spectrum of sample B. D) and E) Native MS spectra of sample C acquired using a collision energy of 30 and 80 V, respectively.

A pilot experiment in a small reaction volume (30 µL) showed that the RRE was well transcribed in all the buffers tested (data not shown). As it does not contain NaAcetate, buffer 1.3 was chosen to scale up *in vitro* transcription of the RRE, followed by a buffer exchange step (by ultrafiltration) in 0.6 M AmAc (sample B). Native MS analysis using a collision energy up to 80 V showed a clear peak distribution (Figure 4.25 C), revealing an improvement in sample quality. However the peak width and high baseline suggests the presence of non-volatile salts still bound to the RRE, implying that the *in vitro* synthesis of the RRE in Na^+ free condition did not completely remove these ions. An alternative possible explanation could be the presence of abortive transcripts produced during the

transcription reaction. Indeed, we noticed that after *in vitro* transcription the degree of RRE purity varied (Figure 4.7 E), depending on the quality and freshness of both the DNA template and the RNA polymerase used.

Therefore, we decided to produce the RRE *in vitro* using the conventional transcription buffer, followed by a purification step on a denaturing gel. The RRE was then electroeluted and precipitated using AmAc (sample C). The native MS spectrum of the RRE acquired at 30 V clearly showed a well resolved charge-state distribution, which upon deconvolution yielded a mass of 117,900 Da (Figure 4.25 D). The use of a collision energy of 80 V during acquisition improved the quality of the signal. Indeed, peaks were sharper and the baseline was reduced. An experimental mass of 117,034 Da was calculated, which differs from the theoretical one of ~ 1,400 Da. (Figure 4.25 E). Although the RRE preparation was still not completely free of non-volatile salts, the sample quality seemed sufficiently good for the native MS analyses of RRE-containing complexes.

4.4.3 Analysis of RRE/Rev complexes by native MS.

As for the analysis of the minimal HIV-1 RNP export complex, we first studied the stoichiometry of RRE/Rev complexes using Rev* to optimize experimental conditions. We note that the RRE sample is not yet optimal because it still contains significant traces of non-volatile salts, which are expected to reduce the quality and resolution of MS spectra collected on RRE-containing complexes.

4.4.3.1 Determination of RRE/Rev* stoichiometry

We previously succeeded in analysing IIABC/Rev* complexes by preparing samples containing IIABC at 5-15 µM concentration (see § 4.3.4.1). Based on this experience, we initially assembled RRE/Rev* complexes by incubating 10 µM concentration of RRE with increasing concentrations of Rev*. For all molar ratios tested (1:4, 1:8 and 1:12) only two charge-state distributions, corresponding to free RRE and Rev*, were detected (i.e., no RRE/Rev* complex was detected; data not shown). Moreover, we observed repeated needle clogging and an increase in baseline noise over time, suggesting aggregation of the sample. Therefore, we decreased the RRE concentration (to 5 µM) and incubated with Rev* in a molar ratio of 1:12. We observed peaks in the acquired spectra at low *m/z* corresponding to free Rev* as well as peaks at ~7500-8500 *m/z* predicted to correspond to RRE/Rev*

complexes. However, the latter signals were too poorly resolved to allow us to assign a reliable experimental mass (data not shown). These results nevertheless encouraged us to continue efforts to optimize conditions for assembling the complex.

Because the acquisition of spectra remained hampered by repeated needle clogging, we decided to assemble complexes at an even lower concentration (2 μM) of RRE, which we incubated with Rev* in a ratio of 1:17.5. Although the quality of the acquired spectrum was not ideal, an ion distribution corresponding to a complex comprising one Rev* monomer bound to the RRE was clearly detected (Figure 4.26 A, B). To favor the detection of complexes of higher stoichiometry, the RRE concentration was further decreased (to 1.75 μM) while the Rev* concentration remained unchanged (resulting in a higher molar ratio of 1:20). Two extra charge-state distributions were detected, corresponding to RRE complexes containing either 2 or 3 monomers of Rev* (Figure 4.26 C). Repeating the experiment using slightly different conditions (RRE at 1.7 μM and an RRE:Rev* ratio of 1:15) allowed us to resolve four different charge-state distributions, corresponding to RRE:Rev* stoichiometries ranging from 1:1 to 1:4 (Figure 4.26 D). To date, we have been unable to detect complexes containing more than 4 Rev monomers bound to the RRE. Our attempts to detect such complexes have been thwarted by the RRE sample quality, which, combined with the small mass increment due to each Rev* monomer, yielded poorly resolved signals. Moreover the affinity of Rev* for the RRE could be insufficient to detect such complexes. In any event, the above results provide the first indication that the analysis of *in vitro* assembled HIV-1 RNP complexes containing full-length RRE is feasible, while at the same time highlighting the importance of sample quality for obtaining well-resolved and informative spectra.

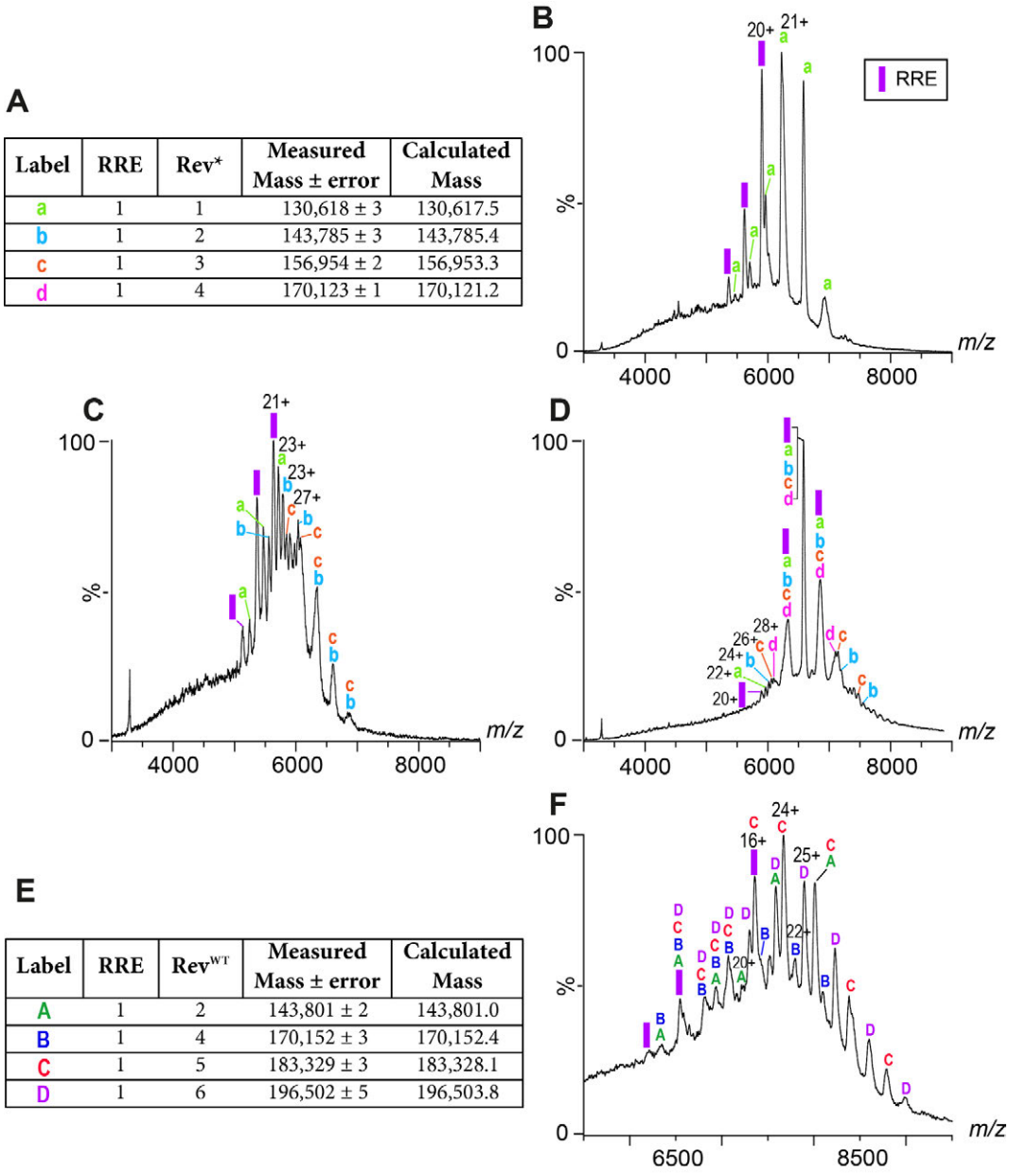
4.4.3.2 Determination of RRE/Rev^{WT} stoichiometry

