Dissertation

in partial fulfilment of the requirements
for the degree “Doctor rerum naturalium (Dr. rer. nat)”
in the Molecular Biology Program
at the Georg August University Göttingen,
Faculty of Biology

Exploration of cargo spectrum and NES patterns recognized by the exportin CRM1

submitted by
Koray Kırlı
from Istanbul (Turkey)

Göttingen, September 2013

Members of the Thesis Committee

Prof. Dr. Dirk Görlich	Max Planck Institute for Biophysical Chemistry
(Thesis mentor and referee)	Dept. of Cellular Logistics
	Göttingen, Germany
Prof. Dr. Reinhard Jahn	Max Planck Institute for Biophysical Chemistry
(Co-referee)	Dept. of Neurobiology
	Göttingen, Germany
Prof. Dr. Kai Tittmann	Department of Bioanalytics
	Georg August University
	Göttingen, Germany

Additional Members of the Examination Board

Prof. Dr. Detlef Doenecke	Universitätsmedizin (UMG)
	Georg August University
	Göttingen, Germany
Prof. Dr. Peter Rehling	Dept. of Biochemistry II
	Georg August University
	Göttingen, Germany
Prof. Dr. Holger Stark	Max Planck Institute for Biophysical Chemistry
	3D Electron Cryo-Microscopy Research Group
	Göttingen, Germany

Date of the disputation: October 29, 2013

The research detailed in this thesis has been performed in the laboratory of Prof. Dr. Dirk Görlich at the Max-Planck-Institute for Biophysical Chemistry in the time from October 2009 to September 2013.

I hereby declare that I completed my thesis entitled

"Exploration of cargo spectrum and NES patterns recognized by the exportin CRM1"

independently and with no other sources and aids than quoted. This dissertation has not been submitted elsewhere for any academic award or qualification.

Koray Kırlı

Göttingen, September 2013

This thesis is dedicated to the ones that suffered

for a brighter future

A.I.K.

M.A.

M.Y.

A.C.

A.A.

E.S.

.

1 TABLE OF CONTENTS

1 TABLE OF CONTENTS 1
2 ACKNOWLEDGEMENTS 4
3 SUMMARY 5
4 LIST OF FIGURES 6
5 LIST OF TABLES 7
6 INTRODUCTION 8
6.1 THE BORDER AND THE GATES 8
6.2 TRANSPORT THROUGH THE NPCs 11
6.2.1 Passive Diffusion Through the NPCs 11
6.2.2 Facilitated Active Transport 12
6.3 DIRECTIONALITY OF THE TRANSPORT 12
6.4 NUCLEAR TRANSPORT RECEPTORS 18
6.4.1 CRM1/Exportin 1 20
6.5 NUCLEAR EXPORT SIGNALS 23
7 RESULTS 27
7.1 A NEW NES PREDICTION ALGORITHM 27
7.1.1 NES Consensus 28
7.1.2 NES Score 29
7.1.3 Disorder Propensities 30
7.1.4 Evaluation of PKI-type NES prediction 31
7.1.5 Evaluation of REV-type NES prediction 35
7.2 IDENTIFICATION OF NES ON eIF2β 36
7.2.1 Prediction of eIF2β NES hits 36
7.2.2 Validation of eIF2β NES hit 36
7.3 IDENTIFICATION OF NES ON spRna1p 41
7.3.1 Previously suggested NESs are buried in the structure 41
7.3.2 Prediction of spRna1p NES hits 42
7.3.3 Validation of spRna1p NES hit 42
7.4 RANGTP DEPENDENT CRM1 BINDERS FROM CYTOSOLIC HELA EXTRACT 49
7.4.1 CRM1 affinity chromatography of cytoplasmic HeLa extract 49

7.4.2 CRM1 Affinity Chromatography for SILAC-Based Mass Spectrometry 50
7.4.3 Mass Spectrometry Analysis of SILAC CRM1 Affinity Chromatography Eluates 53
8 DISCUSSION 61
8.1 A NEW PREDICTION ALGORITHM FOR CRM1 DEPENDENT NESs 61
8.2 MASS SPECTROMETRY ANALYSIS OF HELA CRM1 CARGOES 65
9 OUTLOOK 68
10 MATERIAL AND METHODS 69
10.1 INSTRUMENTS 69
10.2 PREPARATION OF DNA CONSTRUCTS 70
10.2.1 Primer Design 70
10.2.2 Polymerase Chain Reaction (PCR) 70
10.2.3 Mutagenesis PCR 71
10.2.4 DNA Cleavage with Restriction Enzymes 73
10.2.5 DNA Gel Electrophoresis 73
10.2.6 DNA Extraction From Agarose Gels 73
10.2.7 Determination of DNA Concentration 74
10.2.8 Ligation of DNA Fragments into Vectors 74
10.2.9 Electroporation of *E. coli* Cells 74
10.2.10 *E. coli* Strains 74
10.2.11 DNA Purification From *E. coli* Cultures 75
10.2.12 Bacterial Expression Constructs 75
10.3 PROTEIN EXPRESSION AND PURIFICATION 75
10.3.1 Native Protein Expression and Purification 75
10.3.2 Determination of Protein Concentrations 77
10.3.3 SDS-PAGE 77
10.4 BINDING ASSAYS WITH CRM1 79
10.4.1 Binding Assays with Purified Components 79
10.5 PULL DOWN FROM CYTOPLASMIC EXTRACT WITH CRM1 81
10.5.1 Preparation of Cytoplasmic Extracts 81
10.5.2 Preparation of Cytoplasmic SILAC HeLa Extracts 82
10.5.3 CRM1 Affinity Chromatography with Cytoplasmic HeLa Extracts 82
10.5.4 Mass Spectrometry Analysis of Elution Fractions 84
10.5.5 Data and Bioinformatics Analysis 85
10.6 TRANSIENT HELA CELL TRANSFECTIONS 86
11 ABBREVIATIONS 87

12 APPENDIX 89
12.1 PYTHON SCRIPTS 89
12.1.1 PKI-type NES Prediction 89
12.1.2 REV-type NES Prediction 93
12.2 PROTEIN IDENTIFIERS OF THE DATA SETS 96
12.2.1 NESdb Proteins 96
12.2.2 LMB Study Proteins 97
12.2.3 1265 Significant Hits of SILAC MS Data 97
12.2.4 321 Proteins with a Predicted NES 104
13 REFERENCES 106

2 ACKNOWLEDGEMENTS

I would like to thank

Dirk Görlich, who has been a great mentor, for giving me the opportunity to work with him for my thesis. His enthusiasm for science amazed and motivated me through out my doctoral work.

Reinhard Jahn and Kai Tittmann for being part of my thesis committee and for their contributions with valuable comments and discussions.

Henning Urlaub, Samir Karaca, Uwe Plessmann and Monika Raabe for the mass spectrometric analysis.

Samir Karaca for his suggestions and contributions with SILAC based method, and data analysis.

Bastian Hülsmann and Matthias Samwer for the lengthy discussions that broadened by scientific interest, and also for their friendship that made it easy to deal with the down times.

Steffen Frey for being the “room of requirements” of the lab, with many reagents and advices, and answering many questions with a great patience.

Miroslav Nikolov for kindly providing cytoplasmic SILAC HeLa extracts.

Heinz-Jürgen Dehne, Heike Behr, and Jürgen Schünemann for helping me at different stages of my doctoral studies with their excellent technical support.

Gabriele Kopp and Uwe Hoffmann for making the lab a ready to use place, by taking care of many things that one does not easily recognize.

Aksana Labokha for her supervision in the first stages of my doctoral studies.

members of the lab for fruitful discussions during lab and literature seminars, and for the cheerful atmosphere in the lab.

Steffen Burkhardt and Kerstin Grüniger for administrative support,

Metin Aksu, Kevser Gençalp and Samir Karaca that constituted the small Turkish district of the lab with me, for long scientific and non-scientific discussions during our coffee breaks, and also for always keeping me one step closer to home.

my family for their continuous support through out my studies.

Finally, I would like to thank Sinem K. Saka who has been the main inspiration to me, and was always there to cheer me up.

3 SUMMARY

The nuclear envelope (NE) subdivides eukaryotic cells into a nuclear and a cytoplasmic compartment, forcing material exchange between these two compartments to proceed through the nuclear pore complexes (NPCs). While proteins smaller than 30-40 kDa can passively diffuse through the NPCs, larger objects require nuclear transport receptors (NTRs) for efficient transport. NTRs have the privilege of facilitated NPC-passage; they bind transport cargoes and transfer them from one side of the NE to the other. NTRs can act as unidirectional cargo pumps, whereby they utilize the chemical potential of the nucleocytoplasmic RanGTP gradient with high nuclear and low cytoplasmic RanGTP levels.

CRM1 is a major, essential and highly conserved nuclear export receptor. It exports a great variety of cargoes from the nucleus to the cytoplasm. CRM1 also keeps e.g. several translation factors and RanGAP cytoplasmic. The latter is required for maintaining the nucleocytoplasmic RanGTP gradient. CRM1 recognizes many cargoes through so-called leucine-rich nuclear export signal (NES), sequences containing 4-5 hydrophobic residues in a 14-15 residues long stretch. Although NESs are described in the context of primary protein structure, a reliable NES prediction has been a challenge and failed, e.g. for eIF2β and Rna1p (*S.pombe* RanGAP).

Here we present a new NES prediction algorithm based on the recent crystal structures of different NES sequences with CRM1. We classified NES two PKI-type and REV-type with two different consensus definitions. PKI-type NES were graded for CRM1 binding strength and additional filtering was applied with disorder prediction. The REV-type NES was a novel classification based on Rev protein NES, and we show that there are several other examples of this type of NES. The estimation power of the new prediction algorithm was shown on prediction of already known NESs as control, and it also was able to predict the NESs of human eIF2β and S.pombe Rna1p, which was also confirmed experimentally.

Another challenge had been the question of how many different cargo species are actually transported by CRM1. To address this, we optimized affinity chromatography on immobilized CRM1 and used it to retrieve RanGTP-dependent cargoes from a cytoplasmic HeLa extract. This analysis revealed hundreds of new CRM1 cargo candidates, which were further group into functional protein categories. Most of the ribosomal proteins are found in our dataset. Besides them, we find serine threonine kinases, ATP dependent helicases, spliceosomal proteins, translation initiation factors, actin regulators, and E3 ubiquitin ligases. Proteins of metabolic pathways, cell adhesion, phagosome, and proteasome are excluded from the data set.

4 LIST OF FIGURES

Figure 6-1 First known drawing of cells and nuclei by Antonie van Leeuwenhoek, ... 8
Figure 6-2 Organization of NPCs on NE 9
Figure 6-3 Structural elements of vertebrate nucleoporins 10
Figure 6-4 Overview of active nucleocytoplasmic transport through NPCs 14
Figure 6-5 Domain organization of RanGAP homologs 15
Figure 6-6 Comparison of RanGTP and RanGDP structures 17
Figure 6-7 Exportin structures with respective cargoes. 20
Figure 6-8 Free and cargo bound states of CRM1 21
Figure 6-9 Structural definition of NES consensus 26
Figure 7-1 Regular expression for PKI-type NES pattern and an example NES 28
Figure 7-2 Regular expression for REV-type NES pattern and an example NES 29
Figure 7-3 Scoring Scheme for PKI-type NESs 30
Figure 7-4 An output example from PKI- type NES prediction 32
Figure 7-5 NES Scores of the PKI-type hits of the 11 selected proteins 33
Figure 7-6 Aggregated disorder propensities of the PKI-type hits of the 11 selected proteins 34
Figure 7-7 NES scores and disorder propensities of each PKI-type NES hit 34
Figure 7-8 Prediction of NES hits of human eIF2β 36
Figure 7-9 Alignment of hs eIF2β protein sequence with vertebrate orthologs 37
Figure 7-10 RanGTP dependent mmCRM1 binding of eIF2β NES hits 38
Figure 7-11 RanGTP dependent mmCRM1 binding of wild type and NES mutant of human eIF2β 39
Figure 7-12 spRna1p structure and previously suggested NESs 41
Figure 7-13 NES prediction for spRna1p 42
Figure 7-14 RanGTP dependent scCrm1 binding of spRna1p 43
Figure 7-15 RanGTP dependent scCrm1 binding of spRna1p and point mutants 45
Figure 7-16 Phenotypic outcomes of different GFP-NES fusions 46
Figure 7-17 Localization of different GFP spRna1p fusions 47
Figure 7-18 Salt sensitivity of RanGTP dependent mmCRM1 interaction of cargoes from HeLa S10 extract. 49
Figure 7-19 Distribution of protein ratios in heavy and light HeLa extracts 51
Figure 7-20 CRM1 affinity chromatography of cytoplasmic HeLa extracts produced with SILAC method 52
Figure 7-21 SILAC MS analysis of RanGTP dependent CRM1 binders 53
Figure 7-22 SILAC MS data with region markers 54
Figure 7-23 NESdb and LMB Sensitive Hits on SILAC MS data 56
Figure 7-24 Protein groups that are over or under represented in MS data set 57
Figure 10-1 BsaI Mediated Mutagenesis 72

5 LIST OF TABLES

Table 6-1 Mammalian importins and selected cargos..... 19
Table 6-2 Mammalian exportins and selected cargoes..... 19
Table 7-1 Validated NES containing proteins..... 33
Table 7-2 REV-type NESs from NESdb..... 35
Table 7-3 List of translation initiation factors from MS data..... 59
Table 7-4 List of translation elongation and release factors..... 60
Table 10-1 Laboratory Equipment..... 69
Table 10-2 Centrifuges and Rotors..... 69
Table 10-3 Sources of Coding Sequences..... 70
Table 10-4 PfuS Triple Mix Components..... 71
Table 10-5 PCR Reaction Steps..... 71
Table 10-6 Preferred Restriction Enzymes..... 73
Table 10-7 Bacterial Expression Constructs..... 75
Table 10-8 Gradient Gel Solutions..... 78
Table 10-9 Eukaryotic transfection constructs..... 86
Table 12-1 List of Uniprot IDs of human proteins of NESdb..... 96
Table 12-2 List of Uniprot IDs of human ortholog proteins of NESdb..... 97
Table 12-3 List of Uniprot IDs of proteins from LMB Study..... 97
Table 12-4 List of Uniprot IDs of proteins from MS analysis.....104
Table 12-5 321 Uniprot IDs of proteins with a predicted NES.....105

6 INTRODUCTION

The living things are interesting in many ways, and presumably their complex structure and organization is the most fascinating one for the scientists. Antonie van Leeuwenhoek's drawing of the salmon red blood cells marks one of the prominent moments of this fascination. Since the non-mammalian vertebrates retain their nuclei in the erythrocytes, by looking at them from his handcrafted microscope, he drew the first known figure of nucleus (Figure 6-1) (Delphis *et al.*, 1719). It was not called 'nucleus' until Botanist Robert Brown coined the term in 1831 (Oliver, 1913). Since then many important aspects of the nucleus and its function have been revealed, which also brought many new questions.

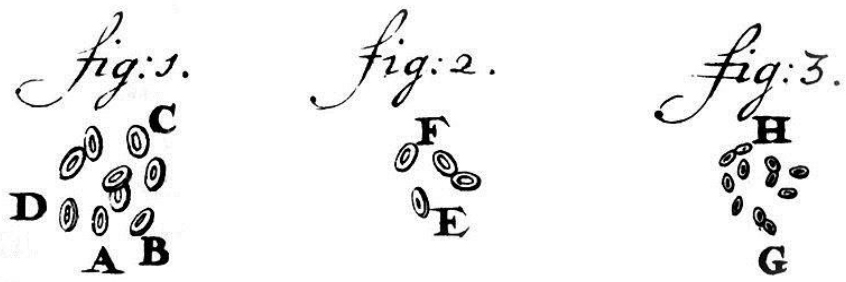

Figure 6-1 First known drawing of cells and nuclei by Antonie van Leeuwenhoek, 1719 (Delphis *et al.*, 1719)

6.1 THE BORDER AND THE GATES

The hallmark of a eukaryotic cell is its compartmentalization into the nucleus and the cytoplasm, which are separated by the nuclear envelope (NE). The nuclear envelope is a double lipid bilayer that is continuous on the cytoplasmic side with the membrane of the endoplasmic reticulum (ER). The perinuclear space in between these membranes is also part of the ER lumen (Subramanian and Meyer, 1997). The evolution of a nucleus enabled the high-end regulation that was required for emergence of very complex multicellular organism (Gorlich and Kutay, 1999).

One of the advantages of the nuclear compartmentalization is that eukaryotes can handle a large amount of genetic material compared to prokaryotes. This enables more coding sequence and regulatory sequence to be accommodated in the genome. A second advantage is the control over the level and the timing of gene expression by regulating the nuclear localization of transcription factors (Kaffman and O'Shea, 1999).

The more striking evolution that comes with the nucleus is the compartmentalization of the cellular information processes. In bacteria DNA, RNA and ribosomes and other proteins take role in conversion of genetic information to functional proteins, and work side by side in a continuous process. Ribosomes start translating the mRNA as soon as its 5' end is synthesized by the RNA

polymerase. In eukaryotes, the NE spatially separates transcription and translation and necessitates localization of specific macromolecules to these compartments.

Since all proteins are produced in the cytoplasm, proteins necessary for DNA maintenance (e.g., histones), transcription (e.g., RNA polymerases), gene expression regulation (e.g., transcription factors) and many others required in the nucleus have to be imported (Bonner, 1975). On the other hand, transcribed and spliced mRNA, tRNA, assembled ribosomal subunits have to be exported to the cytoplasm. Segregation of macromolecules between the nucleus and the cytoplasm poses another challenge for higher eukaryotes. During cell division the NE breaks down and re-forms during telophase. At this point a high load of misplaced macromolecules has to be re-sorted. Thus nucleocytoplasmic transport has to be a very efficient and fast process to keep up with this load (Gorlich and Kutay, 1999).

The need for export and import of cargoes originates not only from the requirement of certain macromolecules and complexes in a specific compartment, but also from the necessity that certain activities should be temporarily or permanently be absent in either the nucleus or the cytoplasm. For example regulation of a gene's expression might depend on import of a specific transcription factor. This regulation necessitates the temporary exclusion of the transcription factor from the nucleus, which can be sustained by nuclear export as in the case of NF-κB/IκBα complexes (Huang *et al.*, 2000).

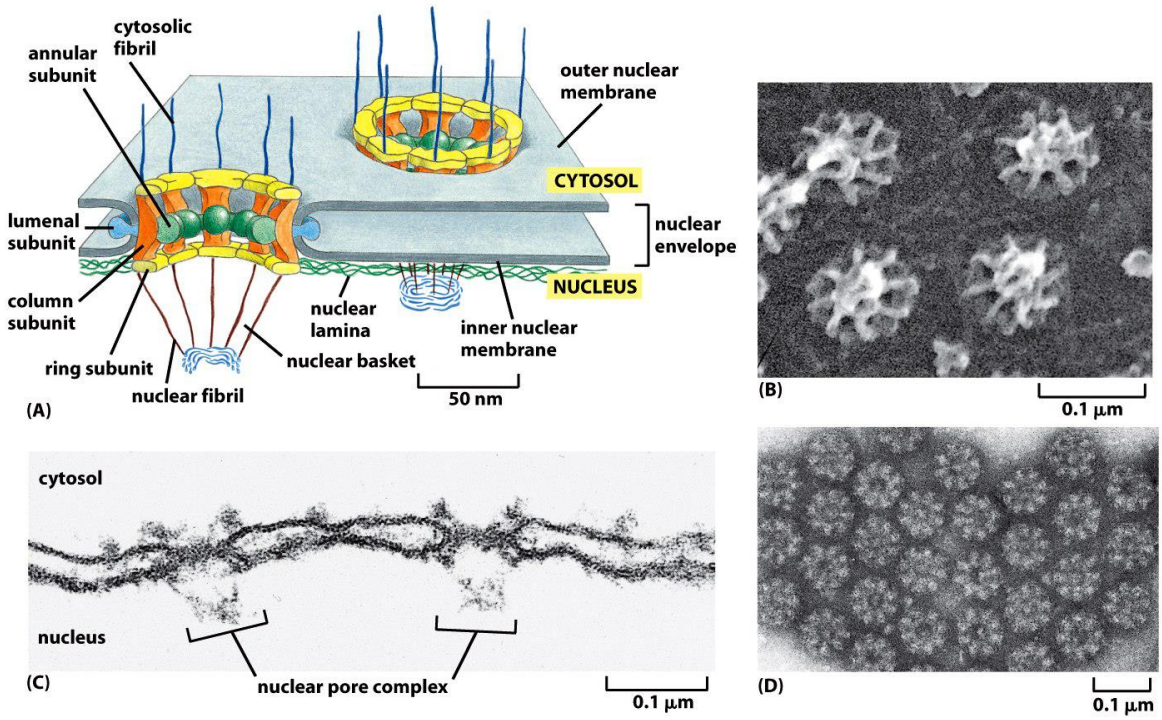

Figure 6-2 Organization of NPCs on NE (Alberts *et al.*, 2007)

(A) Depiction of NPC components. Electron micrograph of (B) NPCs from nuclear side of NE, (C) NPCs from side view, (D) NPCs from cytoplasmic side of NE.

Nuclear envelope is punctured by thousands of very large protein assemblies called nuclear pore complexes (NPCs), and NPCs are the main routes of transport between the nucleus and the cytoplasm (Figure 6-2). One of the first visible features of NPCs was its eight-fold symmetry (Watson, 1959). The total size of the NPC is estimated to be ~66 MDa in yeast (Rout and Blobel, 1993) and ~125 MDa in vertebrates (Reichelt *et al.*, 1990). NPCs restrict the diffusion of large proteins, which can be aided by nuclear transport receptors (NTRs) for NPC passage. It was shown that particles up to ~39 nm in diameter can pass through the NPCs (Pante and Kann, 2002; Au and Pante, 2012). These gigantic protein assemblies are made up of only ~30 different proteins (Figure 6-3) called nucleoporins (Nups) that exist in different copy numbers (Ori *et al.*, 2013). Structural organization of these proteins is still under debate with many proposed models (for a review of models see: Bilokapic and Schwartz, 2012).

Basically, Nups can be divided into two classes; the structural Nups that make up the ring like scaffold sitting on the NE, and Nups with unstructured regions that fill up the gap in the center of the ring and plug the pore. A more comprehensive depiction of structural elements of each vertebrate Nup can be seen on Figure 6-3 (Schwartz, 2005).

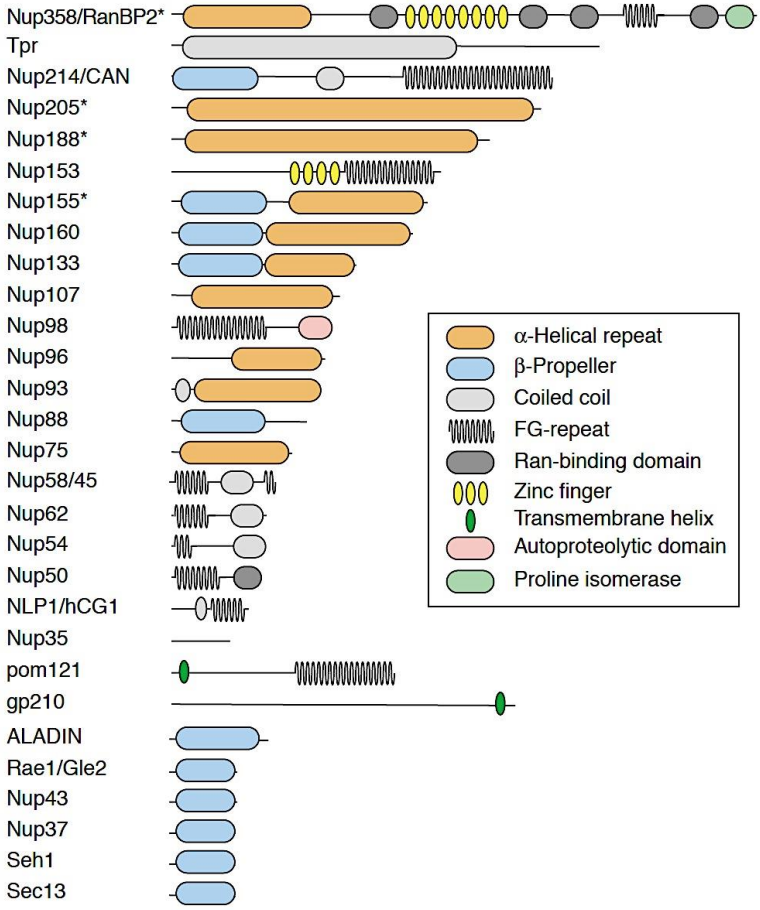

Figure 6-3 Structural elements of vertebrate nucleoporins (Schwartz, 2005)
*α-Helical regions predicted with high certainty

The second group is composed of proteins with dispersed phenylalanine-glycine repeats (FG repeats), which lacks a definite structure (Denning *et al.*, 2003). FG repeats are the sites of interaction with NTRs (Iovine *et al.*, 1995; Radu *et al.*, 1995; Rexach and Blobel, 1995; Bayliss *et al.*, 1999, 2000). Some FG repeat regions show self-interaction, and they can form hydrogels *in vitro*. FG hydrogels can reproduce the two essential functions of NPCs; blocking passively diffusing cargoes (inert cargoes) and enriching NTR· cargo complexes (Frey *et al.*, 2006; Frey and Gorlich, 2007, 2009; Labokha *et al.*, 2013). This observation is in line with the previously suggested "selective phase model" which suggests the NPC permeability barrier being a hydrogel made of a meshwork of FG domains (Ribbeck and Gorlich, 2001).

6.2 TRANSPORT THROUGH THE NPCs

NPCs are the main routes of macromolecule exchange between the nucleus and the cytoplasm. Cells invest quite some resources to maintain this exchange with many dedicated proteins that also involve abundant ones. Transport through the NPC differs from protein import into mitochondria, chloroplast or endoplasmic reticulum since proteins are transported through the NPC in a folded form and sometimes as complexes of different macromolecules.

NPCs efficiently block passive diffusion of inert molecules with a diameter ≥5 nm (Mohr *et al.*, 2009). While a small macromolecule like ubiquitin can pass through the NPC freely, a larger macromolecule would need the aid of nuclear transport receptors (NTRs) for efficient transport. Not only size but also charge may contributes to selectivity. Positively charged proteins were suggested to be excluded better than negatively charged proteins due to the positive net charge of the NPC channel proteins (Colwell *et al.*, 2010). The impressive examples of NTR cargoes with large size include ribosomal subunits, Balbiani ring particles, and intact viral capsids (Stevens and Swift, 1966; Franke and Scheer, 1974; Whittaker and Helenius, 1998; Au and Pante, 2012).

6.2.1 Passive Diffusion Through the NPCs

Passive diffusion of molecules between the cytoplasm and the nucleus proceeds through either the NE or the NPCs. Small organic substances like steroids, glycerol or ethanol can pass through the double lipid bilayer. As suggested by the "selective phase model", the FG meshwork in NPC possesses an aqueous passive diffusion barrier (Ribbeck and Gorlich, 2001). Passive diffusion through NPCs is fast for small molecules and does not require a special interaction with the NPC components. As the size gets larger passive diffusion becomes limiting for the efficient translocation; spherical proteins larger that 20-30 kDa (≥5nm in diameter) are already delayed for NPC passage (Paine *et al.*, 1975; Bonner, 1975; Mohr *et al.*, 2009).

6.2.2 Facilitated Active Transport

Macromolecules that cannot overcome the NPC barrier due to a large size or other features like charge are transported in a facilitated manner. Nuclear transport receptors (NTRs) are large molecules (90-150 kDa) that are able to shuttle between nucleus and cytoplasm. NTRs bind to cargoes and help them go through the barrier. Most NTRs are members of the Importin β (Impβ) superfamily. Their multivalent interactions with the FG meshwork allow them to enter the NPC barrier very efficiently. Most NTRs carry cargoes in one direction. They are called importins when they import cargoes from the cytoplasm to the nucleus, and exportins if exporting cargoes from the nucleus to the cytoplasm (Gorlich *et al.*, 1994; Fornerod *et al.*, 1997). Some NTRs, like Exportin 4, can function in both ways with different cargoes (Gontan *et al.*, 2009). NTR aided transfer is so efficient that up to 1000 translocations can take place in a single NPC per second (Ribbeck and Gorlich, 2001).

6.3 DIRECTIONALITY OF THE TRANSPORT

NTRs can shuttle between the cytoplasm and the nucleus, and can bind to their cargos, but these are not enough for a directional transport. For binding to its cargo in a compartment and releasing it in the other one, NTRs require means of sensing the location. All Imp-β like NTRs bind to a small guanine nucleotide binding protein called Ran (Gorlich *et al.*, 1997; Fornerod *et al.*, 1997). Ran stands for Ras-related nuclear protein and is a 25 kDa GTPase (Drivas *et al.*, 1990; Bischoff and Ponstingl, 1991; Melchior *et al.*, 1993a). The GTPase function enables Ran to switch between two states; the GTP bound active state (RanGTP) and the GDP bound silent state (RanGDP). RanGTP is the active state because it binds to Impβ-like NTRs while RanGDP does not. Nucleus and cytoplasm differ in their RanGTP concentration; the nucleus has 1000 fold RanGTP concentration than the cytoplasm (Gorlich *et al.*, 2003). This steep RanGTP gradient acts as the fuel of the directional transport (Gorlich *et al.*, 1996).

Exportins bind to their cargoes in the nucleus and assemble into export complexes with RanGTP, and in the cytoplasm, export complex is disassembled by involvement of other factors (explained below). Free exportin does not re-bind to its cargo, but returns to the nucleus and is ready for another round of transport. The export complex is formed by cooperative interaction, if one of the binders is present (RanGTP or cargo), affinity for the second one is increased, and upon binding complex is stabilized (Kutay *et al.*, 1997; Petosa *et al.*, 2004; Monecke *et al.*, 2013).

On the other hand, importins form complexes with their cargoes in the cytoplasm where RanGTP levels are very low. When import complex passes through the NPC, it is disassembled upon RanGTP binding to importin in the nucleus. RanGTP binding is strong and incompatible with cargo

binding, and prevents importin-cargo interaction. The importin· RanGTP complex returns to the cytoplasm and after dissociation of RanGTP, it is ready for the next cargo. Transport of NTRs alone or as complexes through the FG meshwork is reversible and does not require energy (Kose *et al.*, 1997; Nakielny and Dreyfuss, 1998; Schwoebel *et al.*, 1998; Ribbeck *et al.*, 1999; Englmeier *et al.*, 1999; Nachury and Weis, 1999; Zeitler and Weis, 2004).

Both for import and export cycles there is a net flux of RanGTP from the nucleus to the cytoplasm. Cells employ a transport receptor called nuclear transport factor 2 (NTF2) to efficiently carry RanGDP from the cytoplasm back to the nucleus (Ribbeck *et al.*, 1998). NTF2 is not an Impβ-like NTR by the structure definition (Bullock *et al.*, 1996). It is a 15 kDa protein that is found as homodimer in the cell. The dimer can bind to two RanGDPs. NTF2 cargo release is linked to the conversion of RanGDP to RanGTP as it enters to the nucleus. This transport cycles are summarized in Figure 6-4.

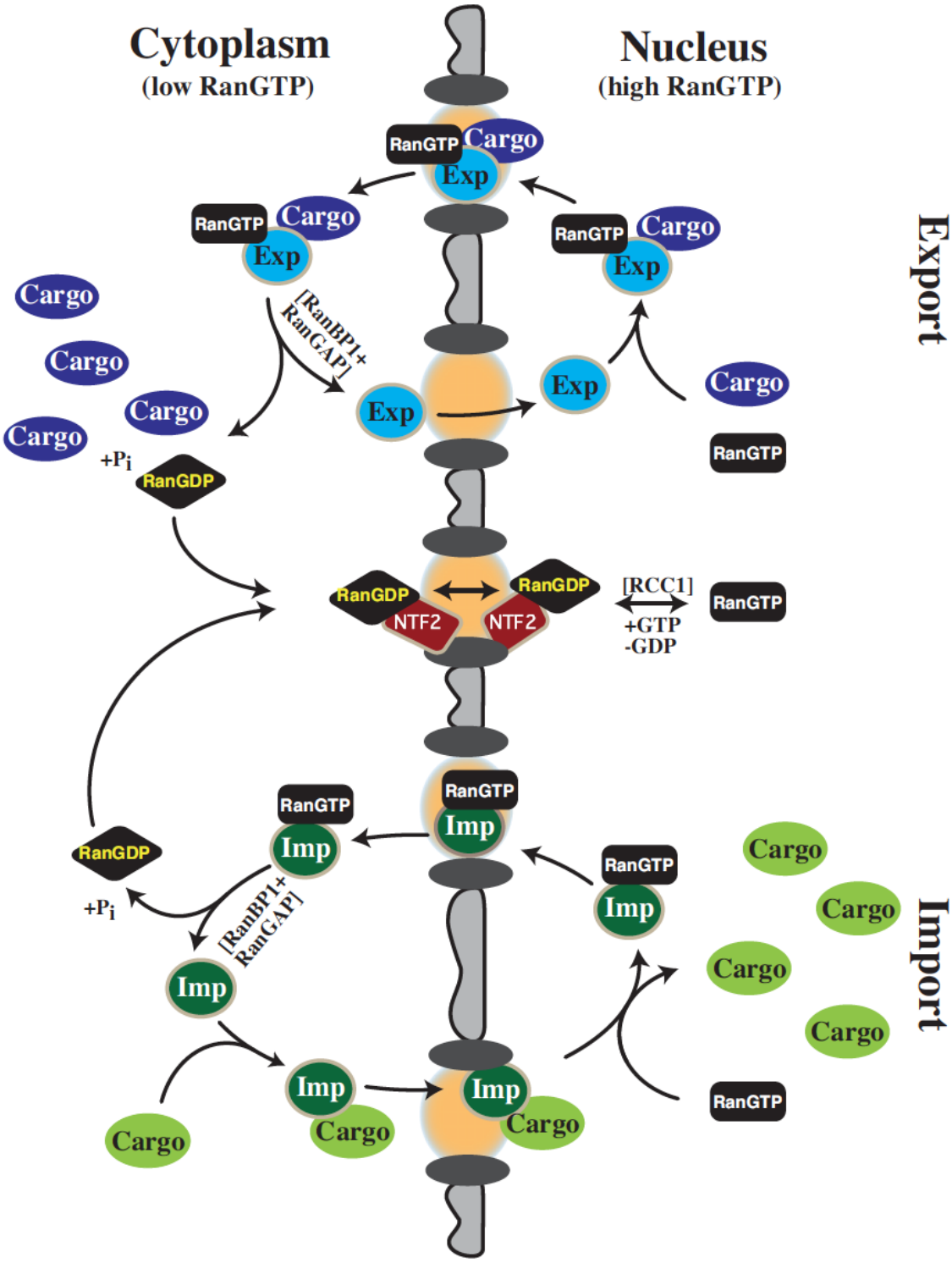

Figure 6-4 Overview of active nucleocytoplasmic transport through NPCs modified from (Gorlich and Kutay, 1999)

This elegant RanGTP gradient has other players on the backstage. Although Ran is a GTPase, it has a very low intrinsic activity. For an efficient hydrolysis, RanGTP needs stimulation of its GTPase activating protein RanGAP. RanGAP can increase the GTPase activity of Ran by 10^5 fold (Bischoff *et*

al., 1994). Human RanGAP, RanGAP1, has a modular organization; an N-terminal leucine rich repeat (LRR) region and a C-terminal domain that gets sumoylated.

The N-terminal LRR domain has the GTPase activating activity, and the C-terminal domain gets sumoylated by Ubc9 and triggers RanGAP interaction with cytoplasmic side of NPCs via Nup358 (Mahajan *et al.*, 1997) (Gareau *et al.*, 2012). While human RanGAP1 is localized to the NPCs, yeast and *S.pombe* orthologs Rna1p lack the C-terminal domain and it is localized to the cytoplasm. In all homologs, LRR domain is followed by a poly glutamic acid region (Figure 6-5) (Hopper *et al.*, 1990) (Melchior *et al.*, 1993b).

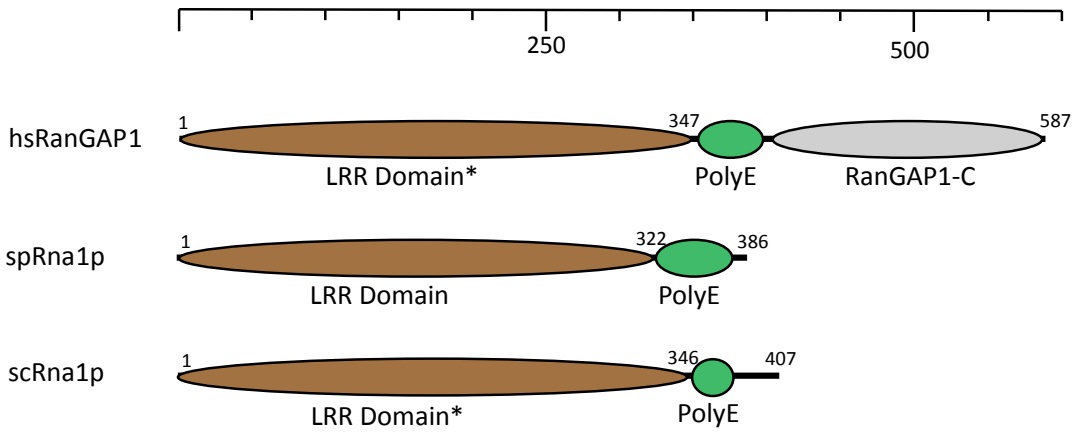

Figure 6-5 Domain organization of RanGAP homologs

Domains assigned by homology are indicated with '*'. LRR stands for leucine rich repeat domain and responsible for GTPase activation, PolyE represents the poly glutamic acid region. Domains are drawn to the scale (50 amino acids). End of LRR domain and protein are also indicated on the domain representations.

RanGAP can act on RanGTP, but RanGTP in export complexes is not accessible for RanGAP since the binding surface on Ran is overlapping for RanGAP and NTRs (Paraskeva *et al.*, 1999) (Seewald *et al.*, 2002) (Monecke *et al.*, 2009). It requires the assistance of RanBP1 or RanBP2/Nup358 which bind to the C-terminal loop on RanGTP in the export complex and destabilize it. (Yokoyama *et al.*, 1995; Bischoff and Gorlich, 1997; Koyama and Matsuura, 2010). When RanGAP and RanBP1 bind RanGTP, its GTPase activity is stimulated ~10^6 fold; RanBP1 contributes about 10 fold to the activation by RanGAP (Bischoff *et al.*, 1995).

Conversion of RanGDP to RanGTP is stimulated by the nuclear protein Ran guanine nucleotide exchange factor (RanGEF), also called RCC1 (regulator of chromosome condensation 1). RCC1 acts specifically only on Ran and stimulates the exchange of nucleotide (Bischoff and Ponstingl, 1991). RCC1 interaction does not have any preference towards GTP or GDP bound Ran, but the high

molar ratio of GTP to GDP in the cell drives RanGDP conversion to RanGTP (Bischoff and Ponstingl, 1991).

RanGAP, RanBP1 and RanBP2 take role in stimulation of GTPase activity of Ran. These proteins are kept cytoplasmic, or on the cytoplasmic side of the NPC (Hopper *et al.*, 1990; Yokoyama *et al.*, 1995; Richards *et al.*, 1996; Matunis *et al.*, 1996; Mahajan *et al.*, 1997; Saitoh *et al.*, 1997). On the other hand, RCC1 is chromatin bound, and constraints RanGTP generation to nucleus (Ohtsubo *et al.*, 1989).

Ran can act as a switch, because it undergoes drastic conformational changes in more than one position upon GTP hydrolysis. The core is mostly stable, but 3 regions show rearrangement upon GTP hydrolysis; switch-I (residues 30 to 47), switch-II (residues 65 to 80), and C terminal switch-III (residues 177 to 216)(Figure 6-6). In RanGTP structure, a Mg^{+2} ion and hydrogen bonds coordinate β and γ-phosphates of the GTP. Conformational change is triggered by the hydrolysis of the phosphodiester bond, and thus rearrangements in the network of hydrogen bonds. Switch-I is relocated completely and gains a α-helical structure. Switch-II undergoes a smaller conformational change than switch-I, but this change is significant since it is in close proximity of the nucleotide. The C terminal switch-III is the part that shows the most extreme change in the structure. C terminal switch-III is a long linker followed by a α-helical extension and the acidic stretch “DEDDDL”. In RanGDP structure the C terminal switch-III is folded back on the globular Ran core and it is in contact with switch-I. Although the acidic stretch is missing from the crystal structure, it most probably contacts the basic patch. In RanGTP structure, the changes in switch-I are transmitted to the C terminal switch-III, and contribute to its displacement from globular Ran core (Milburn *et al.*, 1990; Scheffzek *et al.*, 1995; Vetter *et al.*, 1999b).