As RRE/Rev* complexes caused repeated needle clogging even at low concentrations, we directly assembled the RRE/Rev^{WT} complex using reduced concentrations of components. When the RRE at 2.5 μM was incubated with Rev^{WT} in a 1:4 molar ratio, only signals corresponding to the free RRE were detected (data not shown). However, increasing this ratio to 1:12 yielded signals corresponding to RRE/Rev^{WT} complexes. Four different charge-state distributions were detected, corresponding to complexes containing one molecule of RRE bound to 2, 4, 5 or 6 Rev^{WT} monomers (Figure 4.26 E, F). Low intensity signals

corresponding to free RRE were also detected, whereas a peak corresponding to a RRE:Rev_3 complex was notably absent.

A

Label	RRE	Rev*	Measured Mass ± error	Calculated Mass
a	1	1	130,618 ± 3	130,617.5
b	1	2	143,785 ± 3	143,785.4
c	1	3	156,954 ± 2	156,953.3
d	1	4	170,123 ± 1	170,121.2



E

Label	RRE	Rev^{WT}	Measured Mass ± error	Calculated Mass
A	1	2	143,801 ± 2	143,801.0
B	1	4	170,152 ± 3	170,152.4
C	1	5	183,329 ± 3	183,328.1
D	1	6	196,502 ± 5	196,503.8

Figure 4.26: RRE/Rev* and RRE/Rev^{WT} complexes analyzed by native MS. A) and E) Summary of complexes detected. Samples were incubated in 200 mM AmAc. A) Spectrum acquired with RRE at 2 µM and a [Rev*]/[RRE] ratio of 1:17.5. A RRE/Rev* complex with a stoichiometry of 1:1 (a) was detected. B) Spectrum of RRE at 1.75 µM and a [Rev*]/[RRE] ratio of 1:20. RRE/Rev* complexes with stoichiometry of 1:1 (a), 1:2 (b) and 1:3 (c) were detected. C) Spectrum acquired with RRE at 1.75 µM and a [Rev*]/[RRE] ratio of 1:15. RRE/Rev* complexes with stoichiometry of 1:1 (a), 1:2 (b), 1:3 (c) and 1:4 (d) were detected. F) Spectrum acquired with RRE at 2.5 µM and a [Rev^{WT}]/[RRE] ratio of 1:12. RRE/Rev^{WT} complexes with a stoichiometry of 1:2 (A), 1:4 (B), 1:5 (C) and 1:6 (D) were detected.

From the above data we conclude that the RRE can accommodate at least 6 copies of Rev^{WT}. These experiments establish the feasibility of analysing HIV-1 RNP complexes by native MS. Moreover, they suggest that native MS spectra of even higher quality can be achieved by further improving RRE sample quality and by additional optimization of RRE/Rev^{WT} complex assembly conditions.

5. CONCLUSIONS AND PERSPECTIVES

ABSTRACT

This project aimed to determine the stoichiometry of the RRE/Rev/CRM1/Ran RNP complex by native mass spectrometry. The use of a 66-nt RRE fragment (IIABC) to simplify the challenging study of this RNP has been critical to set up and optimize the sample preparation that allowed us to determine with great accuracy the stoichiometry of the IIABC containing RNPs (small RNP export complex). The key methodological breakthrough has been the preparation of RNA and RNPs molecules free from non-volatile salts and therefore compatible with native mass spectrometry analysis. Adapting the preparation to the full-length RRE we could also obtain informative spectra of RRE containing RNPs. Nevertheless we tackled several technical limits that hindered the success of analysis of the complete RRE/Rev/CRM1/Ran RNP complex. In conclusion this study shows the strengths and limitations of native mass spectrometry and its potential for future development as a tool for analyzing RNP complexes.

RÉSUMÉ

Ce projet visait à déterminer la stœchiométrie du complexe RRE/Rev/CRM1/Ran RNP par spectrométrie de masse native. Pour simplifier l'étude de cette RNP, l'utilisation d'un fragment du RRE de 66 nt (IIABC) a été critique pour mettre en place et optimiser la préparation des échantillons qui nous a permis de déterminer avec une grande précision la stœchiométrie des RNP contenant le IIABC (small RNP export complex). La découverte méthodologique clé a été la préparation de molécules d'ARN et des RNP exemptes de sels non volatiles et donc compatibles avec l'analyse de spectrométrie de masse native. En adaptant cette préparation aux RRE entier, nous avons également obtenu des spectres informatifs des RNP contenant le RRE. Néanmoins, nous nous sommes heurtés à plusieurs limites techniques qui ont entravé le succès de l'analyse du complexe entier RRE/Rev/CRM1/Ran. En conclusion, cette étude montre les points forts et les limites de la spectrométrie de masse native et son potentiel de développement futur en tant qu'outil d'analyse des complexes RNP.

5.1 RECAP OF THE RESULTS

The goal of my PhD project was to investigate the architecture of the RRE/Rev/CRM1/Ran^{GTP} complex assembled *in vitro*. More specifically, the primary objective was to determine the Rev:RRE and Rev:CRM1 stoichiometry using native mass spectrometry. To achieve this goal I had to overcome several technical challenges. The approach, as explained in the overall strategy (§ 4.1), was to dissect the full RNP export complex into smaller RNP complexes (Figure 4.1) to facilitate the analysis.

As expertise in the production of Rev and RNA was lacking in the laboratory, I spent the first year of my PhD optimizing the preparation of these samples. By initially gaining experience with two oligomerization-defective Rev mutants (Rev* and Rev^{V16D}), which are more soluble than the wild-type form, I managed to establish a purification protocol for Rev^{WT} that allowed me to obtain sufficient amounts of this protein for biochemical and biophysical studies (§ 4.2.1.3). Working at the IBS Cell-Free technical platform with the help of Lionel Imbert and in collaboration with Marco Marcia (EMBL), I managed to generate and purify large quantities of the 66 nt RRE fragment IIABC as well as the entire 350 nt RRE (§ 4.2.2). I used electrophoretic mobility shift assays (EMSAs) to evaluate the assembly of different complexes *in vitro* and to optimize various experimental conditions (e.g., concentration of components, mixing order, buffer conditions). Using the 34 nt IIB fragment I verified the RNA-binding activity of the three forms of Rev and confirmed the functionality of the Rev NES motif by assembling the minimal quaternary IIB/Rev^{WT}/CRM1/Ran^{GTP} complex (§ 4.2.3).