NTR· RanGTP structure with Impβ, Transportin, and CAS also show that RanGTP is in contact with N termini of these NTRs (Chook and Blobel, 1999; Vetter *et al.*, 1999a; Matsuura and Stewart, 2004). RanGTP interacts through switch II, basic patch, and some other loops, and most of these regions would not be accessible to NTRs in RanGDP conformation due to C terminal switch-III.

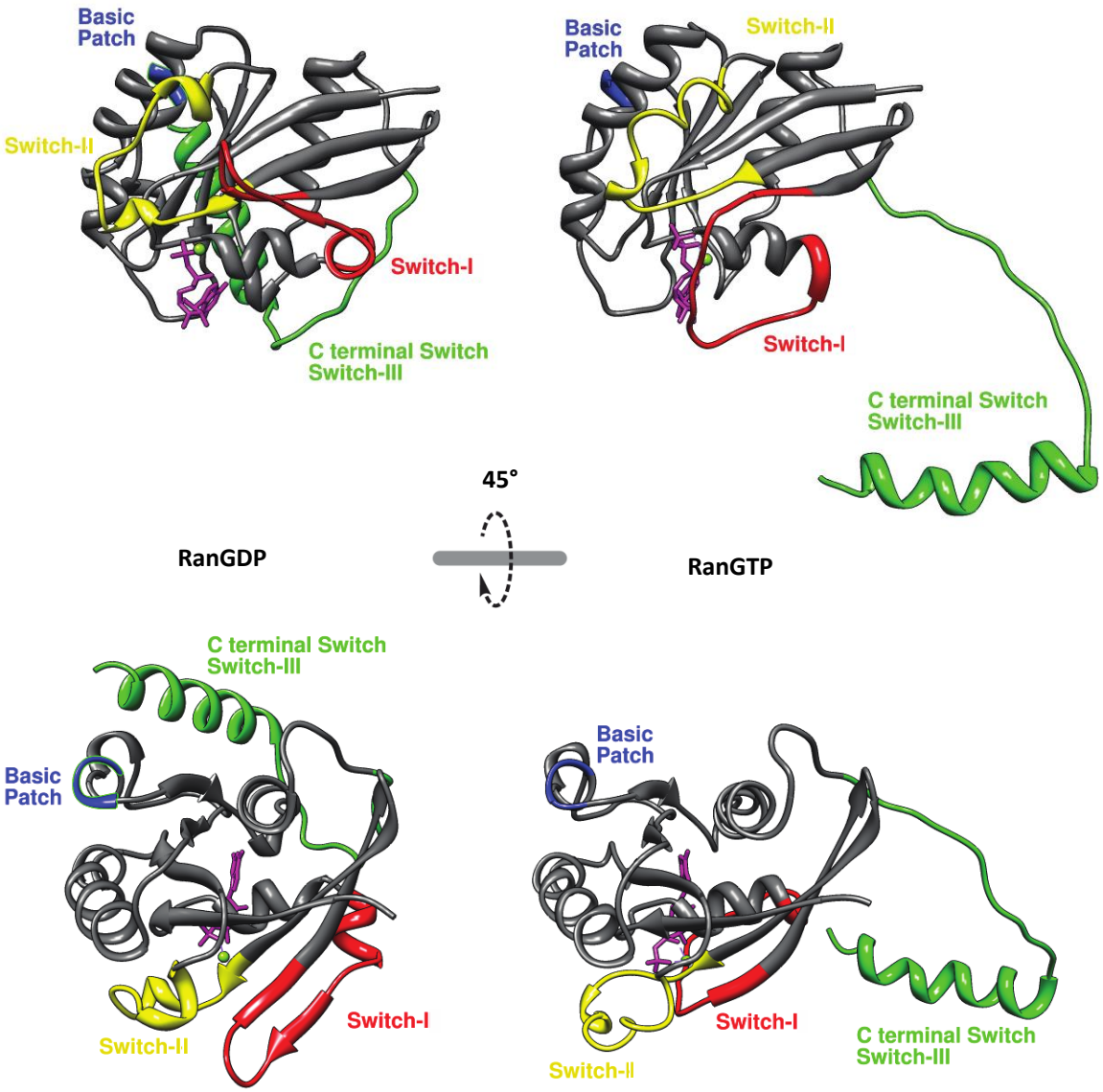

Figure 6-6 Comparison of RanGTP and RanGDP structures

RanGDP structure is from crystal structure with PDB-ID 3GJ0 (Partridge and Schwartz, 2009). RanGTP structure is part of Ran-GPPNHP-RanBD1 crystal structure with PDB-ID 1RRP (Vetter *et al.*, 1999b). From both structures, the overlapping part between amino acids 8 and 207 was visualized with ribbon representation. GDP and GTP were shown as purple sticks. For clarity, RanBP2/RanDB1 domain omitted from RanGTP structure. In both structures some parts of Ran sequence was missing. The overlapping part of two structures; residues 8-207 were used in this representation. Parts of Ran structure that undergo significant changes were indicated on the structure: amino acids 30 to 47 was marked as switch-I (red); 65 to 80 as switch-II (yellow); and 177-207 as C terminal switch-III (green). In the protein sequence, C terminal switch is continued with an acidic stretch, and was missing in the crystal structures. In RanGDP conformation this acidic stretch packs against a basic patch (blue).

6.4 NUCLEAR TRANSPORT RECEPTORS

Impβ-like NTRs are structurally very similar, they are composed of so-called HEAT repeats (Gorlich *et al.*, 1997), named after the proteins huntingtin, elongation factor 3, protein phosphatase 2A, lipid kinase TOR that were the first examples of this structural element (Andrade and Bork, 1995). HEAT repeats are composed of two antiparallel α-helices of 10 to 20 amino acids and linked by a short loop. NTR structure is made up of 18-20 HEAT repeats that line up sequentially with an angular shift that gives rise to a right-handed solenoid (Cingolani *et al.*, 1999; Chook and Blobel, 1999; Matsuura and Stewart, 2004; Monecke *et al.*, 2009). Hydrophobic side chains sustain interactions in and between HEAT repeats. Packing is uniform in a way that the first helix of the HEAT repeat faces outside of NTR circle, and the second one faces inside. This organization confers flexibility to the NTRs and helps them to adapt different conformations (cargo bound and non-bound) (Stewart, 2003). Linear arrangement of HEAT repeats also results in a large protein surface that is needed for interaction with RanGTP, respective cargoes and also FG repeats of the NPC. NTRs recognize many different classes of cargos either to import (Table 6-1), or to export (Figure 6-4).

Impβ like NTRs share many features. They are made up of the same structural elements, they have acidic isoelectric points (pI 4.0-6.0), yet they have very low overall sequence homology (8 - 15 %). The only significant homology is found in the N-terminal region that accounts for interaction with RanGTP (Gorlich *et al.*, 1997).

NTR	Selected Cargoes	References
Importin β (Impβ-1)	Ribosomal Proteins	Gorlich *et al.*, 1995
	HIV Rev, HIV Tat	Huber *et al.*, 1998a
	Histones	Jakel and Gorlich, 1998
	Snurportin1.UsnRNPs	Jakel *et al.*, 1999;
with Importin 7	histone H1	Truant and Cullen, 1999
with Importin α	Classical NLS-cargoes	Muhlhausser *et al.*, 2001
Transportin 1+2	hnRNP protiens	Pollard *et al.*, 1996
(Trn, Impβ-2)	Ribosomal proteins	Jakel and Gorlich, 1998
	TAP/NFX1	Truant *et al.*, 1999
	Histones	Muhlhausser *et al.*, 2001
	c-Fos	Arnold *et al.*, 2006
Transportin SR 1+2	SR proteins	Kataoka *et al.*, 1999
(TrnSR, Trn 3)		
Importin 4	Ribosomal proteins	Mosammaparast *et al.*, 2001
	Histones	Jakel *et al.*, 2002

Importin 5	Ribosomal Proteins	Jakel and Gorlich, 1998
	Histones	Mosammaparast *et al.*, 2001
Importin 7	Ribosomal Proteins	Jakel and Gorlich, 1998
	Histones	Muhlhausser *et al.*, 2001
	ERK2, SMAD3, MEK1	Chuderland *et al.*, 2008
Importin 8	SRP19	Dean *et al.*, 2001
	Argonaute proteins	Weinmann *et al.*, 2009
Importin 9	Ribosomal Proteins	Muhlhausser *et al.*, 2001
	Histones	Jakel *et al.*, 2002
Importin 11	UbcM2	Plafker and Macara, 2000b
	rpL12	Plafker and Macara, 2002
Importin 13	hUBC9, MGN/Y14	Mingot *et al.*, 2001
	TF NF-Y	Kahle *et al.*, 2005
	CHRAC-15/17	Walker *et al.*, 2009
	NC2 Complex	Kahle *et al.*, 2009
Exportin 4	Sox2, SRY	Gontan *et al.*, 2009

Table 6-1 Mammalian importins and selected cargos

NTR	Selected Cargoes	References
CRM1 (Exportin 1)	Leucine rich export signals	Fischer *et al.*, 1995
	HIV Rev.RRE containing RNAs	Fornerod *et al.*, 1997
	Snurportin1	Paraskeva *et al.*, 1999
	Signal recognition particle	Trotta *et al.*, 2003
	Nmd3.60S Ribosomal Subunit	Alavian *et al.*, 2004
	PHAX.UsnRNAs	Ohno *et al.*, 2000
CAS (Exportin 2)	Importin αs	Kutay *et al.*, 1997
Exp-t (Exportin 3)	tRNA	Kutay *et al.*, 1998
Exportin 4	eIF5A	Lipowsky *et al.*, 2000
	SMAD3	Kurisaki *et al.*, 2006
Exportin 5	aa-tRNA.eEF1A	Bohnsack *et al.*, 2002
	dsRNA.dsRNA binding proteins	Brownawell and Macara, 2002
	pre-miRNAs	Bohnsack *et al.*, 2004
Exportin 6	Actin.profilin	Stuven *et al.*, 2003
Exportin 7	p50RhoGAP, 14-3-3σ	Mingot *et al.*, 2004
Importin 13	eIF1A	Mingot *et al.*, 2001

Table 6-2 Mammalian exportins and selected cargoes

6.4.1 CRM1/Exportin 1

CRM1 (chromosomal region maintenance 1) was first found in *Schizosaccharomyces pombe* genetic screen with a cold sensitive mutation that resulted in deformed chromosomes, and was not recognized as a nucleocytoplasmic transport related protein (Adachi and Yanagida, 1989). Later it was found as the target of toxin Leptomycin B (Nishi *et al.*, 1994).

The leucine rich nuclear export signals (NESs) were discovered in HIV-1 Rev protein and protein kinase A inhibitor (PKI), although it was known that a mediator was involved in the nuclear export of these proteins, the identity of the respective NTR was not clear (Fischer *et al.*, 1995; Izaurralde and Mattaj, 1995; Wen *et al.*, 1995; Gorlich and Mattaj, 1996). Later CRM1 was identified as the nuclear transport receptor of these proteins with NESs (Fornerod *et al.*, 1997; Fukuda *et al.*, 1997; Neville *et al.*, 1997; Ossareh-Nazari *et al.*, 1997). Since then, the library of proteins that are exported by CRM1 via an NES grew tremendously. A curated database of CRM1 cargoes with validations at different experimental settings has more than 250 entries from various species.

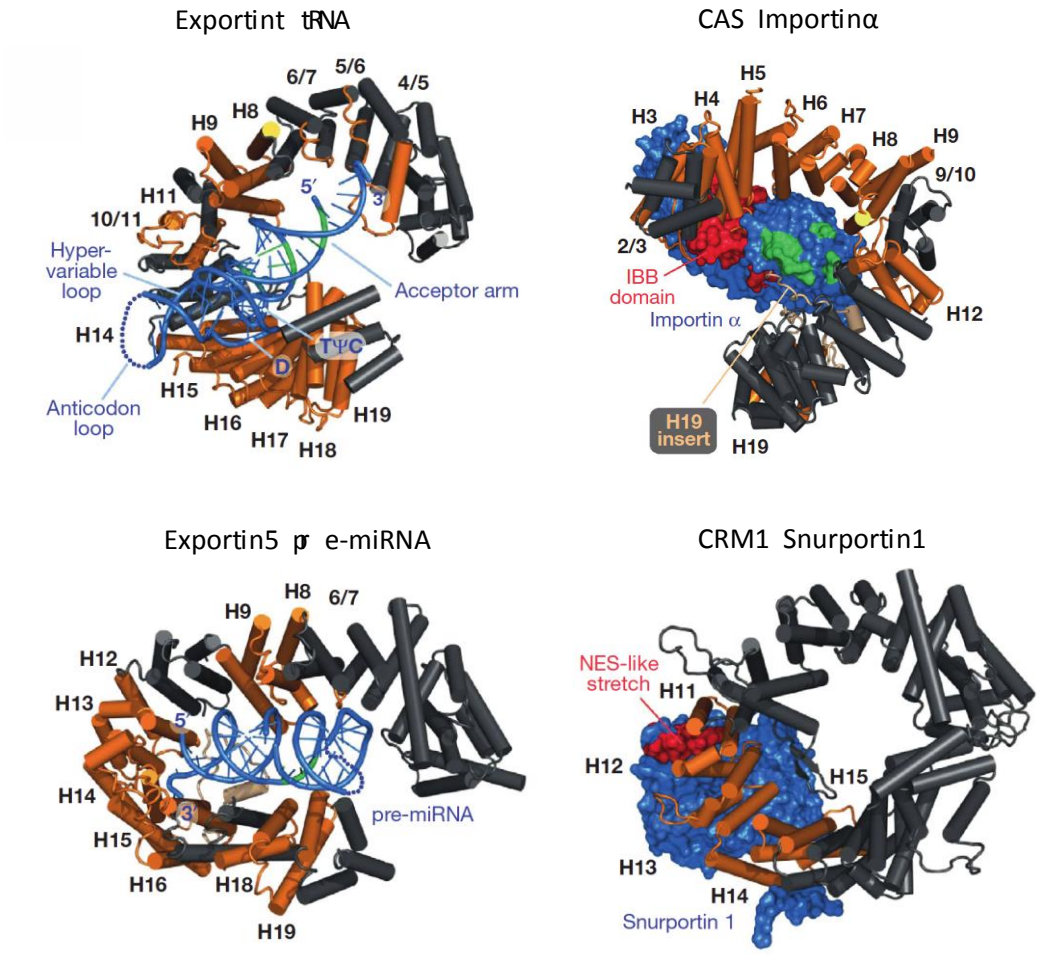

Figure 6-7 Exportin structures with respective cargoes adapted from (Güttler and Görlich, 2011).

Crystal structures of 4 exportins cargo RanGTP complexes are shown without RanGTP. Cargoes are in blue and their contacting helices in NTRs are colored orange. Residues interacting with RanGTP on cargoes are marked green.

The large number of cargoes nominates CRM1 as the most promiscuous NTR of the cell. It recognizes various cargos that are structurally and functionally distinct. This feature of CRM1 can be attributed to its cargo recognition mechanism that is different from other exportins. Several exportins have been crystalized in complex with RanGTP and the respective cargoes. Comparison of these structures point out that exportins other than CRM1 wrap their cargos with the inner surface of the solenoid NTR structure, while cargo binding of CRM1 is on its outer surface with a limited interaction area (Cargo interaction surfaces of NTRs are colored orange in Figure 6-7)(Güttler and Görlich, 2011). One should note that the interaction surface of CRM1 with Snurportin 1 is far larger than the interaction surface with the NES only.

Snurportin 1· CRM1 structure was the first crystalized CRM1 cargo complex, due to its high stability. This is sustained by interaction surfaces in addition to the N-terminal NES (Monecke *et al.*, 2009). CRM1 interacts with Snurportin 1 tighter than with its other export substrates, because CRM1 is not only the export factor of Snurportin 1, but also acts as the disassembly factor for imported Snurportin 1· U snRNP complex (Huber *et al.*, 1998b). The exported Snurportin 1 would be ready for another cycle of U snRNP import. CRM1 interactions with other cargoes are less stable and might have even smaller interaction surface with CRM1. The interaction surface of CRM1 with NESs is a hydrophobic cleft build by 4 neighboring α-helices.

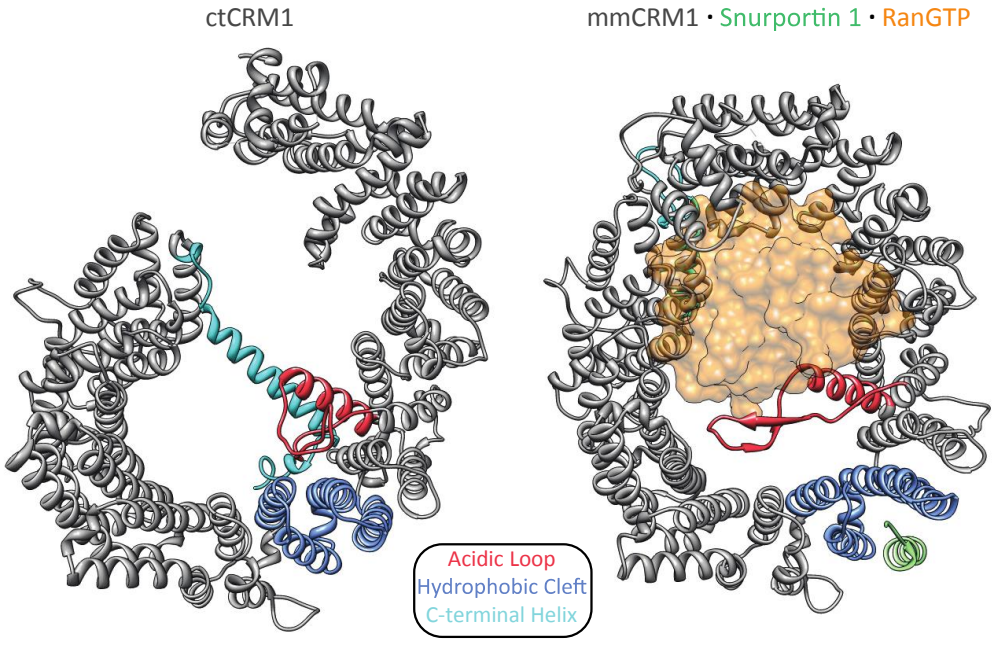

Figure 6-8 Free and cargo bound states of CRM1

Structural comparison of free CRM1 structure from *Chaetomium thermophilum* (PDB ID 4FGV), and human Snurportin 1 and RanGTP bound mmCRM1 structure (PDB ID 3GJX). RanGTP was represented as orange transparent surface, and Snurportin 1 NES was represented as green ribbon. For clarity, rest of the Snurportin 1 structure was omitted. 3 regions that show great flexibility and important for the stabilization of different states were colored. Acidic loop (ctCRM1$^{421\text{-}460}$, mmCRM1$^{423\text{-}464}$) is red, hydrophobic cleft (ctCRM1$^{507\text{-}591}$, mmCRM1$^{510\text{-}594}$) is blue, and C-terminal helix (ctCRM1$^{1033\text{-}1077}$, mmCRM1$^{1021\text{-}1071}$) is cyan.

Comparison of exportin structures also reveals another aspect of CRM1 export mechanism. In the cases of exportins CAS, Exportin-t and Exportin 5, cargoes interact not only with their respective NTRs but also with RanGTP. On the other hand, CRM1 serves as a platform that RanGTP and Snurportin 1 bind on separate surfaces. RanGTP and Snurportin 1 are not in direct contact. Although these interactions are spatially separated, they favor the same structural conformation of CRM1, and bind to CRM1 in a cooperative manner (Monecke *et al.*, 2009).

CRM1 has two different conformations, the relaxed conformation that is free of a cargo and RanGTP, and the strained conformation that is stabilized by RanGTP and NES binding (Dong *et al.*, 2009; Monecke *et al.*, 2009; Dian *et al.*, 2013; Monecke *et al.*, 2013). The main functional difference between the two conformations is the opening of the hydrophobic cleft on CRM1. The distance between the helices 11A and 12A are considerably different, and in the stained conformation they are separated enough to accommodate an NES in between. This strained conformation is stabilized by RanGTP binding with two mechanisms. First, the C terminal helix that stabilizes the relaxed conformation is displaced, and N and C-terminus of CRM1 are brought together. Second, the acidic loop is reorganized to form a β hairpin. In RanGTP bound conformation, the acidic loop extends towards the center of CRM1 and reaches to the other side of the toroid structure and touches helix 15B (Figure 6-8).

High RanGTP concentration in the nucleus drives CRM1 to strained conformation that is ready to accept the incoming NESs, and binding of NES further stabilizes this conformation. Upon arrival to the cytoplasm, RanBP1 or RanBP2 binding to RanGTP disassembles the export complex and RanGAP converts RanGTP to RanGDP. With low levels of RanGTP in the cytoplasm, cargo is not able to bind to CRM1. CRM1 goes back to the nucleus and performs another cycle of export.

6.5 NUCLEAR EXPORT SIGNALS

Nucleocytoplasmic transport is made possible by a reversible binding of cargo to its respective NTR under regulation of RanGTP. The toroid shape of importin β like transport receptors have large protein surface to fulfill this function. NTRs interact with FG repeat proteins of the NPC with their outer surface. RanGTP interacts with the N-terminal B helices, and sits in the inner gap of the toroid. For most NTRs, the inner surface of this toroid is also the binding platform for the transported cargoes. CRM1 is an exception where NES binding hydrophobic cleft is on the outer surface. The large inner surface of NTRs confers many possibilities for interaction with respective cargoes. A nuclear export signal (NES) is a short amino acid stretch that directs proteins to the cytoplasm utilizing the essential NTR CRM1. Investigation of NESs revealed many aspects of NES-mediated transport.

Not all NES containing cargoes are constitutively exported from the nucleus, meaning that CRM1-mediated export can be a regulated transport. Many ways of NES-dependent export regulation have been suggested. Regulated accessibility of NESs (Li *et al.*, 1998; Stommel *et al.*, 1999a; Seimiya *et al.*, 2000; Heerklotz *et al.*, 2001; Kobayashi *et al.*, 2001; Craig *et al.*, 2002), phosphorylation (Engel *et al.*, 1998; Ohno *et al.*, 2000; McKinsey *et al.*, 2001; Zhang and Xiong, 2001; Brunet *et al.*, 2002) and also by oxidation, e.g., disulfide bond formation (Yan *et al.*, 1998; Kudo *et al.*, 1999b; Kuge *et al.*, 2001).

The concept of an NES was first suggested relying on the observations that some proteins continuously shuttle between the cytoplasm and the nucleus (Wen *et al.*, 1994; Fischer *et al.*, 1995; Gerace, 1995). These proteins included hnRNP A1 (Pinol-Roma and Dreyfuss, 1992), HIV-1 Rev protein (Kalland *et al.*, 1994; Meyer and Malim, 1994), cAMP-dependent protein kinase (PKA) (Harootunian *et al.*, 1993; Fantozzi *et al.*, 1994), some transcription factors (Madan and DeFranco, 1993) and hsc70 (Mandell and Feldherr, 1990), and they had the potential to bear an NES besides a nuclear localization signal (NLS). A thorough analysis of two of these proteins, Rev and PKA revealed the first NESs.

Inactive PKA holoenzyme consists of two regulatory and two catalytic subunits, and is localized to the cytoplasm. Binding of cAMP to the regulatory subunits triggers the dissociation of the monomeric catalytic subunits, which then can diffuse into the nucleus. Activity of catalytic subunit is strictly regulated and is inactivated by binding of 74 amino acids long protein kinase inhibitor (PKI). Binding of PKI not only inhibits the enzymatic function but also leads to nuclear exclusion of the catalytic subunit (Fantozzi *et al.*, 1994). The sequence that was responsible for nuclear exclusion was a 10 amino acids stretch on PKI. Fusion of fluorescently labeled proteins to this

fragment restricted their localization to the cytoplasm, and identified it as the first NES (Wen *et al.*, 1994).

Rev is an essential protein for virus reproduction (Cullen, 1992). It takes role in export of viral RNA from the nucleus to the cytoplasm. Two important sequence elements were discovered on Rev, an RNA stem loop interaction motif called Rev response element (RRE), and a C-terminal leucine rich activation domain. This 10 amino acid long leucine rich activation domain was enough to direct other conjugation partners like BSA to the cytoplasm and identified as the second NES (Fischer *et al.*, 1995).

Later CRM1 was identified as the NTR responsible for the transport of NES harboring cargoes (Fornerod *et al.*, 1997; Fukuda *et al.*, 1997; Neville *et al.*, 1997; Ossareh-Nazari *et al.*, 1997). After the identification of the first NES examples, the library of NES containing CRM1 cargoes grew rapidly. Different groups compiled curated NES libraries, or constructed mutant NES libraries, and by analyzing them, they tried to come up with consensus definitions to predict NES sequences.

The first attempt was done by randomization of Rex activation domain. Rex is the functional equivalent of Rev in T-cell leukemia virus type 1, and it also has an NES termed activation domain. By randomization of the activation domain a library was constructed. Then this library was tested for functionality of the activation domain. By aligning the functional sequences, the prominent residues and their spacing was combined in to the consensus L-$X_{2,3}$-[FILVM]- $X_{2,3}$-L-X-[LI], and this definition led to the term leucine rich nuclear export signal (lrNES) (Gerace, 1995; Bogerd *et al.*, 1996).

la Cour *et al*. compiled the first curated database of NES containing proteins in NESbase 1.0. This database contains 80 NES sequences on 75 proteins (la Cour *et al*, 2003). Only 25 of these NES were defined by the previous lrNES consensus. This database was later used for construction of the first NES prediction algorithm NetNES. Two training sets were generated from the validated NES sequences and NES containing protein sequences excluding NESs. These two sets were used to train a machine-learning algorithm. The allowed hydrophobic residues were increased to L, I, M, V and F at 4 positions, and E, D and S residues were preferred as spacers. This new consensus, [FILVM]-$X_{2,3}$-[FILVM]- $X_{2,3}$-[FILVM]-X-[FILVM] was able to cover 50 of the 75 NESs in the database (la Cour *et al.*, 2004).

A third study was based on a screen of random peptides for their exclusion from the nucleus. This study found 101 different peptides that were export competent, and grouped them into three different classes. Hydrophobic positions were termed as Φ positions. L, I, M, V and F were allowed

at Φ positions, and C, W, A and T were also allowed only at one Φ position. Proline residues in the spacer residues were enough to prevent the export, so proline was excluded from the spacer residues. The class I consensus Φ-$X_{2,3}$-Φ-$X_{2,3}$-Φ-X_2-Φ was the same as the previous consensus, and covered 83 of the 101 functional NESs. Class II consensus Φ-X-Φ-X_2-Φ-X-Φ and Class III consensus Φ-$X_{2,3}$-Φ-$X_{2,3}$-Φ-X_2-Φ were novel and rare, and together they covered 17 of 101 functional NESs. Although these three classes can explain 99 out of 101 artificial NESs, they can only cover 89 of 159 naturally occurring NESs (Kosugi *et al.*, 2008). These definitions were not available as an NES prediction tool.

A second computational approach after NetNES came from Fu *et al.*. They also constructed two data sets of true and false NESs of 60 proteins selected from NESbase 1.0. Consensus was defined with three Φ positions, Φ-$X_{2,3}$-Φ-X-Φ, and Φ positions were limited to L, I, V, M and F. The sequences were analyzed for various parameters these parameters were evaluated by LIBSVM (Chang and Lin, 2001) to find the features that gave the significant differences between true and false NESs. These included negative charges in the inter Φ positions and disorder tendencies, and used in the prediction algorithm NESsential (Fu *et al.*, 2011).

The latest curated library of CRM1 cargoes was compiled by Xu *et al.* in NESdb. This database contains 221 NES containing cargoes from various species (Xu *et al.*, 2012a). Analysis of these NESs were summarized in 3 consensus sequences, Φ-$X_{1,2,3}$-Φ-$[\text{^W}]_2$-Φ-[^W]-Φ (type 1), Φ-$X_{2,3}$-Φ-$[\text{^W}]_3$-Φ-[^W]-Φ (type 2), and Φ-X_2-Φ-X$[\text{^W}]_2$-Φ-$[\text{^W}]_2$-Φ (type 3), where [^W] is any of the 20 amino acids except Trp. Φ positions are either L, I, V, F or M, and A and T residues are allowed only once at either first or second Φ position (Xu *et al.*, 2012b).

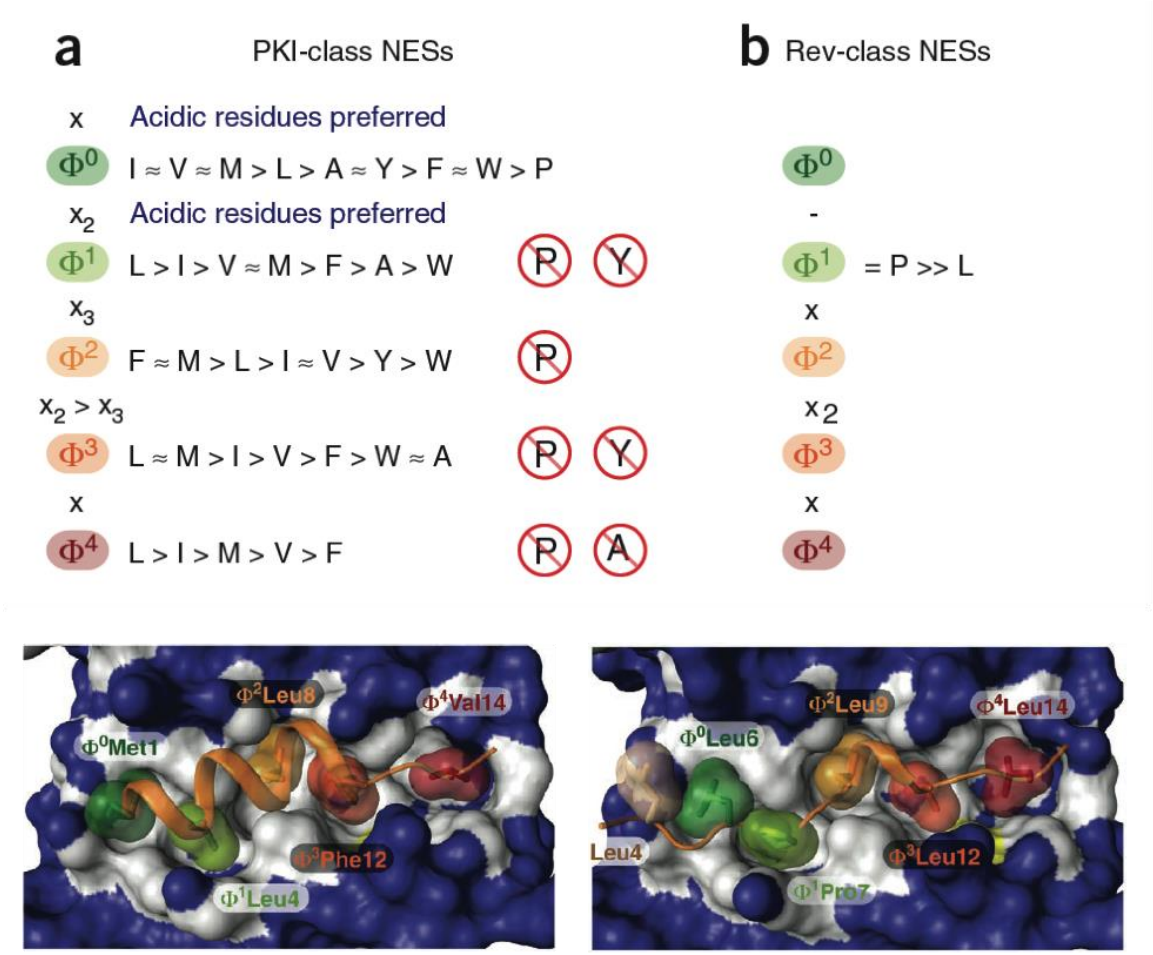

Figure 6-9 Structural definition of NES consensus - modified from (Güttler *et al.*, 2010)

Crystal structures of CRM1 with Snurportin 1 elucidated the true nature of the N-terminal Snurportin 1 NES interaction with CRM1 with 5 Φ positions (Güttler *et al.*, 2010). Güttler et al. replaced the NES with PKI and Rev NESs and obtained two additional crystal structures. PKI NES and Snurportin 1 NES fit into CRM1 hydrophobic cleft with very similar structural orientation, whereas Rev NES is placed in a very different way. This is why these two different NES were separated into PKI type or Rev type NES consensus. Mutation screen of PKI NES Φ positions also revealed preference of Φ positions for different amino acids (Figure 6-9). These new definitions of NESs based on the crystal structures were the basis of the new NES prediction tool that I discuss further in the results and discussion.

7 RESULTS

The recently determined crystal structures of CRM1 with bound NESs uncovered some essential details as to how this nuclear export receptor can bind to its various cargoes from a broad range of structural and functional groups. It not only provided an understanding for the previously recognized consensus amino acid sequence for CRM1 dependent NESs but also was the basis for a thorough mutational analysis that more clearly defined the amino acid requirements at five Φ positions. These experiment in combination with the available structures provided a clearer picture of the properties that render a linear amino acid sequence into a faithful CRM1 binder. We wanted to make use of the gained information to develop a prediction tool that would identify and score potential NESs within a give sequence.

7.1 A NEW NES PREDICTION ALGORITHM

The widely accepted consensus amino acid sequence for CRM1-dependent nuclear export signals (NESs) Φ-$x_{(2-3)}$-Φ-$x_{(2-3)}$-Φ-x-Φ (Φ for hydrophobic residues, x for any amino acid) is better understood in the context of the later solved CRM1-RanGTP-Cargo crystal structures (Dong *et al.*, 2009; Monecke *et al.*, 2009; Güttler *et al.*, 2010).

We wanted to apply the new experimental findings to generate an improved NES prediction tool. To achieve this goal, we focused on the NES prototypes with the same Φ residue spacing as in these crystal structures and considered a previously published systematic mutational analysis for each of these Φ positions (Güttler *et al.*, 2010). The outcome of the latter study resulted in a scoring matrix to estimate CRM1 binding strength of a given sequence.

The consensus NES definition contains several critical hydrophobic residues. Since hydrophobic residues are often buried in the folded core of the protein structure, there is a high probability of finding NES hits that are not accessible for CRM1 interaction. To sort out such potential false positives, we applied two types of filtering. First we predicted the disorder propensity for the found hit, and the 6 amino acids before and after that region. This prediction is used for assessing the possibility of the hit being exposed for an interaction. Second, protein sequence is searched for domain homology since folded domains are less likely to contain a disordered stretch of amino acids that can act as an NES. At the end, high scoring NESs that are not in a folded domain and that have high disorder propensity were considered as good hits.

7.1.1 NES Consensus

Crystal structures of CRM1 with NESs show that there are at least two different arrangements of 5 Φ residues that can fit into the hydrophobic cleft (Güttler *et al.*, 2010). The first one is the more common NES pattern that is in agreement with the PKI NES Φ residue arrangement. The second one follows the REV NES Φ residue arrangement, and described as a new class of NES consensus for the first time. I will refer to these two types as PKI-type and REV-type NES.

7.1.1.1 PKI-type NES consensus

To scan the given protein sequences for NES hits, a pattern-matching algorithm called regular expression is used. To construct the PKI-type regular expression, following statements are used.

- PKI type Φ residues follow a Φ_1-$x_{(3)}$-Φ_2-$x_{(2-3)}$-Φ_3-x-Φ_4 spacing (la Cour *et al.*, 2004) (Güttler *et al.*, 2010)
- Proline residues are not allowed in the spacer regions between Φ_1 and Φ_4 (Kosugi *et al.*, 2008).
- Φ_0 and neighboring negatively charged amino acids contribute positively to the binding (Güttler *et al.*, 2010).

These statements were combined into the PKI-type NES regular expression (Figure 7-1).

Φ0	Φ1			Φ2				Φ3		Φ4	
. . . .	[LIVMFWAY]	[^P]	[^P]	[^P]	[FMLIVYW]	[^P]{2,3}		[LMIVFWAY]	[^P]	[LIMVFPWY]	
N L E A	L	Q	K	K	L	E	E	L	E	L	

Figure 7-1 Regular expression for PKI-type NES pattern and an example NES

Each bracket-enclosed expression represents a position with allowed amino acids. Dot (.) represents any residue. Residue after '^' sign is not allowed at that position. Curly brackets indicate the allowed repeat numbers for the previous pattern (e.g. [^P](2,3) means 2 to 3 amino acids stretch without any proline). Underneath the regular expression the NES from Map kinase kinase 1 is placed with matching positions.

The first 4 amino acids including the Φ_0 position did not have any prerequisites during the pattern search, since any amino acid (represented by '.' in a regular expression) can be matched. Contribution of these residues was graded later in the NES Score. Allowed amino acids in the Φ positions are explained in the NES Score section.

7.1.1.2 REV-type NES consensus

Early studies tried to "squeeze" the REV NES into a PKI-type consensus. The actual CRM1·Rev-NES structure revealed however a different binding conformation between Φ0 and Φ2 and Φ_1 pocket was occupied by a proline and not by a more typical hydrophobic amino acid (Güttler *et al.*, 2010).

A regular expression by this new structural definition was constructed for REV-type NES consensus (Figure 7-2).

Figure 7-2 Regular expression for REV-type NES pattern and an example NES
Each bracket-enclosed expression represents a position with allowed amino acids. Dot (.) represents any residue. Underneath the regular expression the NES from Rev protein is placed with matching positions.

NES score was calculated only for PKI-type NES hits, and for REV-type NES Φ positions, only a limited set of favored hydrophobic amino acids were allowed.

7.1.2 NES Score

An NES scoring scheme for PKI-type NES hits was designed based on previously published CRM1 binding assay with point mutants of PKI NES (Güttler *et al.*, 2010). For Φ residues 1 to 4, each position was given an incremental score (S1 to S4) of 1 to 10, based on the ranking of amino acid preference. These amino acids include tyrosine, tryptophan, phenylalanine, and alanine in addition to the previously recognized hydrophobic residues leucine, valine, methionine, and isoleucine. Alanine is only accepted as Φ_1 or Φ_3 residue (Figure 7-3).

Previous NES consensus definitions used 4 Φ positions (Φ_1 to Φ_4), which were important for CRM1 interaction. With the previously published crystal structures of NES bound CRM1, a previously unrecognized NES position was defined and named Φ_0 (Güttler *et al.*, 2010). Although it was defined recently, an analysis showed that 76% of validated NES, which fit into previous NES consensus, has a hydrophobic residue in Φ_0 position (Xu *et al.*, 2012b). This position clearly contributes to binding, and allows the construction of high affinity CRM1 binders.