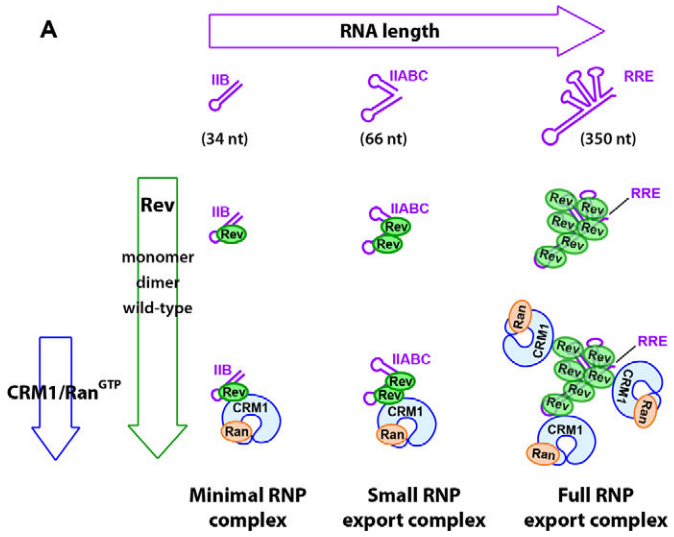
EMSA analyses confirmed that the three forms of Rev oligomerize on the IIABC fragment, with the observed binding cooperativity increasing with the number of functional Rev oligomerization interfaces (Rev* < Rev^{V16D} < Rev^{WT}) (§ 4.2.4.1). Interestingly, the different forms of Rev exhibited a difference in the relative mobility of complexes containing more than one Rev molecule. This finding suggests a possible difference in the conformation of the different IIABC/Rev complexes, leading us to hypothesize that IIABC adopts a more open conformation when bound to multiple molecules of Rev* compared to Rev^{V16D} and Rev^{WT}. IIABC/Rev complexes, assembled using different forms of Rev, can recruit CRM1 in the presence of Ran^{GTP}. As expected, the higher degree of cooperativity of

Rev^{V16D} and Rev^{WT} increases the sample heterogeneity compared to Rev*, making the former proteins more challenging to analyze (§ 4.2.4.2).

Working in close collaboration with Elisabetta Boeri Erba, I could successfully analyse the above complexes by native MS and determine their stoichiometry (Figure 5.1). First, I developed and extensively optimized a protocol to produce IIABC in a manner compatible with native MS analysis, resulting in high quality spectra that yielded an accurate mass for this RNA molecule (§ 4.3.2.2). Different IIABC/Rev complexes prepared with increasing ratios of Rev were analysed using the three forms of Rev. We could unambiguously establish that IIABC can accommodate up to six copies of Rev*, four of Rev^{V16D} and three of Rev^{WT} (§ 4.3.4, Figure 5.1 B). These data, combined with relative mobility analysis of the same complexes, strengthen the hypothesis that Rev^{WT} binding to IIABC induces a more compact conformation of the RNA, reducing accessibility of secondary binding sites to additional molecules of Rev.

Native MS analysis of the small RNP export complex was more challenging. Using Rev* I was able to assemble homogenous complexes and could demonstrate that a IIABC/$(Rev^*)_2$ complex can recruit only a single molecule of CRM1 (§ 4.3.5.1, Figure 5.1 C). With Rev^{V16D} and Rev^{WT} the mass assignments were more complicated because of sample polydispersity and the low intensity of signals. The most informative spectra obtained suggest that two molecules of CRM1 can be recruited to IIABC/Rev complexes containing 4 copies of Rev^{V16D} or 3-4 copies of Rev^{WT}. However, these data need further confirmation since we were unable to perform MS-MS analyses due to the low signal intensity (§ 4.3.5.3, Figure 5.1 C).

To analyze RRE-containing complexes I extensively optimized the preparation of the RNA sample until obtaining acceptable spectra for the unbound RRE, allowing me to subsequently proceed with native MS analysis on RRE/Rev complexes (§ 4.4.2.2). The data acquired show that the RRE can accommodate at least 4 molecules of Rev* and 6 molecules of Rev^{WT} (§ 4.4.3).



B

RRE	IIABC	Rev^{WT}	Rev^{V16D}	Rev*	Figure
-	1	1 to 3	-	-	4.19
-	2	2, 3, 6	-	-	4.19
-	1	-	1 to 4	-	4.18
-	1	-	-	1 to 6	4.17
1	-	2, 4, 5, 6	-	-	4.26
1	-	-	-	1 to 4	4.26

C

IIABC	Rev^{WT}	Rev^{V16D}	Rev*	CRM1	Ran^{GTP}	Figure
-	1	-	-	1	-	4.13
1	1	-	-	1	1	4.22
- or 1	3	-	-	2	2	4.22
2	3 or 4	-	-	2	1	4.22
2	4	-	-	3	3	4.22
-	-	1	-	1	1	4.21
1	-	4	-	2	1 or 2	4.21
-	-	-	1	1	1	4.13
1	-	-	1 or 2	1	1	4.20

Figure: 5.1: Stoichoimetry determination by native MS of HIV-1 RNP complexes. A) Schematic representation of the different viral RNPs used in this study. Summary of the stoichiometries of the different Rev/IIABC complexes (B) and HIV-1 RNP complexes (C) established by native MS during this study.

5.2 METHODOLOGICAL IMPROVEMENTS FOR THE ANALYSIS OF RNPS BY NATIVE MASS SPECTROMETRY

The key technical breakthrough made during this study was the preparation of a IIABC sample compatible with native MS analysis. The fact that IIABC is almost free of non-volatile salts allowed us to resolve signals corresponding to IIABC/Rev complexes that differed in mass by only one molecule of Rev, even when samples were highly heterogeneous, such as those obtained at the highest Rev* concentration tested (§ 4.3.4.1).

During this study we could also identify other critical parameters to obtain stable complexes for the acquisition of high-quality native MS spectra, such as the concentration, incubation time and mixing order of the different components, as well as the buffer composition. The concentration of IIABC and Rev mixed to prepare the sample, in addition to the molar ratio used, proved to be crucial to obtain stable complexes. In particular, we observed that to assemble stable IIABC/Rev* complexes, at low ratios, the initial concentration of components must be high, using IIABC in a concentration 10 to 15 times higher than the one recommended for the acquisition. Conversely, when working with Rev^{WT} in order to obtain stable samples and avoid aggregation, the concentration of components must be not so high. Concerning the incubation time we observed that IIABC

and Rev should be incubated for at least 20-30 minutes at room temperature before acquisition or addition of CRM1 and Ran, to obtain stable complexes. As general rules regarding the mixing order, we observed that in all cases Rev must be added after the IIABC and, when assembling the small RNP export complexes, it was important to add Ran before CRM1. The addition of DTT (1 µM) also greatly facilitates the spraying of CRM1-containing samples as well as the use of MgAcetate (1 to 2.5 µM) to stabilize Ran^{GTP}.