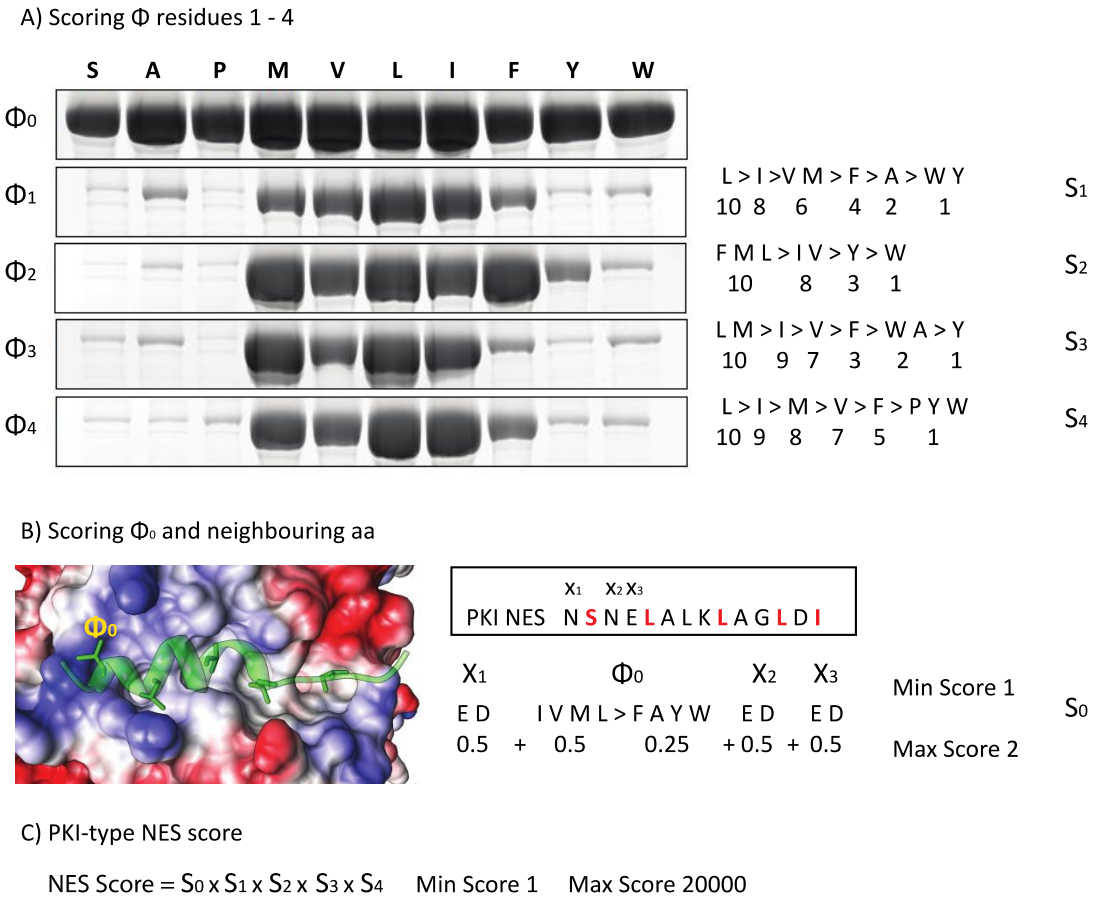

Figure 7-3 Scoring Scheme for PKI-type NESs

A) Correlation between Φ position occupation and CRM1-binding strength and conversion of this correlation into scores (Güttler *et al.*, 2010). B) PDB structure entry 3NBY. CRM1 surface colored according to coulomb potential (red is negative and blue is positive) and Φ_0L PKI sequence backbone is colored transparent green, and pocket fitting Φ residue side chains are colored solid green. Scoring scheme for Φ_0 and negative residues around it. C) Calculation of NES score based on Φ position specific scores.

Negatively charged residues around Φ_0 also contribute to this binding with electrostatic interactions, evident from the positive charges around Φ_0 binding pocket, and also from previous studies (Figure 7-3B) (Güttler *et al.*, 2010). Since contribution of this part was not as crucial as the other Φ residues, its effect to the score was limited with a coefficient of 2. If the S_0 score was less that 1, it was overridden by 1 to not to affect the final score negatively.

The final PKI-type NES score was calculated by multiplying all 5 sub-scores to represent the cooperative binding of Φ pockets. This can yield score of 1 as minimum and score of 20000 as maximum.

7.1.3 Disorder Propensities

For an NES to bind CRM1, we reasoned that not only the NES itself but also a small region following and preceding the actual NES should have disorder tendencies. Following this reasoning

we analyzed three regions for disorder propensity by IUPred; the six amino acids before the hit, the candidate NES, and the six amino acids after the hit. IUPred gives a disorder propensity value for each amino acid ranging from 0 (complete order) to 1 (complete disorder), and these three regions got one value each by averaging the disorder propensity over the analyzed region. If the NES hit is at the extreme N or C-terminus, the disorder propensity is set to 1 for the preceding or proceeding part.

For simplicity, disorder values were categorized into three sections. The first category was indicated by ”1” and covered averaged disordered propensities of 0 to 0.25. Second category was indicated by ”2” and covered averaged disordered propensities of 0.25 to 0.5. Third category was indicated by ”3” and covered averaged disordered propensities of 0.5 to 1.0. IUPred regards values lower than 0.5 as order and values higher than 0.5 as disorder. The lower half was separated into two categories because previous studies showed that a large portion of linear motifs also resided in the second range (Fuxreiter *et al.*, 2007).

Additional information was fetched from SMART domain database (Schultz *et al.*, 1998). Since not all annotated domains are folded domains and prediction of exact domain borders are not accurate, such domain prediction was used with caution. When multiple sequences were analyzed, domain prediction was exempted from constraints. It was used as a visual inspection tool of individual hits, since a final reasoning requires analysis of the predicted domains.

7.1.4 Evaluation of PKI-type NES prediction

We wanted to know if the algorithm would allow prediction of functional NESs within a given amino acid sequence. To this end we selected 11 proteins the NES of which have been experimentally characterized before by others. Their primary sequences were retrieved from databases and fed into the NES prediction algorithm. Sequence analysis revealed putative PKI-type NESs, which were subsequently ranked according to our scoring criteria. Interestingly, highest scoring NES hits largely matched the sequences that have been validated experimentally. This indicated that the algorithm was indeed capable of predicting functionally relevant NES.

The PKI-type NES prediction algorithm was written in Python (12.1.1), which is a programming language that is widely used by the bioinformatics community and for which many code libraries are already available. The Input file is a fasta formatted protein sequence or several sequences. The algorithm iterates over the given sequences and outputs the predicted NES borders and sequences with disorder propensity, domain prediction, an NES score for each hit sequence. An example output is shown in Figure 7-4 for human MAP kinase kinase 1 (MP2K1_Human).

MP2K1_HUMAN							
Disorder			Sequence	Start	End	Domain	NES Score
B	NES	A					
3	2	2	N**L**EA**L**QKK**L**EE**L**E**L**	29	42	n.i.d.	10000
1	1	1	SGLVMARKLIHLEI	90	103	in S_TKc	5400
1	1	1	GLVMARKLIHLEIKP	91	105	in S_TKc	144
1	1	1	IKPAIRNQIIRELQV	103	117	in S_TKc	4480
1	1	1	CNSPYIVGFYGAFY	121	134	in S_TKc	20
1	1	1	IPEQILGKVSIAVI	161	174	in S_TKc	1152
1	1	1	ILGKVSIAVIKGLTY	165	179	in S_TKc	480
2	1	1	THYSVQSDIWSMGL	238	251	in S_TKc	4800
2	1	2	QSDIWSMGLSLVEM	243	256	in S_TKc	560
2	1	2	DIWSMGLSLVEMAV	245	258	in S_TKc	4200
3	2	1	R**P**PM**A**IFE**L**LD**Y**I**V**	305	318	in S_TKc	140
1	1	1	ERADLKQLMVHAFI	348	361	in S_TKc	1800
1	1	2	EEVDFAGWLCSTIGL	367	381	n.i.d.	3600

Figure 7-4 An output example from PKI- type NES prediction

MP2K1 is the abbreviation for MAP kinase kinase 1 and was shown to have an N-Terminal NES (Fukuda *et al.*, 1996). From the three disorder values (B) represents the six amino acids before the NES, (NES) represents the predicted hit, and (A) represents the six amino acids after NES. S_TKc is the abbreviation for SMART domain Serine/Threonine protein kinases, catalytic domain. 'n.i.d' stands for 'not in any domain'. Hits with a NES disorder prediction of 1 are shaded gray. For other hits, putative Φ positions are marked bold.

Performance of the NES score and disorder filtering for PKI-type NES prediction was evaluated on a set of previously defined NES dependent CRM1 cargos. 11 proteins of NES instance examples from ELM database were used for evaluation (Table 7-1).

Protein	NES Sequence	Reference
Spn1_Human Snurportin 1	^{1}**M**EE**L**SQA**L**ASS**F**S**V**^{14}	(Monecke *et al.*, 2009)
Apc_Human Adenomatous polyposis coli protein	^{64}G**Q**IDLLER**L**KE**L**N**L**^{77}	(Henderson, 2000)
Ccnb1_Xenla G2/mitotic-specific cyclin-B1	^{104}L**P**DE**L**CQA**F**SD**V**L**I**^{117}	(Yang *et al.*, 1998)
Ctnd1_Human Catenin delta-1	^{940}G**Q**ES**L**EEE**L**DV**L**V**L**^{953}	(van Hengel *et al.*, 1999)
Ipka_Human (PKI) cAMP-dependent protein kinase inhibitor α	^{34}N**S**NE**L**ALK**L**AG**L**D**I**^{47}	(Johnson *et al.*, 1999)
Rang_Human Ran-binding protein 1	^{176}H**A**EK**V**AEK**L**EA**L**S**V**^{189}	(Richards *et al.*, 1996)

Mp2k1_Xenla MAP kinase kinase 1	[29]N**L**EA**L**QKK**L**EE**L**E**L**[42]	(Fukuda *et al.*, 1996)
Nep_I34a1 Nuclear export protein	[8]S**F**QDI**L**LR**M**SK**M**Q**L**[21]	(O'Neill *et al.*, 1998)
P53_Human Cellular tumor antigen p53	[336]E**R**FE**M**FRE**L**NEA**L**E**L**[350]	(Stommel *et al.*, 1999b)
Per2_Mouse Period circadian protein homolog 2	[456]S**V**QE**L**TEQ**I**HR**LL**M[469]	(Vielhaber *et al.*, 2001)
Rex_Htl1a Protein Rex	[78]S**M**DA**L**SAQ**L**YSS**L**S**L**[92]	(Bogerd *et al.*, 1996)

Table 7-1 Validated NES containing proteins

11 proteins were previously shown to have a CRM1 dependent NES. Regions of NESs are indicated with the starting and ending amino acid numbers, and the NES sequence. NES sequences were validated by the indicated reference. Putative Φ residues are marked bold.

These hits were evaluated for PKI-type NES content with the prediction algorithm. When NES scores from each protein were analyzed, the score from the true hit was the highest score or was one of the highest (Figure 7-5). For proteins RanBP1, cAMP-dependent protein kinase inhibitor α, and Rex there was only one hit identified. For all three cases this was the previously validated NES.

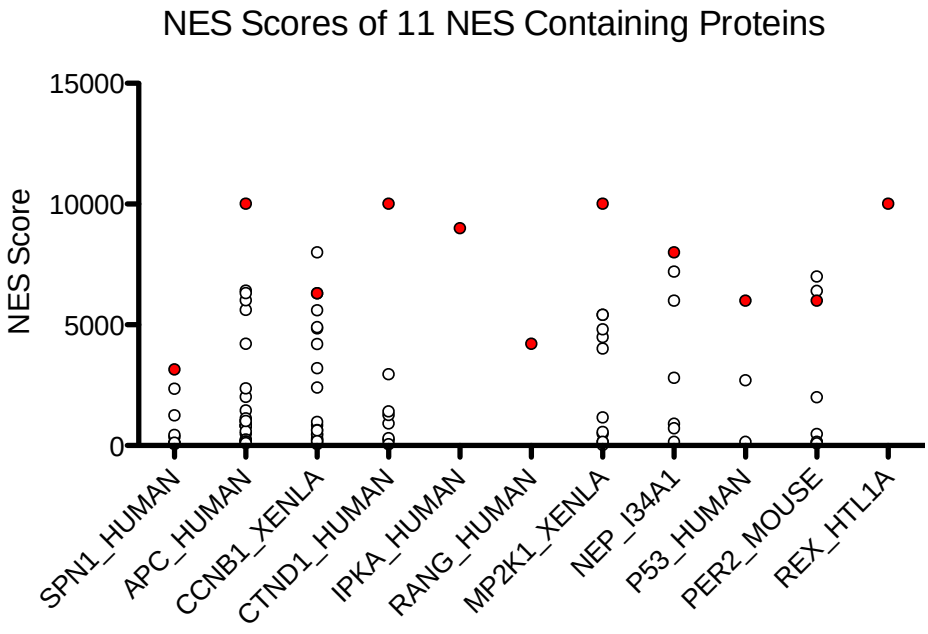

Figure 7-5 NES Scores of the PKI-type hits of the 11 selected proteins

NES scores were calculated for predicted PKI-type NES hits. Each hit from each protein is represented with a circle. The previously validated NESs are indicated with red color.

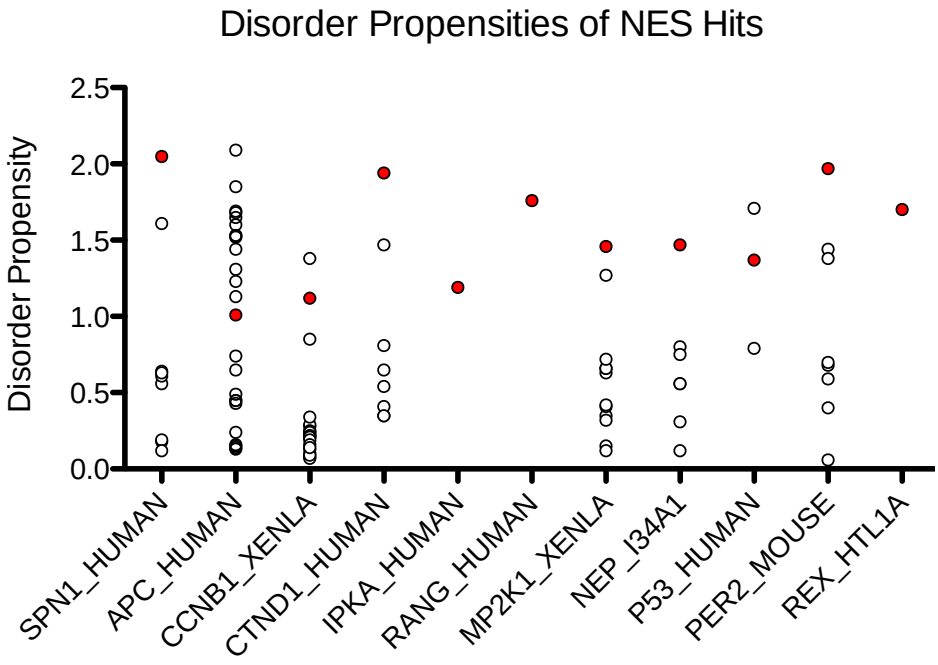

Figure 7-6 Aggregated disorder propensities of the PKI-type hits of the 11 selected proteins

The averaged disorder propensities of 6 amino acids before the hit, the NES hit, and 6 amino acids after the hit were summed (minimum of 0 and maximum of 3). NES scores were calculated for predicted PKI-type NES hits. Each hit from each protein is represented with a circle. The previously validated NESs are indicated with red color.

To analyze the disorder propensities of the hits, not only the hit itself but also the 6 amino acids before and after were considered. This led to an aggregated disorder score that had a value between 0 (order) and 3 (disorder). Disorder values for the previously confirmed hits were among the top ones except protein APC (Figure 7-6). APC is a very large protein with 2843 amino acids. C-terminal half has a high disorder propensity, and it is predicted to have many NES hits with weak NES score. If all hits from 11 proteins are plotted both with NES scores and disorder propensities, one can see the clear distinction between true and false hits (Figure 7-7).

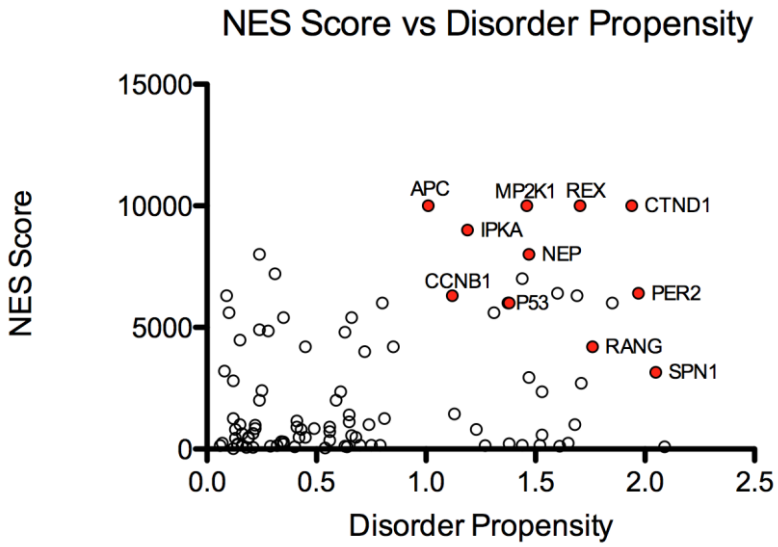

Figure 7-7 NES scores and disorder propensities of each PKI-type NES hit

7.1.5 Evaluation of REV-type NES prediction

The CRM1 crystal structure with the bound Rev NES revealed an unusual placement of Φ residues in CRM1 hydrophobic cleft. This was a new NES definition, so we wanted to see if there were previously annotated NES, which would fit into REV-type NES.

A second algorithm was written with Python (12.1.2). The algorithm iterates over the given sequences and outputs the predicted NES borders and sequences with disorder propensity, domain prediction for each hit sequence. Since there was no Φ position mutation study for this type of NES, thus we did not calculate an NES score for this type. Instead only a limited selection of hydrophobic residues was allowed at each Φ position (Figure 7-2).

To extend the number of examples and show that this pattern also exists in other proteins, 236 NES containing proteins from the curated database NESdb (Xu *et al.*, 2012a) were analyzed for their REV-type NES match. The hits matching with the annotated NESs are listed.

Protein	NES Sequence	Reference
Rev_Hv1h3 Protein Rev	^{75}**LP**P**L**ER**L**T**L**^{83}	(Meyer and Malim, 1994) (Güttler *et al.*, 2010)
Tf3a_Anaae Transcription factor IIIA	^{330}**LP**V**L**EN**L**T**L**^{338}	(Fridell *et al.*, 1996)
Ddx6_Xenla ATP-dependent RNA helicase ddx6	^{151}**IP**L**L**ER**L**D**L**^{159}	(Smillie and Sommerville, 2002)
gi 159024820 Nonstructural protein NS5	^{335}**VP**M**V**TQ**M**A**M**^{343}	(Rawlinson *et al.*, 2009)
Nf2l2_Human NF-E2-related factor 2	^{194}**IP**E**L**QC**L**N**I**^{202}	(Li *et al.*, 2006)
Q99AM3_HHV8 B-cell specific latent nuclear protein	^{552}**VP**L**V**IK**L**R**L**^{560}	(Munoz-Fontela *et al.*, 2005)
Fbx7_Human F-box only protein 7	^{326}**LP**D**V**FG**L**V**V**^{334}	(Nelson and Laman, 2011)

Table 7-2 REV-type NESs from NESdb

Interestingly there were 6 proteins with REV-type NESs other than Rev protein. This suggests that REV-type NESs might indeed represent a more general binding mode that is used by several proteins.

7.2 IDENTIFICATION OF NES ON eIF2β

For certain known CRM1 cargoes, the hitherto available prediction tools failed to identify a *bona fide* NES. We reasoned that our new algorithm might be more powerful than previous tools and used it to predict putative NESs on two proteins that are of general interest for our lab, human eIF2β and *Schizosaccharomyces pombe* Rna1p.

The first protein that was analyzed with the prediction algorithm was human eukaryotic translation initiation factor 2 subunit β (eIF2β). eIF2β is part of the trimeric eIF2 complex that is responsible for bringing the initiator methionine-tRNA to 40S ribosomal subunit. eIF2β was shown to accumulate in the nucleus upon Leptomycin B treatment, indicating CRM1 dependent nuclear exclusion (Bohnsack *et al.*, 2002). Since existing bioinformatics tools failed to predict testable NESs on eIF2β, Chandini Kadian from our lab was trying to experimentally narrow down the CRM1 interaction site on this protein.

7.2.1 Prediction of eIF2β NES hits

The primary sequence of the protein was analyzed with the PKI and REV-type prediction algorithms. There was no REV-type hit, but 3 PKI-type hits were predicted (Figure 7-8). Out these 3 hits, the second one was considered as a significant hit since it had the highest NES score (2880) and a high disorder propensity. This NES was not noticed before, because it had an alanine residue in its Φ_1 position, and alanine was not considered as a suitable amino acid for Φ positions by previous prediction tools.