One significant limitation related to data analysis was our inability, due to the low signal intensity, to perform MS-MS so as to dissociate complexes in order to confirm their stoichiometry. In the future, technical developments in instrumentation may overcome this problem.

The protocol used to replace non-volatile salts bound to IIABC with volatile salts was applied to the RRE and this, in combination with obtaining high sample purity, proved to be crucial for obtaining spectra of satisfactory quality. The experimental mass of the RRE determined using denaturing and native MS was 116,884 and 117,034 Da, respectively. These values differ from the theoretical mass by ~1300-1400 Da, an increase that might be explained by the presence of ~40-45 cations still bound to the RRE (considering the average mass of Na^+ and K^+). Such a number is reasonable given that for IIABC, which is 5 times smaller, ESI-TOF analysis indicated the presence of ~5 cations still bound to IIABC after extensive desalting procedures.

The experience gained by working on the small RNP export complex allowed us to identify critical points for the analysis of the RRE/Rev RNP. However our attempts to obtain spectra for the RRE/Rev/CRM1/Ran^{GTP} complex to allow stoichiometry determination were unfruitful. Improving the RRE sample by removing more non-volatile salts is likely to be critical for obtaining such spectra. On the other hand the correct folding of the RRE might require the presence of cations such as Na^+, K^+ and Mg^{2+} and so their complete removal might hamper complex formation or stability. Indeed, SAXS analyses of a ~230 nt RRE-fragment showed an increasing degree of compactness in the presence of increased concentrations of Mg^{2+} (Fang et al., 2013). Thus, care needs to be taken in removing as many non-volatile cations as possible to permit analysis, but not too many so as to compromise the stability of the complex.

5.3 BIOLOGICAL RELEVANCE FOR HIV-1 RNP EXPORT

The "divide-and-conquer" approach, involving the use of different RRE fragments and Rev mutants, allowed us to successfully overcome several technical hurdles anticipated already at the start of this project. The advantage of investigating an RRE fragment such as IIABC is that it greatly facilitated native MS analysis of Rev- and CRM1/Ran-bound complexes. The disadvantage of using IIABC is that the data obtained might not be relevant to the full HIV-1 RNP export complex. Although no previous study has investigated the assembly of a Rev- and CRM1-bound complex using IIABC, our results are nevertheless compatible with previously reported findings and advance our understanding of the stoichiometry and architecture of the full-export complex.

Structural data show that a 39 nt RRE fragment containing the IIB site can accommodate two molecules of a Rev dimeric mutant (Jayaraman et al., 2014). We therefore expected IIABC to accommodate at least two molecules of Rev^{WT} and Rev^{V16D}, a mainly dimeric form of Rev. This was indeed confirmed by native MS (§ 4.3.4.2 and 4.3.4.3).

Comparing the monomeric, dimeric and wild-type forms of Rev reveals that the more oligomerization interfaces are disrupted, the more Rev molecules can bind IIABC. This was initially puzzling to us, but subsequently the analyses of the relative electrophoretic mobility of Rev/IIABC complexes combined with the stoichiometry deduced from native MS data suggested that the binding of Rev^{WT} might induce IIABC to adopt a more compact conformation. In the context of the entire RRE, such conformational changes may be required to properly guide each Rev monomer to the appropriate binding site during the multimerization process; i.e., the RRE may undergo a stepwise series of Rev-induced conformational changes in a highly orchestrated manner. Indeed, the most recent model proposed for RRE/Rev assembly suggests that the oligomerization of Rev on the RRE induces a conformational change in the RRE that exposes additional cryptic secondary Rev binding sites (Bai et al., 2014).

Our analysis of the IIABC/Rev^{WT} complex revealed the presence of higher-order complexes containing 2 molecules of IIABC bound to 2, 3, 4 and 6 copies of Rev^{WT} (§ 4.3.4.3; figure 5.2). Although such complexes might be experimental artefacts, they nevertheless prove the ability of Rev^{WT} to bridge two distant RNA molecules. In addition,

we did not observe such higher-order complexes with Rev* and Rev^{V16D} suggesting that they are related to the intact oligomerization domain of Rev. Such a bridging hypothesis has so far only been inferred from structural data following the identification of a third type of Rev-Rev oligomerization interface, the C-C interface (DiMattia et al., 2016). It would be interesting to check whether mutations distrupting the C-C interface would compromise this bridging ability of Rev^{WT} and abrogate the formation of such higher-order complexes.

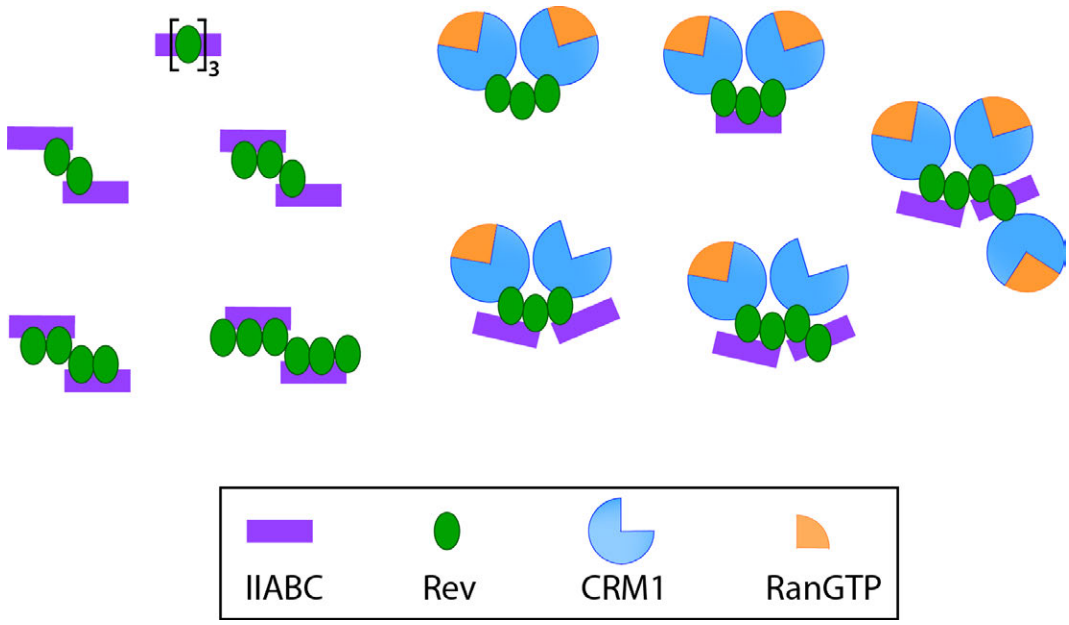


Figure: 5.2: Schematic representation of HIV-1 small RNP export complexes detected by native MS. Schematic representation of the stoichiometries determined by native MS from the analysis of the small RNP export complex assembled using Rev^{WT}. The different cartoons are identified in the box below.

Our data suggest that 2 CRM1 molecules can be recruited to IIABC/Rev complexes containing either 3 or 4 copies of Rev^{WT}, suggesting that an array of 3-4 Rev molecules spaces two NESs sufficiently far apart to allow the engagement of 2 CRM1 molecules without steric clashes (§ 4.3.5.3; figure 5.2). These results are in agreement with the discovery that a 242 nt RRE fragment bound to 6 molecules of Rev^{WT} can stably recruit two molecules of CRM1 *in vitro* (Booth et al., 2014). Moreover the analysis of the small RNP export complex using Rev^{WT} revealed the presence of unexpected complexes (Figure 5.2). For instance we determined masses of RNPs corresponding to a stoichiometry containing 2

copies of CRM1 but only one Ran^{GTP} or protein complexes containing 2 molecules of CRM1 bound to 3 copies of Rev^{WT} in absence of IIABC. We suggest that such complexes are stabilized by the dimerization interface of the two CRM1 molecules, once they have been recruited on the array of Rev^{WT}, interaction that can compensate the absence of Ran^{GTP} or IIABC.

Another important aspect is the heterogeneity of the small RNP export complex. Although these complexes are obtained *in vitro* and using a fragment of the RRE, we couldn't help speculating if multiple complexes with different stoichiometries may exist also *in vivo*. Indeed the RRE is present in two classes of viral RNA, ~4kb or ~9kb long, and the export of these different molecules may involve a different number of copies of Rev and CRM1.

5.4 PERSPECTIVES FOR FUTURE WORK

Up to date no atomic resolution structure of the RRE/Rev/CRM1/Ran^{GTP} is available, studies mostly limited by the high dynamic and heterogeneity of this RNP. Although not the aim of this study, our data obtained with the IIABC showed the feasibility to use such RRE fragment to assemble *in vitro* a HIV-1 RNP complex suitable for structural studies using cryo-EM or X-ray crystallography. Functional studies demonstrated the importance of the C-terminal of Rev in the interaction with CRM1, suggesting that this portion contains a second epitope on the viral protein involved in the specific recognition of the exportin (Hakata et al., 2002). Define at the atomic level the still unknown interactions between Rev and CRM1 would open a window for an *in silico* screening of molecules in order to specifically target this step of the viral cycle.

To continue investigating the stoichiometry of the RRE/Rev/CRM1/Ran^{GTP} complex by native MS, a good compromise between technical limitations and biological relevance might be to work with the ~230 nt RRE-fragment, which is known to retain export activity *in vivo* (Fang et al., 2013; Mann et al., 1994). By applying the purification protocol we developed, it is possible that a ~230 nt RRE might bind up to ~2 times fewer cations than the full-length RRE, significantly increasing the resolution of the native MS spectra.

Our difficulties in analysing the full RNP export complex might also be related to the absence of other cellular co-factors required during the export of the viral RNP. For

instance, RanBP3 is a nuclear co-factor of CRM1 known to enhance the binding of certain cargoes, including Rev (Englmeier et al., 2001; Lindsay et al., 2001). Another co-factor playing a role in the RRE/Rev assembly process is the DEAD-box helicase DDX1, which has been shown to promote Rev oligomerization on the RRE *in vitro* (Hammond et al., 2017). It is conceivable that the addition of these proteins to the complex might lead to a significant increase in conformational and compositional homogeneity that would yield cleaner native MS spectra.

The preparation of IIABC and RRE in a compatible manner with native MS analysis is a huge step forward in the field and we can anticipate that this method will be successfully applied to RNA molecules of similar size, allowing and improving the analysis of many other RNPs. In the long term one might reasonably expect that technical improvements, resulting in more sensitive instruments, would also make the analysis of such complexes more tractable. Another technical aspect that might be improved is the approach used for data analysis of RNA containing complexes. For some RNPs or for longer RNA molecules it might be not possible to remove all the non-volatile salts without compromising the correct RNA folding. Defining a way to calculate the mass of residual non-volatile salts still bound to the RNA molecule and including such mass in the data analysis would greatly improve data interpretation. Ideally, one would hope that in the not-too-distant future it will become possible to perform native MS on such RNP complexes which are endogenously expressed in cell or tissue samples, as for example has recently been reported for endogenous multi-protein assemblies (Olinares et al., 2016).

In conclusion, my thesis work has demonstrated the feasibility of analysing small- to medium-sized RNP complexes by native MS. Future applications of this technique to RNP assemblies of ever-increasing size and complexity will no doubt become possible as additional improvements are made to sample preparation protocols combined with technical advances in instrumentation and data analysis.

BIBLIOGRAPHY

Aligeti, M., Behrens, R.T., Pocock, G.M., Schindelin, J., Dietz, C., Eliceiri, K.W., Swanson, C.M., Malim, M.H., Ahlquist, P., and Sherer, N.M. (2014). Cooperativity among Rev-Associated Nuclear Export Signals Regulates HIV-1 Gene Expression and Is a Determinant of Virus Species Tropism. Journal of Virology 88, 14207–14221.

Andrade, M.A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. Nat Genet 11, 115–116.

Andrade, M.A., Petosa, C., O'Donoghue, S.I., Müller, C.W., and Bork, P. (2001). Comparison of ARM and HEAT protein repeats1. Journal of Molecular Biology 309, 1–18.

von Appen, A., and Beck, M. (2016). Structure Determination of the Nuclear Pore Complex with Three-Dimensional Cryo electron Microscopy. Journal of Molecular Biology 428, 2001–2010.

Arts, E.J., and Hazuda, D.J. (2012). HIV-1 Antiretroviral Drug Therapy. Cold Spring Harbor Perspectives in Medicine *2*, a007161–a007161.

Bai, Y., Tambe, A., Zhou, K., and Doudna, J.A. (2014). RNA-guided assembly of Rev-RRE nuclear export complexes. Elife 3, e03656.

Bartel, D.P., Zapp, M.L., Green, M.R., and Szostak, J.W. (1991). HIV-1 rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell 67, 529–536.

Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). α Helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex. Science 273, 1547.

Beck, M., and Hurt, E. (2016). The nuclear pore complex: understanding its function through structural insight. Nature Reviews Molecular Cell Biology 18, 73–89.

Bernad, R., van der Velde, H., Fornerod, M., and Pickersgill, H. (2004). Nup358/RanBP2 Attaches to the Nuclear Pore Complex via Association with Nup88 and Nup214/CAN and Plays a Supporting Role in CRM1-Mediated Nuclear Protein Export. Molecular and Cellular Biology 24, 2373–2384.

Bischoff, F.R., and Görlich, D. (1997). RanBP1 is crucial for the release of RanGTP from importin β-related nuclear transport factors. FEBS Letters 419, 249–254.

Bischoff, F.R., and Ponstingl, H. (1991). Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. Nature 354, 80–82.

Bischoff, F.R., Klebe, C., Kretschmer, J., Wittinghofer, A., and Ponstingl, H. (1994). RanGAP1 induces GTPase activity of nuclear Ras-related Ran. Proceedings of the National Academy of Sciences 91, 2587–2591.

Booth, D.S., Cheng, Y., and Frankel, A.D. (2014). The export receptor Crm1 forms a dimer to promote nuclear export of HIV RNA. Elife 3, e04121.

Burdisso, P., Suarez, I.P., Bologna, N.G., Palatnik, J.F., Bersch, B., and Rasia, R.M. (2012). Second double-stranded RNA binding domain of dicer-like ribonuclease 1: structural and biochemical characterization. Biochemistry 51, 10159–10166.