IF2B_HUMAN							
Disorder			Sequence	Start	End	Domain	NES Score
B	NES	A					
3	2	2	R**K**KD**A**SDD**L**DD**L**N**F**	62	75	n.i.d.	1000
2	2	3	D**I**DE**A**EEG**V**KD**L**K**I**	90	103	n.i.d.	2880
2	2	2	R**D**YT**Y**EEL**L**NRV**F**N**I**	172	186	n.i.d.	270

Figure 7-8 Prediction of NES hits of human eIF2β

IF2B is the Uniprot ID for eukaryotic translation initiation factor 2 subunit β (eIF2β). From the 3 disorder values (B) represents the 6aa before NES, (NES) represents the predicted hit, and (A) represents the 6aa after NES. 'n.i.d' stands for 'not in any domain'. Putative Φ residues are marked bold.

7.2.2 Validation of eIF2β NES hit

We also analyzed the candidate NES for conservation among close species. To see the variance of NES hit sequence among vertebrate homologs, human eIF2β sequence was blasted against vertebrates, and aligned with the top hits (Figure 7-9). All hits followed the allowed sequences for Φ positions except one. *Xenopus laevis* had 2 homologs, and one of them (Q619h4_xenla) had a threonine residue instead of an alanine in the Φ_1 position.

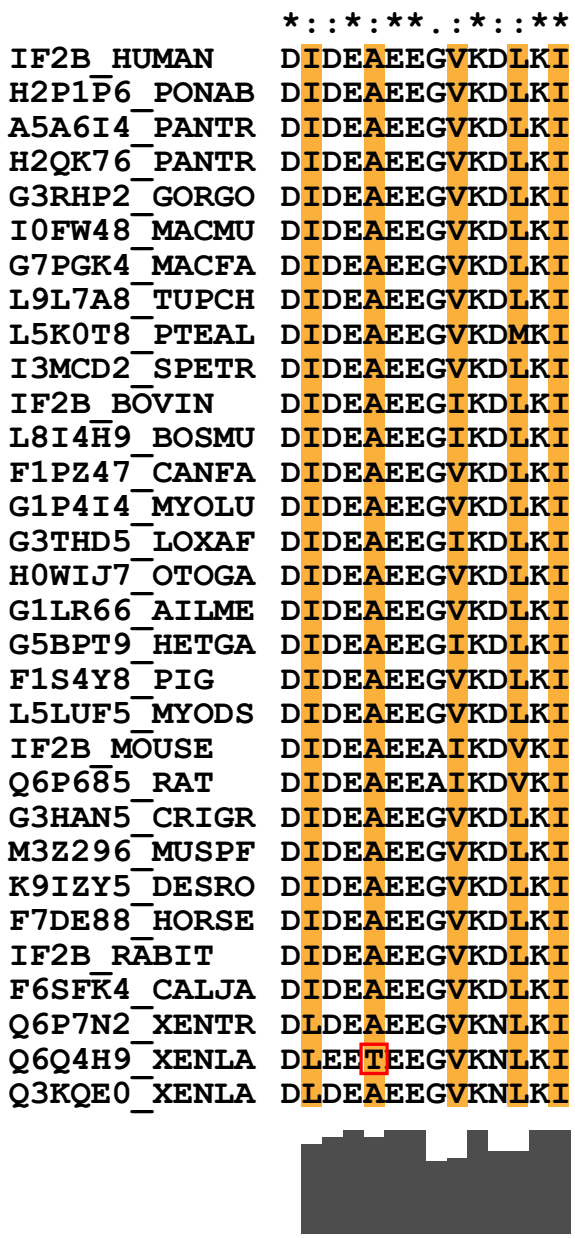

Figure 7-9 Alignment of hs eIF2β protein sequence with vertebrate orthologs

eIF2β orthologs were retrieved from Uniprot database and alignment was done with ClustalX 2.0 default settings. Φ residues are indicated with orange background. Proteins are named with Uniprot IDs.

To validate this putative NES, it was expressed in *E.coli* as His10-ZZ-Tev fusion to use in the RanGTP dependent CRM1 binding assays. To test the effect of the alanine in the Φ_1 position, a Φ_1 A⇒L mutant version of the putative NES was designed.

To tests its CRM1 binding, residues corresponding to *Xenopus laevis* eIF2β NES hit were also expresses as His10-ZZ-Tev fusion. Human eIF2β NES hit, the Φ_1 A⇒L mutant, the Φ_1 A⇒T *Xenopus laevis* ortholog, and the PKI NES were incubated with mmCRM1 either in the presence or absence of $Ran_{5\text{-}180}$ Q69L GTP.

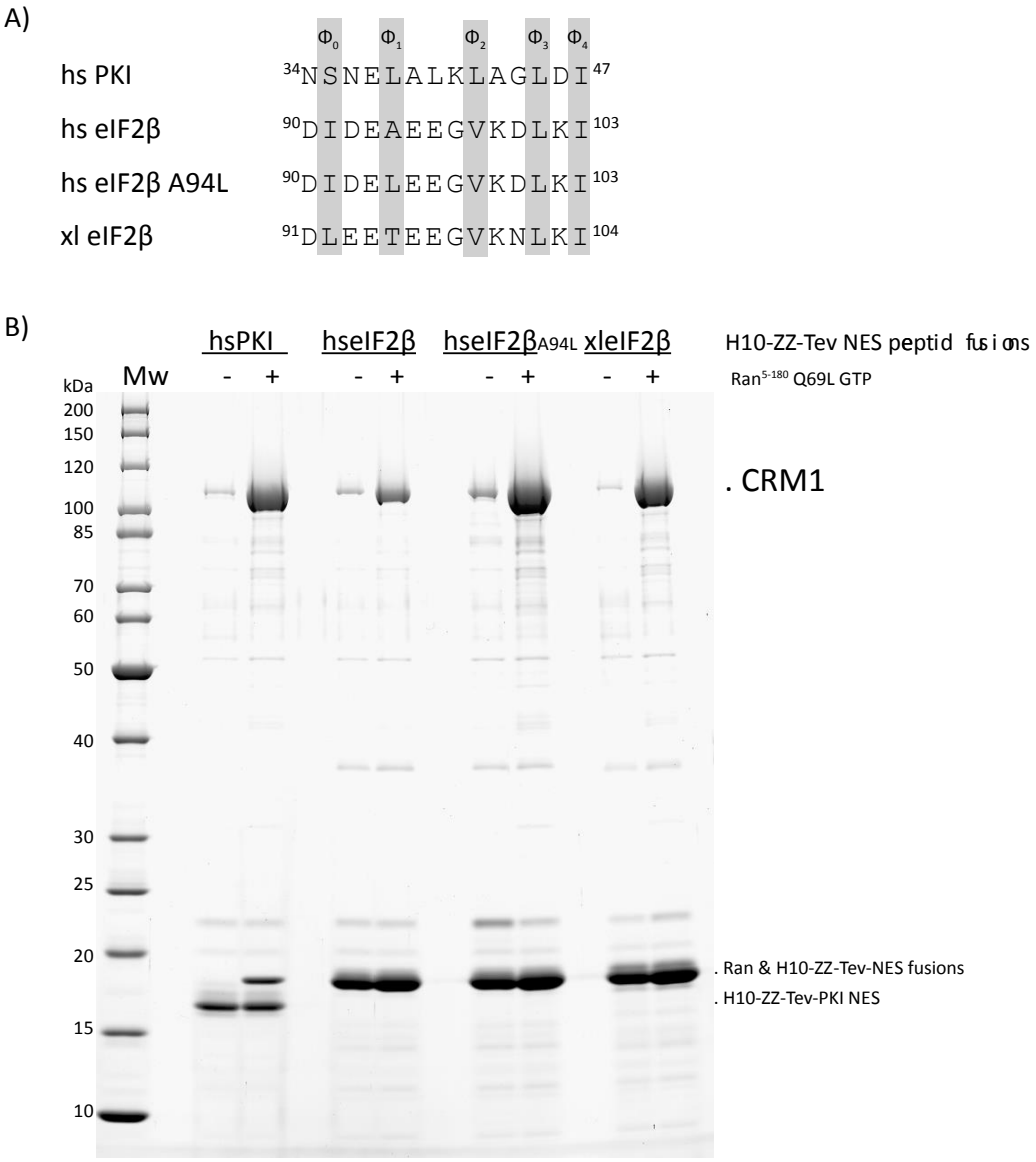

Figure 7-10 RanGTP dependent mmCRM1 binding of eIF2β NES hits

The binding reactions were performed in 500 µl volume with 2 µM mmCRM1 and 2 µM His10-ZZ-Tev-NES fusion. For the reactions with Ran, 3 µM Ran_{5-180} Q69L GTP was added. Final buffer concentration was adjusted to 50 mM Tris/HCl 7.5, 120 mM NaCl, 2 mM $Mg(OAc)_2$, 5 mM DTT. After 2 hours at 4°C, ZZ-affibody beads were added to pull down the NES peptides and the bound proteins.

A) Alignment of PKI NES to human and *Xenopus laevis* eIF2β. Φ residues are indicated above the PKI NES and their alignments with NES hits are shaded gray. Start and end residue numbers of NESs are indicated in the full-length protein context. B) SDS-PAGE analysis of RanGTP dependent mmCRM1 binding of NES hits. H10 stands for N-terminal 10 histidine residues, Z (in ZZ) stands for IgG-binding domain of the Staphylococcal protein A, and Tev stands for Tobacco Etch Virus protease recognition sequence. Protein ladder is abbreviated with 'Mw' for molecular weight, and protein sizes are indicated on the left side of the corresponding bands. Samples without RanGTP are indicated with a '–' sign, and samples with RanGTP are indicated with '+' sign.

Beads were eluted with 100 µl SDS sample buffer, and 10 µl of elution was analyzed with SDS-PAGE. PKI NES served as the positive control and showed RanGTP dependent CRM1 interaction. The candidate NES of eIF2β was weaker than PKI in the CRM1 interaction, but still showed a significant binding. This binding was enhanced when the Φ_1 residue was mutated from alanine to

leucine, indicating that this was truly a Φ pocket binding position. Interestingly the *Xenopus laevis* homolog of the NES was also functional with a threonine in Φ_1 position (Figure 7-10).

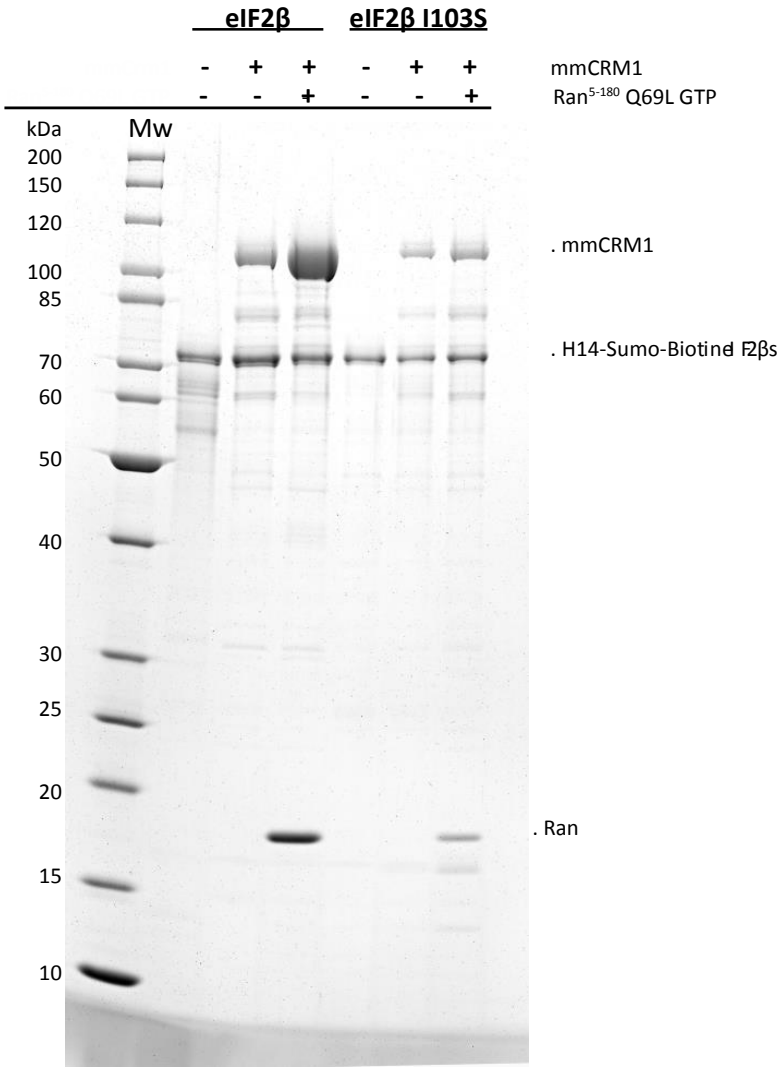

Figure 7-11 RanGTP dependent mmCRM1 binding of wild type and NES mutant of human eIF2β

H10 stands for N-terminal 10 histidine residues, Z (in ZZ) stands for IgG-binding domain of the Staphylococcal protein A. Protein ladder is abbreviated with 'Mw' for molecular weight, and protein sizes are indicated on the left side of the corresponding bands. Samples without RanGTP are indicated with a '–' sign, and samples with RanGTP are indicated with '+' sign.

It was also important to test the validity of the NES in the context of full-length protein. To test this, the Φ_4 position was mutated from isoleucine to serine, and both the wild type human eIF2β and the I103S mutant were expressed as biotinylated full-length proteins. Proteins were incubated with mmCRM1 either in the presence or absence of RanGTP, and the final salt concentration was adjusted to 100 mM NaCl. Reaction was incubated with Streptavidin-Agarose beads. Elution was done with 37°C SDS sample buffer which is enough for elution of mmCRM1 and Ran but cannot fully disrupt streptavidin-biotin interaction. Elutions were analyzed with SDS-PAGE (Figure 7-11).

Taken together, with the prediction algorithm we identified 3 NES hits for human eIF2β and analyzed the most prominent hit in terms of NES score. This NES hit was indeed able to bind CRM1. By mutation of the Φ4 position in the full-length protein, we were able to abolish the interaction of full-length eIF2β with CRM1. The predicted hit had and unusual alanine residue in Φ_1, confirmed by the leucine mutant that had a stronger affinity for CRM1. Thus, our algorithm proved useful to predict a functional NESs that has escaped previous attempts of bioinformatics analysis.

7.3 IDENTIFICATION OF NES ON spRna1p

Second protein of interest was spRna1p, which is the RanGAP ortholog in *Schizosaccharomyces pombe*. spRna1p has an N-terminal leucine rich repeat (LRR) domain followed by a poly glutamic acid region (Figure 6-5). It is kept cytoplasmic in *S.pombe* (Melchior *et al.*, 1993b). Previous studies suggest Crm1 mediated export but the NES region was not experimentally identified for spRna1p.

7.3.1 Previously suggested NESs are buried in the structure

The *Saccharomyces cerevisiae* homolog of Rna1p (scRna1p) was previously shown to interact with Crm1 and also two NESs were identified. By homology, 2 corresponding NESs were also suggested for spRna1p (Feng *et al.*, 1999). In the same year, the N-terminal LRR region of *S.pombe* Rna1p protein was crystallized (Hillig *et al.*, 1999). In order to see if the suggested regions would serve as an NES also in spRna1p, we analyzed the crystal structure for the accessibility of the previously suggested NES regions.

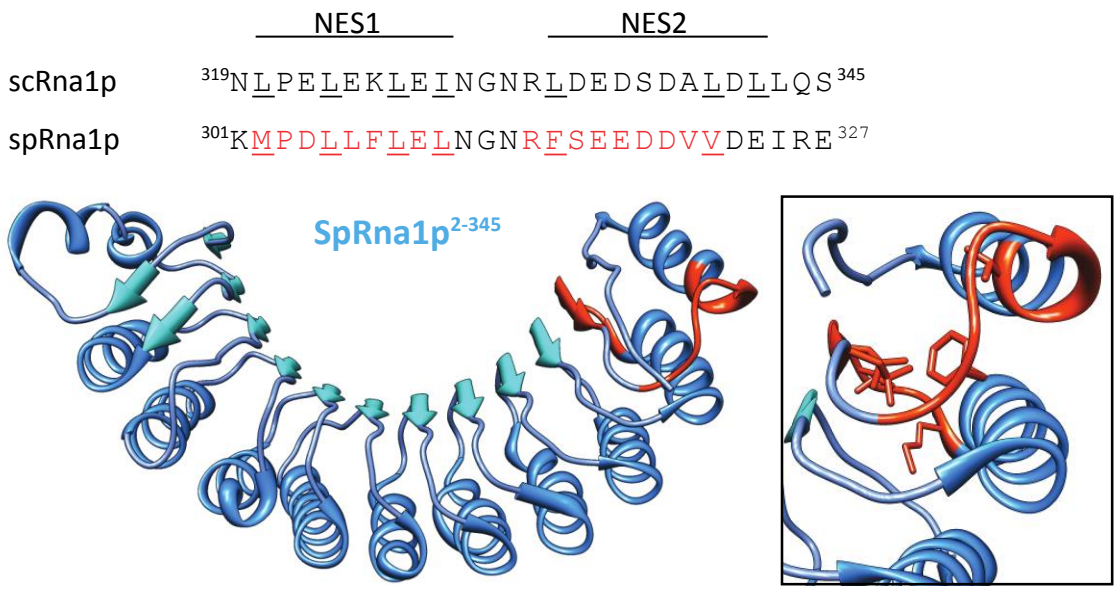

Figure 7-12 spRna1p structure and previously suggested NESs

scRna1p and spRna1p sequences aligned for the region covering previously suggested NESs on scRna1p. On top of the sequences, 2 NESs of scRna1p; NES1 and NES2 are indicated. Corresponding regions are marked orange-red on spRna1p sequence. Residues important for Crm1 interaction are underlined black on scRna1p sequence. Corresponding residues on spRna1p sequence are underlined orange-red. These orange-red regions are also shown on the spRna1p structure (PDB structure 1K5D Chain C). Side chains for underlined residues are show in orange red.

The two suggested NES regions were part of the crystalized spRna1p. They are located within well folded leucine-rich repeat regions and the hydrophobic residues that would be important for Crm1 interaction are clearly buried in the structure (Figure 7-12). For recognizing such "NES", Crm1 would have to locally unfold its cargo, which is not a very plausible scenario. We therefore

reasoned that the true NES might have escaped detection and so we used the new prediction algorithm to analyze the primary sequence of spRna1p for alternative NES hits.

7.3.2 Prediction of spRna1p NES hits

spRna1p is 386 aa long and the first 340 aa are made up of 8 leucine rich repeats (LRRs). Since NESs can also be rich in leucine residues, it is highly likely to find an NES hit on LRR region of the protein. When the sequence was analyzed with the prediction algorithm, 7 PKI-type and 1 REV-type NES hits were identified. Indeed an NES was predicted for all LRRs except the first one. Since 7 of them were in the previously crystallized LRR domain of the protein, we focused on the last hit. This was the very last 14 residues of the protein and had an unusual alanine residue in its Φ_3 position. It had no hydrophobic residue in Φ_0 position but the neighboring residues were acidic. In total it had a low NES score, but it was the only hit that had a high disorder propensity (Figure 7-13).

RNA1_SCHPO							
Disorder			Sequence	Start	End	Domain	NES Score
B	NES	A					
1	1	1	EIPEALRLLLQALL	78	91	n.i.d.	600
2	2	2	AGAKIARALQELAV	138	151	in LRR	5600
1	1	1	RPEGIEHLLLEGLAY	200	214	n.i.d.	800
1	1	1	EGLAYCQELKVLDL	210	223	in LRR	1000
1	1	1	AVVDAFSKLENIGL	263	276	in LRR	1800
1	1	1	GLQTLRLQYNEIEL	275	288	in LRR	2700
1	1	1	MPDLLFLEL	302	310	n.i.d.	REV-type
3	3	3	E**D**KE**L**ADE**L**SK**A**H**I**	373	386	n.i.d.	1800

Figure 7-13 NES prediction for spRna1p

Schpo is the Uniprot abbreviation for organism *Schizosaccharomyces pombe*. From the 3 disorder values (B) represents the 6aa before NES, (NES) represents the predicted hit, and (A) represents the 6aa after NES. 'n.i.d' stands for 'not in any domain'. Hits that reside in the previously crystalized region are shaded gray.