Charpentier, B., Stutz, F., and Rosbash, M. (1997). A dynamic in Vivo view of the HIV-IRev-RRE interaction. Journal of Molecular Biology 266, 950–962.

Chillón, I., Marcia, M., Legiewicz, M., Liu, F., Somarowthu, S., and Pyle, A.M. (2015). Native Purification and Analysis of Long RNAs. Meth. Enzymol. 558, 3–37.

Chook, Y.M., and Blobel, G. (1999). Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. Nature 399, 230–237.

Cochrane, A.W., Chen, C.H., Kramer, R., Tomchak, L., and Rosen, C.A. (1989). Purification of biologically active human immunodeficiency virus rev protein from Escherichia coli. Virology 173, 335–337.

Cole, J.L., Gehman, J.D., Shafer, J.A., and Kuo, L.C. (1993). Solution oligomerization of the rev protein of HIV-1: Implications for function. Biochemistry 32, 11769–11775.

Cook, A., Bono, F., Jinek, M., and Conti, E. (2007). Structural Biology of Nucleocytoplasmic Transport. Annual Review of Biochemistry 76, 647–671.

Cook, K.S., Fisk, G.J., Hauber, J., Usman, N., Daly, T.J., and Rusche, J.R. (1991). Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate. Nucleic Acids Res. 19, 1577–1583.

Cullen, B.R. (2003). Nuclear mRNA export: insights from virology. Trends in Biochemical Sciences 28, 419–424.

Cullen, B.R., and Malim, M.H. (1991). The HIV-1 Rev protein: prototype of a novel class of eukaryotic post-transcriptional regulators. Trends Biochem. Sci. 16, 346–350.

Daugherty, M.D., D'Orso, I., and Frankel, A.D. (2008). A Solution to Limited Genomic Capacity: Using Adaptable Binding Surfaces to Assemble the Functional HIV Rev Oligomer on RNA. Molecular Cell 31, 824–834.

Daugherty, M.D., Booth, D.S., Jayaraman, B., Cheng, Y., and Frankel, A.D. (2010a). HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. Proceedings of the National Academy of Sciences 107, 12481–12486.

Daugherty, M.D., Liu, B., and Frankel, A.D. (2010b). Structural basis for cooperative RNA binding and export complex assembly by HIV Rev. Nature Structural & Molecular Biology 17, 1337–1342.

Dian, C., Bernaudat, F., Langer, K., Oliva, M.F., Fornerod, M., Schoehn, G., Müller, C.W., and Petosa, C. (2013). Structure of a Truncation Mutant of the Nuclear Export Factor CRM1 Provides Insights into the Auto-Inhibitory Role of Its C-Terminal Helix. Structure 21, 1338–1349.

DiMattia, M.A., Watts, N.R., Stahl, S.J., Rader, C., Wingfield, P.T., Stuart, D.I., Steven, A.C., and Grimes, J.M. (2010). Implications of the HIV-1 Rev dimer structure at 3.2 A resolution for multimeric binding to the Rev response element. Proceedings of the National Academy of Sciences 107, 5810–5814.

DiMattia, M.A., Watts, N.R., Cheng, N., Huang, R., Heymann, J.B., Grimes, J.M., Wingfield, P.T., Stuart, D.I., and Steven, A.C. (2016). The Structure of HIV-1 Rev Filaments Suggests a Bilateral Model for Rev-RRE Assembly. Structure 24, 1068–1080.

Dong, X., Biswas, A., Süel, K.E., Jackson, L.K., Martinez, R., Gu, H., and Chook, Y.M. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. Nature 458, 1136–1141.

Drivas, G.T., Shih, A., Coutavas, E., Rush, M.G., and D'Eustachio, P. (1990). Characterization of four novel ras-like genes expressed in a human teratocarcinoma cell line. Molecular and Cellular Biology 10, 1793–1798.

Edgcomb, S.P., Aschrafi, A., Kompfner, E., Williamson, J.R., Gerace, L., and Hennig, M. (2008). Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev-RRE complexes. Protein Science 17, 420–430.

Elinav, H., Wu, Y., Coskun, A., Hryckiewicz, K., Kemler, I., Hu, Y., Rogers, H., Hao, B., Ben Mamoun, C., Poeschla, E., et al. (2012). Human CRM1 Augments Production of Infectious Human and Feline Immunodeficiency Viruses from Murine Cells. Journal of Virology 86, 12053–12068.

Engelsma, D., Bernad, R., Calafat, J., and Fornerod, M. (2004). Supraphysiological nuclear export signals bind CRM1 independently of RanGTP and arrest at Nup358. EMBO J. 23, 3643–3652.

Fadouloglou, V.E., Lin, H.-T.V., Tria, G., Hernández, H., Robinson, C.V., Svergun, D.I., and Luisi, B.F. (2015). Maturation of 6S regulatory RNA to a highly elongated structure. FEBS J. 282, 4548–4564.

Fang, X., Wang, J., O'Carroll, I.P., Mitchell, M., Zuo, X., Wang, Y., Yu, P., Liu, Y., Rausch, J.W., Dyba, M.A., et al. (2013). An Unusual Topological Structure of the HIV-1 Rev Response Element. Cell 155, 594–605.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64–71.

Fernandes, J.D., Booth, D.S., and Frankel, A.D. (2016). A structurally plastic ribonucleoprotein complex mediates post-transcriptional gene regulation in HIV-1: Structural Plasticity of the HIV-1 Export RNP. Wiley Interdisciplinary Reviews: RNA 7, 470–486.

Fischer, U., Huber, J., Boelens, W.C., Mattajt, L.W., and Lührmann, R. (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell 82, 475–483.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997a). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell 90, 1051–1060.

Fornerod, M., van Deursen, J., van Baal, S., Reynolds, A., Davis, D., Murti, K.G., Fransen, J., and Grosveld, G. (1997b). The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. EMBO J 16, 807.

Frankel, A.D., and Young, J.A.T. (1998). HIV-1: Fifteen Proteins and an RNA. Annu. Rev. Biochem. 67, 1–25.

Fu, S.-C., Huang, H.-C., Horton, P., and Juan, H.-F. (2013). ValidNESs: a database of validated leucine-rich nuclear export signals. Nucleic Acids Research 41, D338–D343.

Fung, H.Y.J., and Chook, Y.M. (2014). Atomic basis of CRM1-cargo recognition, release and inhibition. Seminars in Cancer Biology 27, 52–61.

Fung, H.Y.J., Fu, S.-C., Brautigam, C.A., and Chook, Y.M. (2015). Structural determinants of nuclear export signal orientation in binding to exportin CRM1. Elife 4, e10034.

Görlich, D., and Mattaj, I.W. (1996). Nucleocytoplasmic transport. Science 271, 1513–1518.

Görlich, D., Panté, N., Kutay, U., Aebi, U., and Bischoff, F.R. (1996). Identification of different roles for RanGDP and RanGTP in nuclear protein import. EMBO Journal 15, 5584–5594.

Grüter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., and Izaurralde, E. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Mol. Cell 1, 649–659.

Güttler, T., and Görlich, D. (2011). Ran-dependent nuclear export mediators: a structural perspective: RanGTPase-driven nuclear export. The EMBO Journal 30, 3457–3474.