7.3.3 Validation of spRna1p NES hit

The first experiment to validate the NES candidate was a C-terminal truncation of spRna1p. Both full-length spRna1p and spRna1p lacking the last 3 amino acids (spRna1p$^{\Delta C3}$) were expressed as His10-ZZ-Tev fusions. The last 3 amino acids covered the Φ_3 and Φ_4 positions. The candidate NES hit and PKI NES were also expressed with the same tag. We performed binding assays with scCrm1 either in the absence or presence of RanGTP (Figure 7-14).

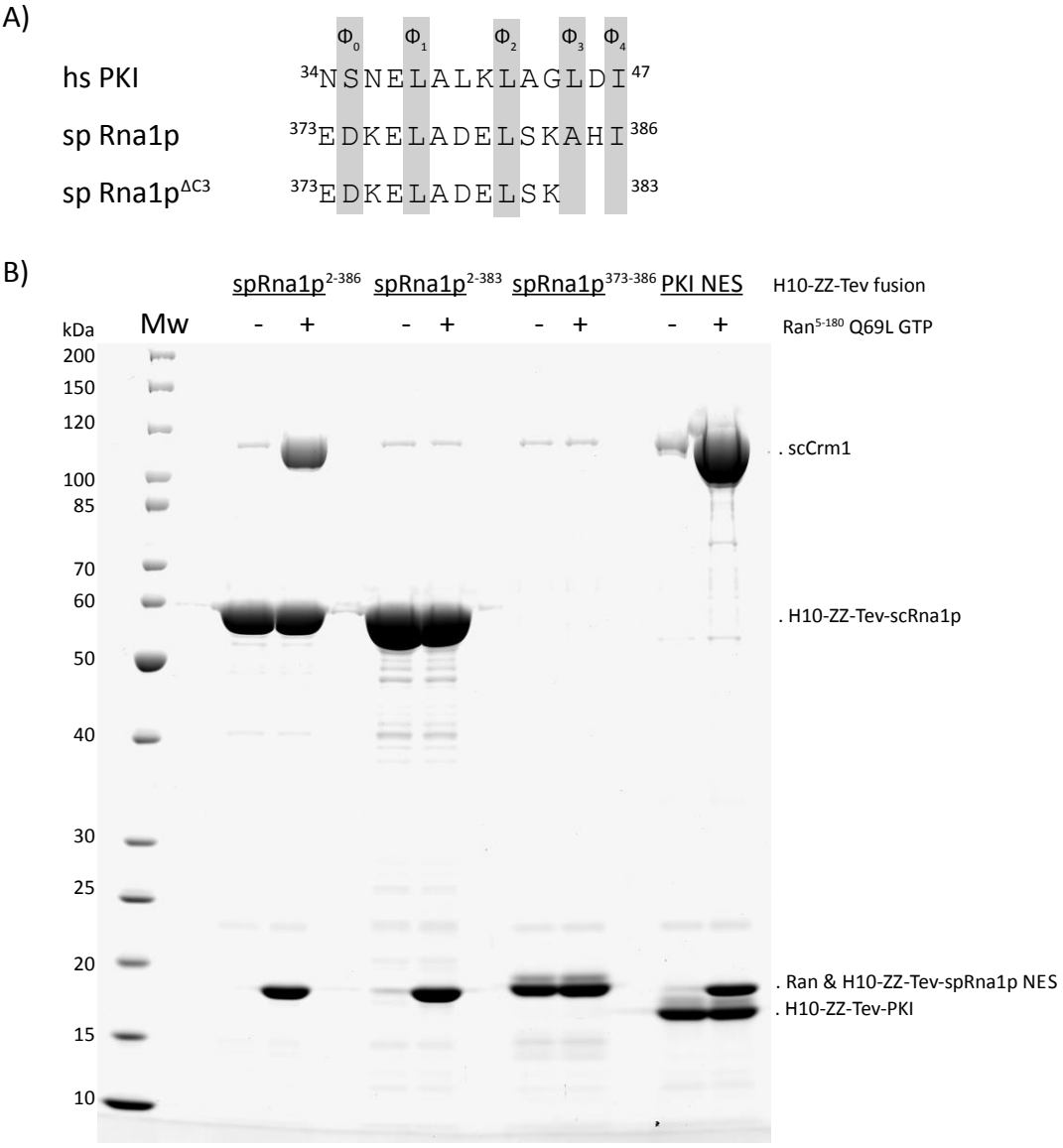

Figure 7-14 RanGTP dependent scCrm1 binding of spRna1p

The reactions were performed in 500 µl volume with 2 µM scCrm1 and 2 µM H10-ZZ-Tev-NES fusion cargoes. For the reactions with RanGTP, 3 µM Ran_{5-180} Q69L GTP was added. Final buffer concentration was adjusted to 50 mM Tris/HCl 7.5, 60 mM NaCl, 2 mM $Mg(OAc)_2$, 5 mM DTT. After 2 hours of incubation at 4°C, ZZ-affibody beads were added to pull down the cargo and the bound proteins. Elutions were analyzed with SDS-PAGE.

A) Alignment of PKI NES to spRna1p. Φ residues are indicated above the PKI NES and their alignments with NES hits are shaded gray. Start and end residue numbers of NESs are indicated in the full-length protein context. B) SDS-PAGE analysis of RanGTP dependent *Saccharomyces cerevisiae* Crm1 (scCrm1) binding of NES hits. H10 stands for N-terminal 10 histidine residues, Z (in ZZ) stands for IgG-binding domain of the Staphylococcal protein A, and Tev stands for Tobacco Etch Virus protease recognition sequence. Protein ladder is abbreviated with 'Mw' for molecular weight, and protein sizes are indicated on the left side of the corresponding bands. Samples without RanGTP are indicated with a '–' sign, and samples with RanGTP are indicated with '+' sign.

spRna1p shows a significant RanGTP dependent scCrm1 binding. This binding is lost upon removal of last 3 amino acids from spRna1p. Although scCrm1 binding was gone, Ran was still eluted with the $spRna1p^{\Delta C3}$, due to background binding of Ran to ZZ-affibody beads at low salt conditions. Interestingly, the isolated NES did not show RanGTP-dependent Crm1 binding, suggesting that the functional export signal is bipartite and includes additional parts of the Rna1p molecule.

To further validate the NES-like sequence within the last 14 residues, we rationally designed point mutants of the full-length protein. If it were the NES, binding strength would respond to the mutations in the Φ position. We anticipated the following potential outcomes: Mutation of the predicted Φ_4 position from isoleucine to serine would weaken while mutating the alanine in Φ_1 position to leucine should further strengthen the binding.

The binding assay was done with streptavidin agarose beads, which show less background binding of RanGTP. A biotinylated version of scCrm1 and untagged spRna1p$^{2\text{-}386}$, spRna1p$^{\Delta C3}$, spRna1p$^{2\text{-}386}$ I386S, spRna1p$^{2\text{-}386}$ A384L were used for the assay (Figure 7-15).

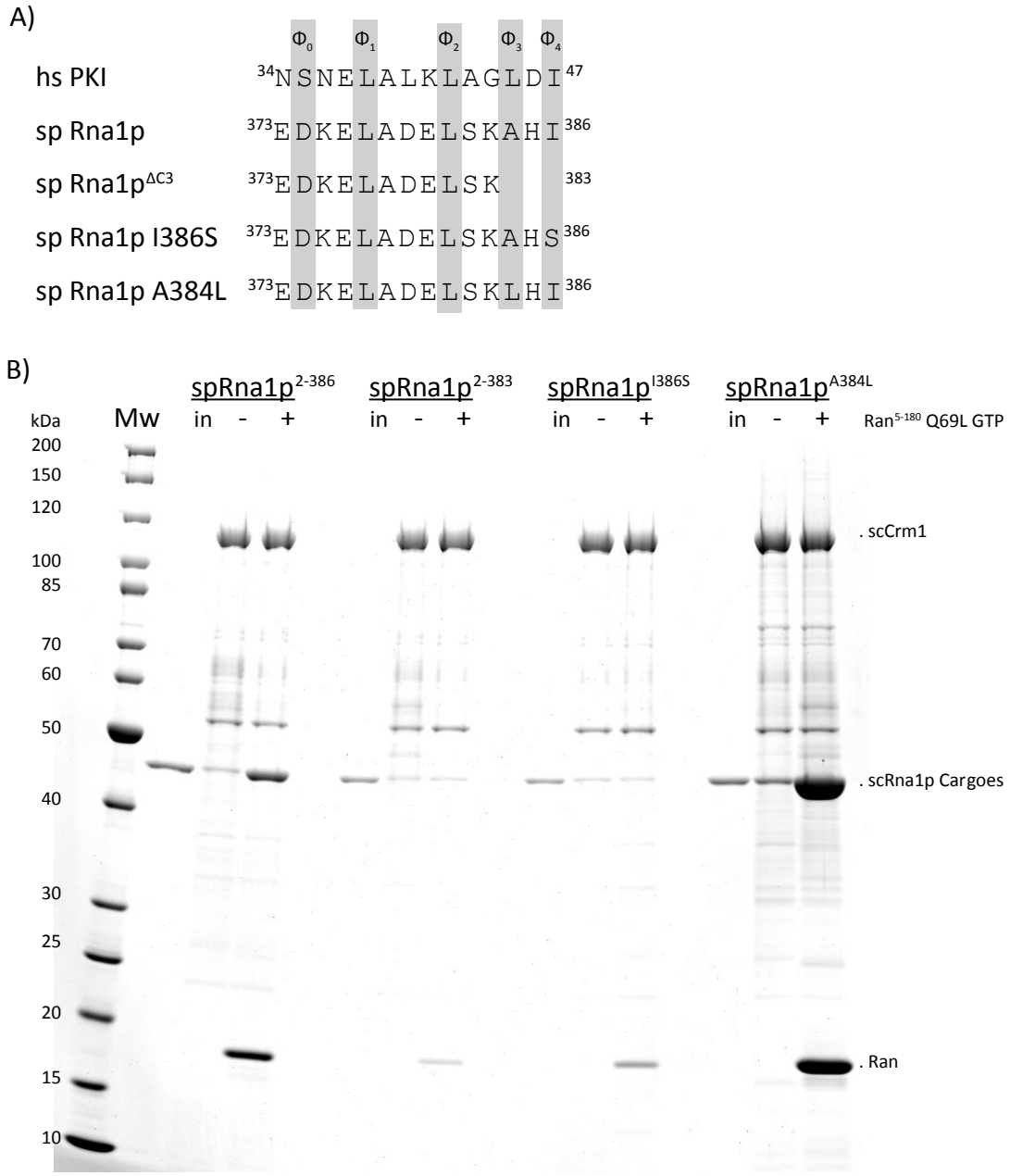

Figure 7-15 RanGTP dependent scCrm1 binding of spRna1p and point mutants

The reactions were performed in 500 µl volume with 2 µM biotinylated scCrm1 and 2 µM cargo. For the reactions with RanGTP, 3 µM Ran_{5-180} Q69L GTP was added. Final buffer concentration was adjusted to 50 mM Tris/HCl 7.5, 60 mM NaCl, 2 mM $Mg(OAc)_2$, 5 mM DTT. After 2 hours of incubation at 4°C, streptavidin-agarose beads were added to pull down the cargo and the bound proteins. Elution was done with 37°C SDS sample buffer which is enough for elution of cargo and Ran but did not fully disrupt streptavidin-biotin interaction. Elutions were analyzed with SDS-PAGE.
A) Alignment of PKI NES to spRna1p wt and mutants. Φ residues are indicated above the PKI NES and their alignments with NES hits are shaded gray. Start and end residue numbers of NESs are indicated in the full-length protein context. B) SDS-PAGE analysis of RanGTP dependent scCRM1 binding of NES hits. H10 stands for N-terminal 10 histidine residues, Z (in ZZ) stands for IgG-binding domain of the Staphylococcal protein A, and Tev stands for Tobacco Etch Virus protease recognition sequence. Protein ladder is abbreviated with 'Mw' for molecular weight, and protein sizes are indicated on the left side of the corresponding bands. Samples without RanGTP are indicated with a '–' sign, and samples with RanGTP are indicated with '+' sign.

Full-length spRna1p was bound to the beads in the presence of RanGTP, and was not bound when RanGTP was absent. . The previously described mutant lacking the 3 C-terminal residues served as an additional control. Interestingly a similar result was obtained upon mutating the Φ_4 position

was mutated from isoleucine to serine. Consistent with this, when the Φ_1 was mutated from alanine to leucine the binding was greatly enhanced. These experiments strongly suggest that the NES hit was identified with the correct Φ spacing.

Testing the localization phenotype of the protein in a cellular context would provide a more stringent test. Therefore the spRna1p NES was also tested by transient transfection of HeLa cell. To test the experimental setup, three PKI NES versions, wt PKI, super PKI (sPKI) with enhanced CRM1 binding, and PKI Φ_4A mutant with weaker CRM1 binding, were fused to GFP. NLS and NES of eIF2β was fused to mCherry and used as control. Each GFP vector was cotransfected with a mCherry-eIF2β NES, and cells were fixed after 24h. DAPI was used to stain DNA. wtPKI showed a prominent nuclear exclusion, and the PKI Φ_4A mutant lost this exclusion. sPKI was localized to nuclear rim. Since sPKI can bind to CRM1 without RanGTP, disassembly of the export complex is inefficient, and it was stalled at the NPCs. CRM1 was blocked by sPKI, which resulted in mislocalization of the control NES fusion (Figure 7-16).

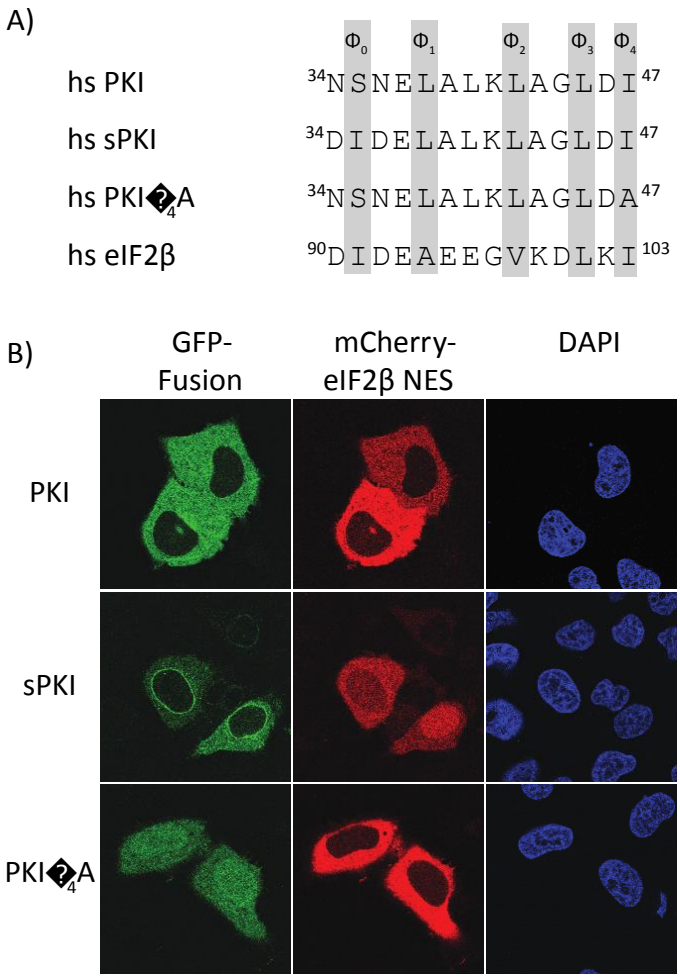

Figure 7-16 Phenotypic outcomes of different GFP-NES fusions

3 versions of PKI were fused to GFP, and cotransfected with the mCherry fusion with eIF2β65-114, which contains both an NLS and an NES. DAPI staining was used for DNA.

To test the effect of the NES hit on spRna1p localization, wt and ΔNES Rna1p were fused to GFP, and cotransfected with NLS and NES of eIF2β fused to mCherry. GFP-spRna1p was fully cytoplasmic and deletion of NES on Rna1p resulted in nuclear leakage of the protein. To enhance the effect of NES deletion, an SV40 NLS was fused between GFP and wt Rna1p and Rna1p ΔNES. GFP-SV40NLS-Rna1p was cytoplasmic, whereas deletion of the C terminal NES resulted in total nuclear accumulation (Figure 7-17).

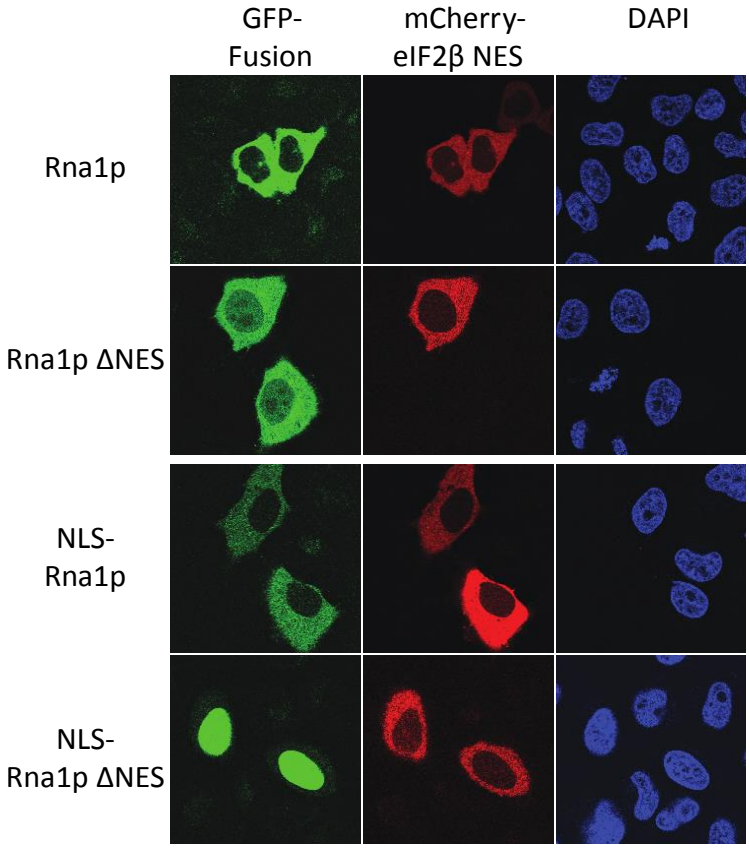

Figure 7-17 Localization of different GFP spRna1p fusions

Full-length spRna1p and ΔNES version was fused to GFP or GFP-SV40 NLS, and cotransfected with mCherry fusion of eIF2β65-114 that contains both an NLS and an NES. DAPI staining was used for DNA.

The very C-terminal 14 residues stretch is the NES of spRna1p as confirmed by the binding experiments and HeLa transfections. The full-length spRna1p has a stronger binding than the NES peptide itself. This is an indication that rest of the protein, either the LRRs or the poly glutamic acid region, or both contribute to its interaction with Crm1.

spRna1p was the second protein that was analyzed for an NES with the new prediction algorithm. There was only one hit that was out of the previously crystalized region. It had a low NES score since it had an alanine in its Φ_3 position. This NES hit with the low NES score was indeed a poor

CRM1 binder, but in the context of full-length protein, it was strong enough to sustain a RanGTP dependent binding to and very efficient export by Crm1.

7.4 RANGTP DEPENDENT CRM1 BINDERS FROM CYTOSOLIC HELA EXTRACT

We were further interested in a more complete picture of the RanGTP dependent CRM1 binders. Although it has been known that CRM1 is the most versatile NTR, it was not clear how large the pool of CRM1 exported cargoes actually is. This pool probably comprises not only direct CRM1 binders, but also interaction partners of direct binders. Direct binders may function as adaptors for export of larger complexes. To address these questions, we designed and optimized affinity chromatography on immobilized CRM1. We used extracts from a human cell line as a starting material and mass spectrometry (MS) for identifying candidate cargoes from the eluates.

7.4.1 CRM1 affinity chromatography of cytoplasmic HeLa extract

To come up with a complete list of CRM1 exported cargoes, we used the cytoplasmic S10 HeLa extract as the source of RanGTP-dependent CRM1 binders. For a low background binding we optimized the binding conditions and the chromatography system. HeLa extract showed quite high background binding with IgG-sepharose, or anti ZZ affibody-silica beads. Tests on streptavidin-agarose beads gave much better results.