Güttler, T., Madl, T., Neumann, P., Deichsel, D., Corsini, L., Monecke, T., Ficner, R., Sattler, M., and Görlich, D. (2010). NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. Nature Structural & Molecular Biology 17, 1367–1376.

Hakata, Y., Yamada, M., Mabuchi, N., and Shida, H. (2002). The Carboxy-Terminal Region of the Human Immunodeficiency Virus Type 1 Protein Rev Has Multiple Roles in Mediating CRM1-Related Rev Functions. Journal of Virology *76*, 8079–8089.

Hammarskjold, M.-L., and Rekosh, D. (2011). A Long-Awaited Structure Is Rev-ealed. Viruses 3, 484–492.

Hammond, J.A., Lamichhane, R., Millar, D.P., and Williamson, J.R. (2017). A DEAD-Box Helicase Mediates an RNA Structural Transition in the HIV-1 Rev Response Element. J. Mol. Biol. 429, 697–714.

Heaphy, S., Finch, J.T., Gait, M.J., Karn, J., and Singh, M. (1991). Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. Proceedings of the National Academy of Sciences of the United States of America 88, 7366–7370.

Hart, G.J., Orr, D.C., Penn, C.R., Figueiredo, H.T., Gray, N.M., Boehme, R.E., and Cameron, J.M. (1992). Effects of (-)-2'-deoxy-3'-thiacytidine (3TC) 5'-triphosphate on human immunodeficiency virus reverse transcriptase and mammalian DNA polymerases alpha, beta, and gamma. Antimicrobial Agents and Chemotherapy *36*, 1688–1694.

van den Heuvel, R.H.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J.J., Langridge, D., van der Oost, J., Hoyes, J., et al. (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Anal. Chem. 78, 7473–7483.

Hoffmann, D., Schwarck, D., Banning, C., Brenner, M., Mariyanna, L., Krepstakies, M., Schindler, M., Millar, D.P., and Hauber, J. (2012). Formation of Trans-Activation Competent HIV-1 Rev:RRE Complexes Requires the Recruitment of Multiple Protein Activation Domains. PLoS ONE 7, e38305.

de Hoffmann, E., and Stroobant, V. (2007). Mass Spectrometry : Principles and Applications (Wiley).

Hsu, A., Granneman, G.R., Cao, G., Carothers, L., El-Shourbagy, T., Baroldi, P., Erdman, K., Brown, F., Sun, E., and Leonard, J.M. (1998). Pharmacokinetic interactions between two human immunodeficiency virus protease inhibitors, ritonavir and saquinavir*. Clinical Pharmacology & Therapeutics *63*, 453–464.

Iwai, S., Pritchard, C., Mann, D.A., Karn, J., and Gait, M.J. (1992). Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. Nucleic Acids Research 20, 6465–6472.

Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I.W., and Görlich, D. (1997). The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. EMBO J 16, 6535.

Jain, C., and Belasco, J.G. (2001). Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. Molecular Cell 7, 603–614.

Jayaraman, B., Crosby, D.C., Homer, C., Ribeiro, I., Mavor, D., and Frankel, A.D. (2014). RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev–Rev response element complex. Elife 3, e04120.

Joseph, J. (2006). Ran at a glance. Journal of Cell Science 119, 3481–3484.

Kabachinski, G., and Schwartz, T.U. (2015). The nuclear pore complex - structure and function at a glance. Journal of Cell Science 128, 423–429.

Kalab, P., Weis, K., and Heald, R. (2002). Visualization of a Ran-GTP Gradient in Interphase and Mitotic Xenopus Egg Extracts. Science 295, 2452.

Kilby, J.M., Hopkins, S., Venetta, T.M., DiMassimo, B., Cloud, G.A., Lee, J.Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., et al. (1998). Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. Nat Med *4*, 1302–1307.

Kjems, J., and Sharp, P.A. (1993). The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6. U5 small nuclear ribonucleoprotein in spliceosome assembly. Journal of Virology 67, 4769–4776.

Kjems, J., Brown, M., Chang, D.D., and Sharp, P.A. (1991). Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. Proceedings of the National Academy of Sciences 88, 683–687.

Kondrat, F.D.L., Struwe, W.B., and Benesch, J.L.P. (2015). Native mass spectrometry: towards high-throughput structural proteomics. Methods Mol. Biol. 1261, 349–371.

Koyama, M., and Matsuura, Y. (2010). An allosteric mechanism to displace nuclear export cargo from CRM1 and RanGTP by RanBP1. The EMBO Journal 29, 2002–2013.

Kudo, N., Matsumori, N., Taoka, H., Fujiwara, D., Schreiner, E.P., Wolff, B., Yoshida, M., and Horinouchi, S. (1999). Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. Proceedings of the National Academy of Sciences 96, 9112–9117.

Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R., and Görlich, D. Export of Importin α from the Nucleus Is Mediated by a Specific Nuclear Transport Factor. Cell 90, 1061–1071.

Lounsbury, K.M., and Macara, I.G. (1997). Ran-binding Protein 1 (RanBP1) Forms a Ternary Complex with Ran and Karyopherin β and Reduces Ran GTPase-activating Protein (RanGAP) Inhibition by Karyopherin β. Journal of Biological Chemistry 272, 551–555.

Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997). A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell 88, 97–107.

Malim, M.H., and Cullen, B.R. (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: Implications for HIV-1 latency. Cell 65, 241–248.

Malim, M.H., Hauber, J., Le, S.-Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature 338, 254–257.

Mann, D.A., Mikaélian, I., Zemmel, R.W., Green, S.M., Lowe, A.D., Kimura, T., Singh, M., Jonathan, P., Butler, G., Gait, M.J., et al. (1994). A Molecular Rheostat. Journal of Molecular Biology 241, 193–207.

Matsuura, Y. (2016). Mechanistic Insights from Structural Analyses of Ran-GTPase-Driven Nuclear Export of Proteins and RNAs. Journal of Molecular Biology 428, 2025–2039.

Monecke, T., Guttler, T., Neumann, P., Dickmanns, A., Gorlich, D., and Ficner, R. (2009). Crystal Structure of the Nuclear Export Receptor CRM1 in Complex with Snurportin1 and RanGTP. Science 324, 1087–1091.

Monecke, T., Haselbach, D., Voss, B., Russek, A., Neumann, P., Thomson, E., Hurt, E., Zachariae, U., Stark, H., Grubmuller, H., et al. (2013). Structural basis for cooperativity of CRM1 export complex formation. Proceedings of the National Academy of Sciences 110, 960–965.

Monecke, T., Dickmanns, A., and Ficner, R. (2014). Allosteric control of the exportin CRM1 unraveled by crystal structure analysis. FEBS Journal 281, 4179–4194.

Mosammaparast, N., and Pemberton, L.F. (2004). Karyopherins: from nuclear-transport mediators to nuclear-function regulators. Trends in Cell Biology 14, 547–556.

Natalizio, B.J., and Wente, S.R. (2013). Postage for the messenger: designating routes for nuclear mRNA export. Trends Cell Biol. 23, 365–373.

Nemergut, M.E., Mizzen, C.A., Stukenberg, T., Allis, C.D., and Macara, I.G. (2001). Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. Science 292, 1540–1543.

Okamura, M., Inose, H., and Masuda, S. (2015). RNA Export through the NPC in Eukaryotes. Genes 6, 124–149.

Pallesen, J., Dong, M., Besenbacher, F., and Kjems, J. (2009). Structure of the HIV-1 Rev response element alone and in complex with regulator of virion (Rev) studied by atomic force microscopy. FEBS J. 276, 4223–4232.

Paraskeva, E., Izaurralde, E., Bischoff, F.R., Huber, J., Kutay, U., Hartmann, E., Lührmann, R., and Görlich, D. (1999). CRM1-mediated recycling of snurportin 1 to the cytoplasm. The Journal of Cell Biology 145, 255–264.

Peterlin, B.M., and Trono, D. (2003). Hide, shield and strike back: how HIV-infected cells avoid immune eradication. Nature Reviews Immunology 3, 97–107.

Petosa, C., Schoehn, G., Askjaer, P., Bauer, U., Moulin, M., Steuerwald, U., Soler-López, M., Baudin, F., Mattaj, I.W., and Müller, C.W. (2004). Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. Molecular Cell 16, 761–775.

Petosa, C. (2012). Ran. In Encyclopedia of Signaling Molecules, (Springer), pp. 1574–1581.

Pond, S.J., Ridgeway, W.K., Robertson, R., Wang, J., and Millar, D.P. (2009). HIV-1 Rev protein assembles on viral RNA one molecule at a time. Proceedings of the National Academy of Sciences 106, 1404–1408.

Port, S.A., Monecke, T., Dickmanns, A., Spillner, C., Hofele, R., Urlaub, H., Ficner, R., and Kehlenbach, R.H. (2015). Structural and Functional Characterization of CRM1-Nup214 Interactions Reveals Multiple FG-Binding Sites Involved in Nuclear Export. Cell Rep 13, 690–702.

Rausch, J., and Grice, S. (2015). HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. Viruses 7, 3053–3075.

Rexach, M., and Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. Cell 83, 683–692.

Ribbeck, K., Lipowsky, G., Kent, H.M., Stewart, M., and Görlich, D. (1998). NTF2 mediates nuclear import of Ran. EMBO J 17, 6587.

Richards, S.A., Lounsbury, K.M., Carey, K.L., and Macara, I.G. (1996). A nuclear export signal is essential for the cytosolic localization of the Ran binding protein, RanBP1. J. Cell Biol. 134, 1157–1168.

Saito, N., and Matsuura, Y. (2013). A 2.1-Å-Resolution Crystal Structure of Unliganded CRM1 Reveals the Mechanism of Autoinhibition. Journal of Molecular Biology 425, 350–364.

Saliou, J.-M., Manival, X., Tillault, A.-S., Atmanene, C., Bobo, C., Branlant, C., Van Dorsselaer, A., Charpentier, B., and Cianférani, S. (2015). Combining native MS approaches to decipher archaeal box H/ACA ribonucleoprotein particle structure and activity. Proteomics 15, 2851–2861.

Sato, M., Motomura, T., Aramaki, H., Matsuda, T., Yamashita, M., Ito, Y., Kawakami, H., Matsuzaki, Y., Watanabe, W., Yamataka, K., et al. (2006). Novel HIV-1 Integrase Inhibitors Derived from Quinolone Antibiotics. J. Med. Chem. *49*, 1506–1508.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Lührmann, R., and Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. EMBO J. 16, 3256–3271.

Sherer, N.M., Swanson, C.M., Hué, S., Roberts, R.G., Bergeron, J.R.C., and Malim, M.H. (2011). Evolution of a Species-Specific Determinant within Human CRM1 that Regulates the Post-transcriptional Phases of HIV-1 Replication. PLoS Pathogens 7, e1002395.

Sherpa, C., Rausch, J.W., Le Grice, S.F.J., Hammarskjold, M.-L., and Rekosh, D. (2015). The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. Nucleic Acids Research 43, 4676–4686.

Smith, A., Brownawell, A., and Macara, I.G. (1998). Nuclear import of Ran is mediated by the transport factor NTF2. Current Biology 8, 1403–1406.

Sobott, F., Hernández, H., McCammon, M.G., Tito, M.A., and Robinson, C.V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. Anal. Chem. 74, 1402–1407.

Sodroski, J., Goh, W.C., Rosen, C., Dayton, A., Terwilliger, E., and Haseltine, W. (1986). A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature 321, 412–417.

Spence, R., Kati, W., Anderson, K., and Johnson, K. (1995). Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. Science *267*, 988.

Stewart, M., Kent, H.M., and McCoy, A.J. (1998). The structure of the Q69L mutant of GDP-ran shows a major conformational change in the switch II loop that accounts for its failure to bind nuclear transport factor 2 (NTF2)1. Journal of Molecular Biology 284, 1517–1527.

Symensma, T.L., Baskerville, S., Yan, A., and Ellington, A.D. (1999). Polyvalent Rev decoys act as artificial Rev-responsive elements. Journal of Virology 73, 4341–4349.

Tan, R., Chen, L., Buettner, J.A., Hudson, D., and Frankel, A.D. (1993). RNA recognition by an isolated alpha helix. Cell 73, 1031–1040.

Thomas, S.L., Oft, M., Jaksche, H., Casari, G., Heger, P., Dobrovnik, M., Bevec, D., and Hauber, J. (1998). Functional analysis of the human immunodeficiency virus type 1 Rev protein oligomerization interface. Journal of Virology 72, 2935–2944.

Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G., and Cullen, B.R. (1992). Identification of a high-affinity RNA-binding site for the human immunodeficiency virus type 1 Rev protein. Proceedings of the National Academy of Sciences 89, 758–762.

Tsamis, F., Gavrilov, S., Kajumo, F., Seibert, C., Kuhmann, S., Ketas, T., Trkola, A., Palani, A., Clader, J.W., Tagat, J.R., et al. (2003). Analysis of the Mechanism by Which the Small-Molecule CCR5 Antagonists SCH-351125 and SCH-350581 Inhibit Human Immunodeficiency Virus Type 1 Entry. Journal of Virology *77*, 5201–5208.

Van Ryk, D.I., and Venkatesan, S. (1999). Real-time kinetics of HIV-1 Rev-Rev response element interactions. Definition of minimal binding sites on RNA and protein and stoichiometric analysis. J. Biol. Chem. 274, 17452–17463.

Vercruysse, T., and Daelemans, D. (2013). HIV-1 Rev multimerization: mechanism and insights. Current HIV Research 11, 623–634.

Vetter, I.R., Arndt, A., Kutay, U., Görlich, D., and Wittinghofer, A. (1999). Structural View of the Ran–Importin β Interaction at 2.3 Å Resolution. Cell 97, 635–646.

Wen, W., Meinkotht, J.L., Tsien, R.Y., and Taylor, S.S. (1995). Identification of a signal for rapid export of proteins from the nucleus. Cell 82, 463–473.

Wilm, M., and Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. Anal. Chem. 68, 1–8.

Wingfield, P.T., Stahl, S.J., Payton, M.A., Venkatesan, S., Misra, M., and Steven, A.C. (1991). HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. Biochemistry 30, 7527–7534.

Xu, D., Grishin, N.V., and Chook, Y.M. (2012). NESdb: a database of NES-containing CRM1 cargoes. Molecular Biology of the Cell 23, 3673–3676.

Zemmel, R.W., Kelley, A.C., Karn, J., and Butler, J.G. (1996). Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. Journal of Molecular Biology 258, 763–777.