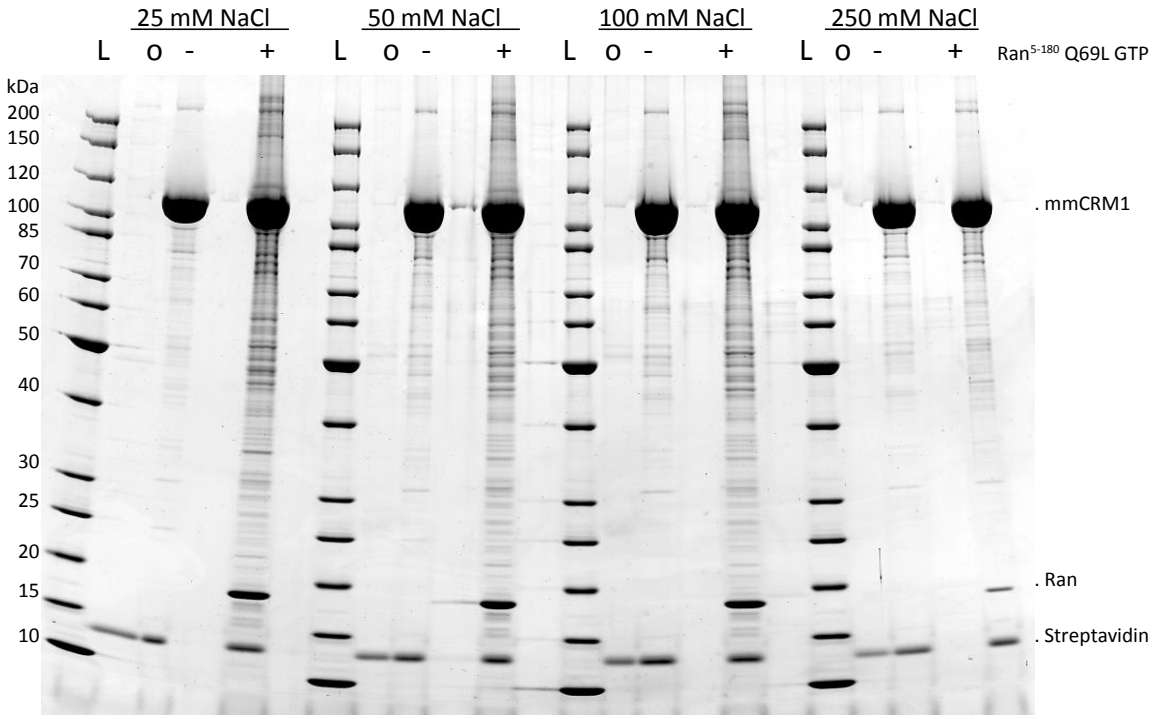

Figure 7-18 Salt sensitivity of RanGTP dependent mmCRM1 interaction of cargoes from HeLa S10 extract.

Affinity chromatography of HeLa S10 extract with Biotin-CRM1 immobilized streptavidin-agarose beads.
For each reaction 0.5 nmol of biotin-CRM1 was immobilized on 20 µl of streptavidin-agarose beads, and washed with free biotin. Binding reaction was performed in a final volume of 500 µl for 3h at 4°C with rotation. For each binding experiment 100 µl of cytoplasmic HeLa S10 extract was used. For samples with RanGTP, 2 µM Ran_{5-180} Q69L GTP was added. 20 µl beads were boiled in 100 µl SDS sample buffer to retrieve all bound material.
'L' represents the protein ladder, and the corresponding sizes of the ladder bands are indicated on the left side of the first ladder band. Absence or presence of RanGTP is indicated with '-' or '+' sign. Background binding of HeLa extract to

the streptavidin-agarose beads without CRM1 is indicated with 'o' sign. Affinity chromatography was done at different salt concentrations as indicated above the gel.

For CRM1 affinity chromatography, a biotinylated version of mmCRM1 was bacterially expressed and immobilized to the streptavidin-agarose beads. The beads were then washed with free biotin to block remaining biotin binding sites. These beads were used in affinity chromatography of cargos from cytoplasmic HeLa S10 extract either in the absence or presence of RanGTP. To optimize the incubation conditions, a salt screen was performed at 25, 50, 100 and 250 mM NaCl concentrations.

At 250 mM NaCl concentration, the RanGTP dependent binding was mostly gone, as also evident from the intensity of the Ran band. RanGTP dependence of the binders was most impressive at low salt concentrations (Figure 7-18).

7.4.2 CRM1 Affinity Chromatography for SILAC-Based Mass Spectrometry

The most convenient way of analyzing these global protein pools was mass spectrometry. That's why we further analyzed Ran- and Ran+ lanes with MS to resolve the highly complex pool of proteins in the elutions. CRM1 binding can be a salt sensitive interaction. PKI NES would bind only in the presence of RanGTP when 100 mM NaCl is used in the binding conditions. As salt concentration goes up, this binding will become weaker. And if salt concentration goes down, a RanGTP independent CRM1 interaction will emerge. That would mean that affinity chromatography at low salt conditions might favor cargo binding also in the absence of RanGTP. Although a very prominent difference was visible between RanGTP+ and RanGTP- lanes at 25 mM NaCl condition, it still required comparison of protein levels between two samples to clearly identify the RanGTP dependent binders (Figure 7-18). Non-quantitative MS analysis is suited for identification of the proteins in a given sample, however for comparison of protein levels in two different samples, it is not sufficient.

Many different MS methods have been applied for quantification of the protein amounts in the sample (Wilm, 2009). Among them, stable isotope labeling by amino acids in cell culture (SILAC) has proven to be a useful technique in analysis of many proteomics studies (Ong and Mann, 2006). SILAC is a metabolic labeling method where natural amino acids lysine and arginine ('light') are replaced with their ^{2}H, ^{13}C and ^{15}N labeled forms ('heavy') in cell culture. With light and heavy cells I refer to cells grown in the corresponding media. Cytoplasmic extracts prepared from cells grown in either light or heavy medium were kindly provided by Miroslav Nikolov from Mass Spectrometry Research Group, MPI-BPC.

SILAC based comparison of Ran- and Ran+ samples depends on the assumption that metabolic labeling by heavy amino acids does not change the protein composition of the cell extract. To test this assumption, identical volumes of light and heavy extracts were mixed, and run on SDS-PAGE and the complete lane was analyzed with MS (Figure 7-19).

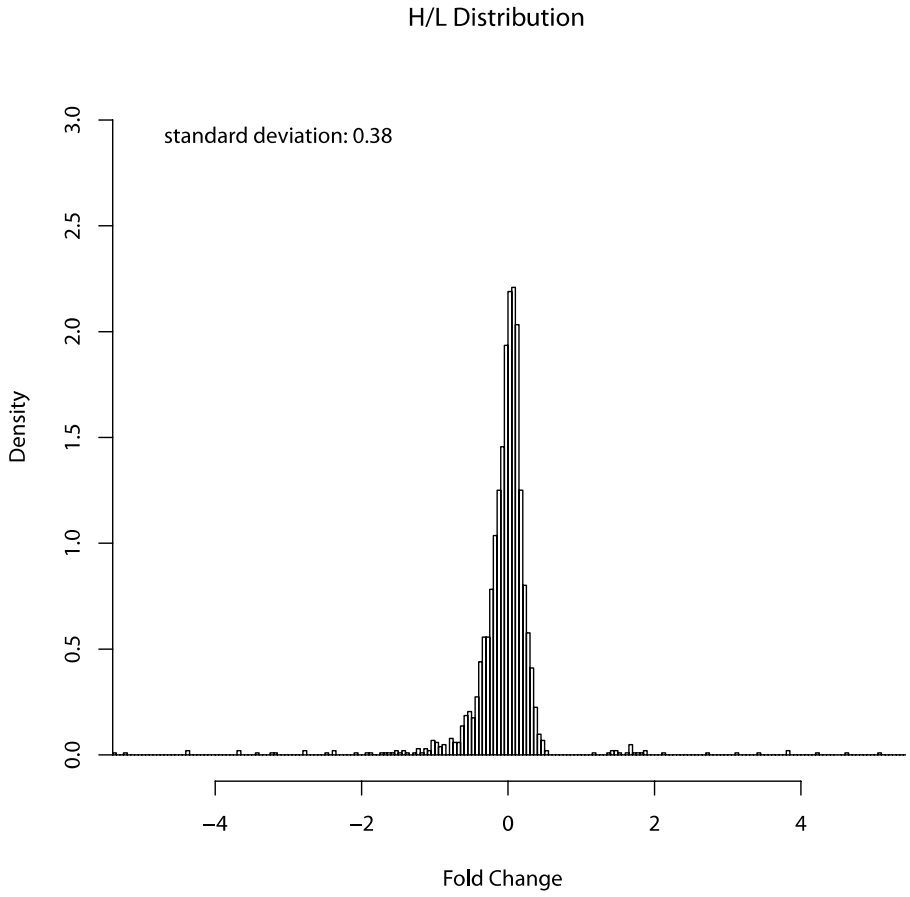

Figure 7-19 Distribution of protein ratios in heavy and light HeLa extracts
SILAC ratios were calculated from a 1 to 1 mixture of heavy and light extracts, and distribution of $\log_2$(Ratios) were subjected to a Gaussian distribution analysis for the standard deviation.

More than 2800 proteins were identified from cytoplasmic HeLa extracts. Although this does not cover whole cytoplasmic proteome, it contained sufficient number of proteins to obtain reliable statistics on the proteins levels. Proteins were present in very similar amount in both heavy and light extracts with a standard deviation of 0.38 for $\log_2$(Heavy/Light) values.

For each reaction 0.5 nmol of biotin-mmCRM1 was immobilized on 20 µl of streptavidin-agarose beads, and washed with free biotin. Binding reaction was done in a final volume of 500 µl for 3h at 4°C with rotation. For each binding experiment 100 µl of light or heavy cytoplasmic HeLa extract was used. For samples with RanGTP, 2 µM $Ran_{5\text{-}180}$ Q69L GTP was added. Elution was done with 37°C SDS sample buffer which is enough for elution of cargoes and Ran but does not fully dissociate biotinylated mmCRM1 from streptavidin, and the most intense band on SDS-PAGE was

suppressed. Since the total eluate had considerably low CRM1 amounts, it was more representative for the CRM1 binders. This also enables an easier identification on MS for proteins co-migrating with CRM1 band. 10 µl of eluate was analyzed on SDS-PAGE (Figure 7-20).

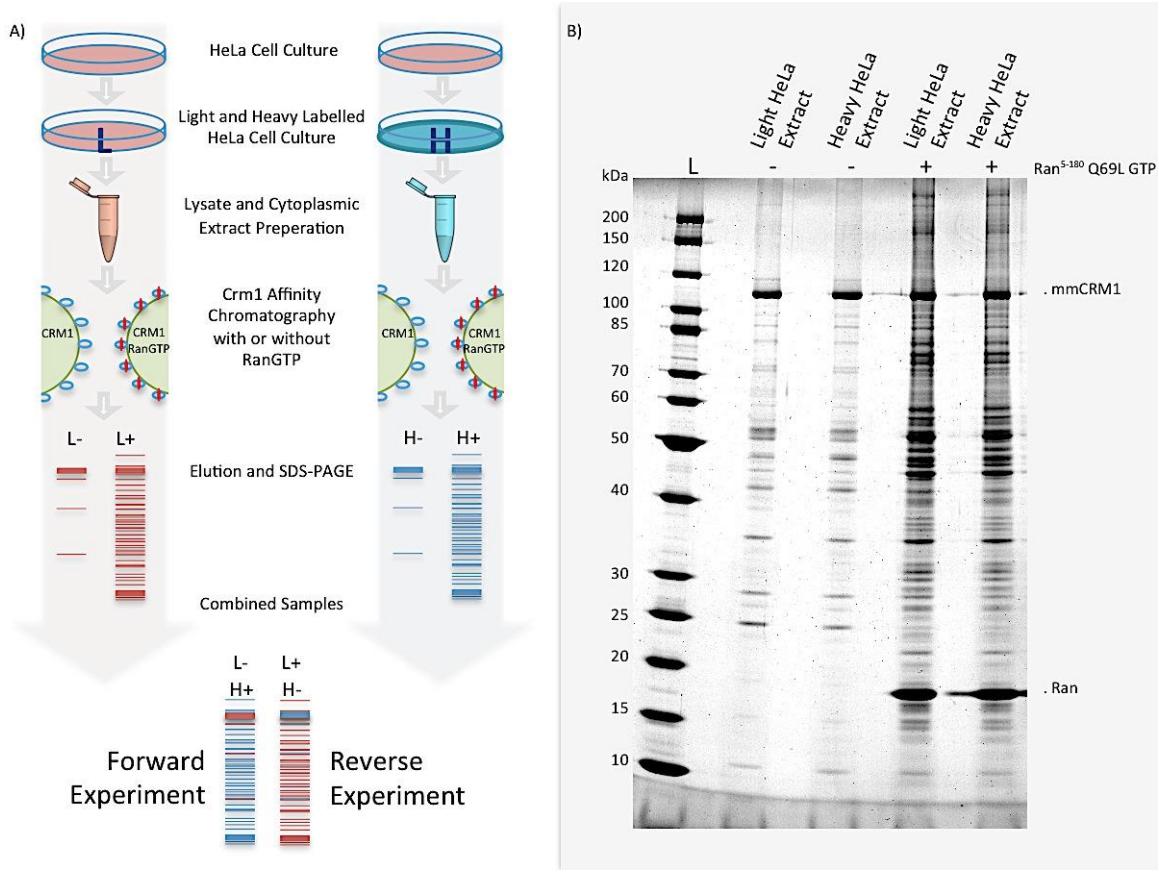

Figure 7-20 CRM1 affinity chromatography of cytoplasmic HeLa extracts produced with SILAC method

A) Schematic depiction of SILAC CRM1 affinity chromatography experiment. B) Affinity chromatography of light and heavy labeled cytoplasmic HeLa S10 extract with Biotin-mmCRM1 immobilized streptavidin-agarose beads. 'L' represents the protein ladder, and the corresponding sizes of the ladder bands are indicated on the left side of the first ladder band. Absence or presence of RanGTP is indicated with '-' or '+' sign. Extracts produced from light and heavy labeled cells are indicated as "Light HeLa Extract" or "Heavy HeLa Extract".

With SILAC based methods, it is possible to analyze a light sample together with a heavy sample at the same time with MS. To compare binding of cargoes in the presence or absence of RanGTP to CRM1, different Ran states of light and heavy elutions are combined. Elution of CRM1 binders from light extract in the absence of RanGTP (L- sample) was mixed with Elution of CRM1 binders from heavy extract in the presence of RanGTP (H+ sample), and vice versa, Elution of CRM1 binders from heavy extract in the absence of RanGTP (H- sample) was mixed with Elution of CRM1 binders from light extract in the presence of RanGTP (L+ sample). The former is called "forward experiment" and the latter is called "reverse experiment". Since MS analysis can distinguish between peptides from heavy and light extracts by residue specific mass difference, in a single analysis L- sample was compared to H+ sample and also H- sample was compared to L+ sample. With this experimental setup we not only minimized the error of MS analysis by a streamlined

processes of samples to be compared, but also repeated the experiment and MS analysis 2 times (forward and reverse experiments). Mass spectrometry analysis was carried out by Samir Karaca from the Mass Spectrometry Research Group of the MPI-BPC.

7.4.3 Mass Spectrometry Analysis of SILAC CRM1 Affinity Chromatography Eluates

Forward and reverse experiments were analyzed together. In total there were 3070 proteins with unique Uniprot identifiers. The identified protein levels in Ran+ samples were compared to Ran- samples, a ratio value for each protein was obtained, and values were analyzed as $\log_2$ values to reflect the fold changes in protein amounts (Figure 7-21).

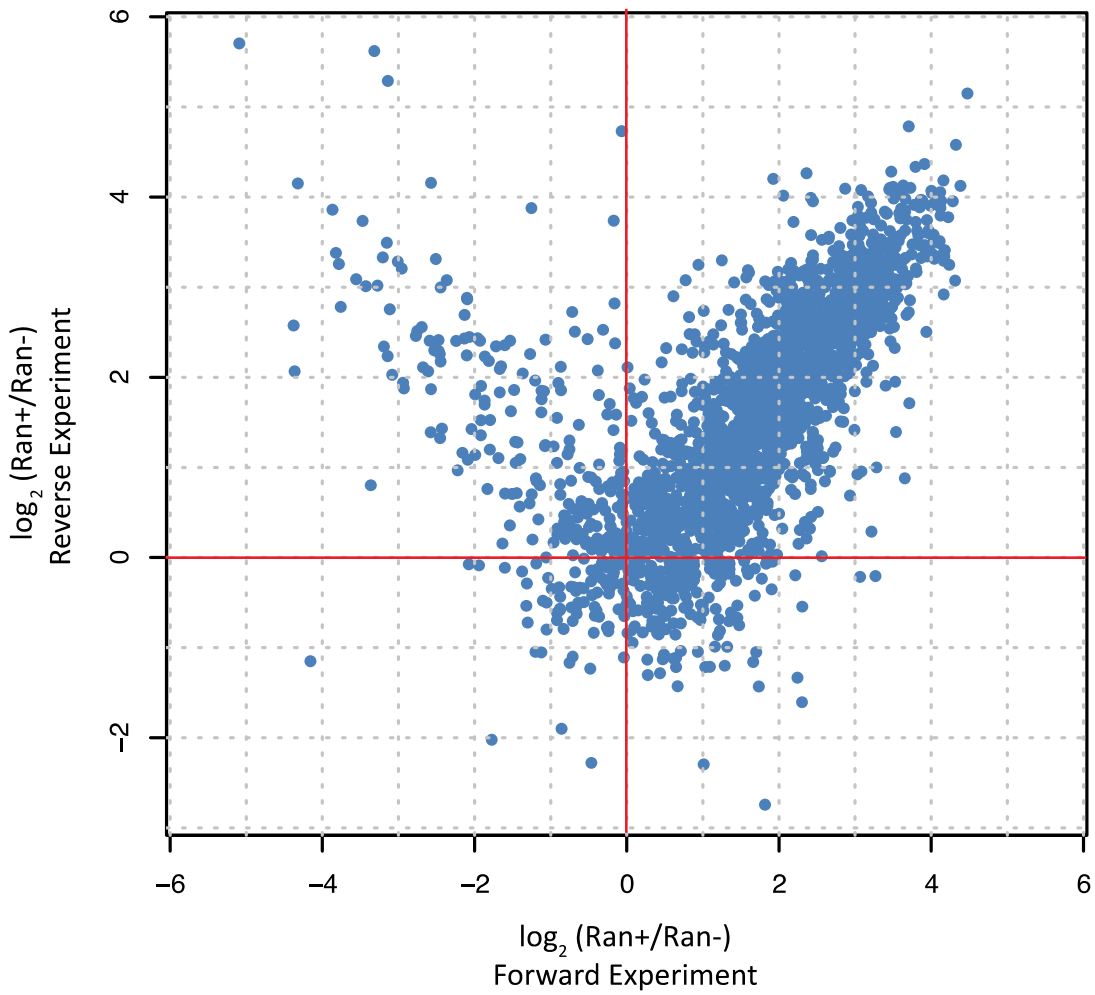

Figure 7-21 SILAC MS analysis of RanGTP dependent CRM1 binders

Ratios of proteins levels obtained by CRM1 affinity chromatography with and without RanGTP were plotted. $\log_2$ of values were used to represent fold changes. Ratios from forward experiment were plotted on x-axis and ratios from reverse experiment were plotted on y-axis. x=0 and y=0 lines were shown in red.

The data was divided with two lines that pass through the 0 values of both experiments. If a protein was bound to the beads both with and without RanGTP, the ratio would be 1, and $\log_2 1$ would yield 0. x values higher than 0 show RanGTP dependent binding in forward experiment, and

y values higher than 0 show RanGTP dependent binding in reverse experiment. The upper left and lower right quartiles show the hits that have different enrichments on mmCRM1 beads in reverse and forward experiments.

The data was divided into further sections. A central circle with formula $x^2+y^2 = 2$ contains most of the data points from quartile 1, 3 and 4. This circle marked the non-specific data points of the analysis that were considered as the background. There were 2 regions that were outside of this circle, one in quartile 1 and one in quartile 2 (Figure 7-22).

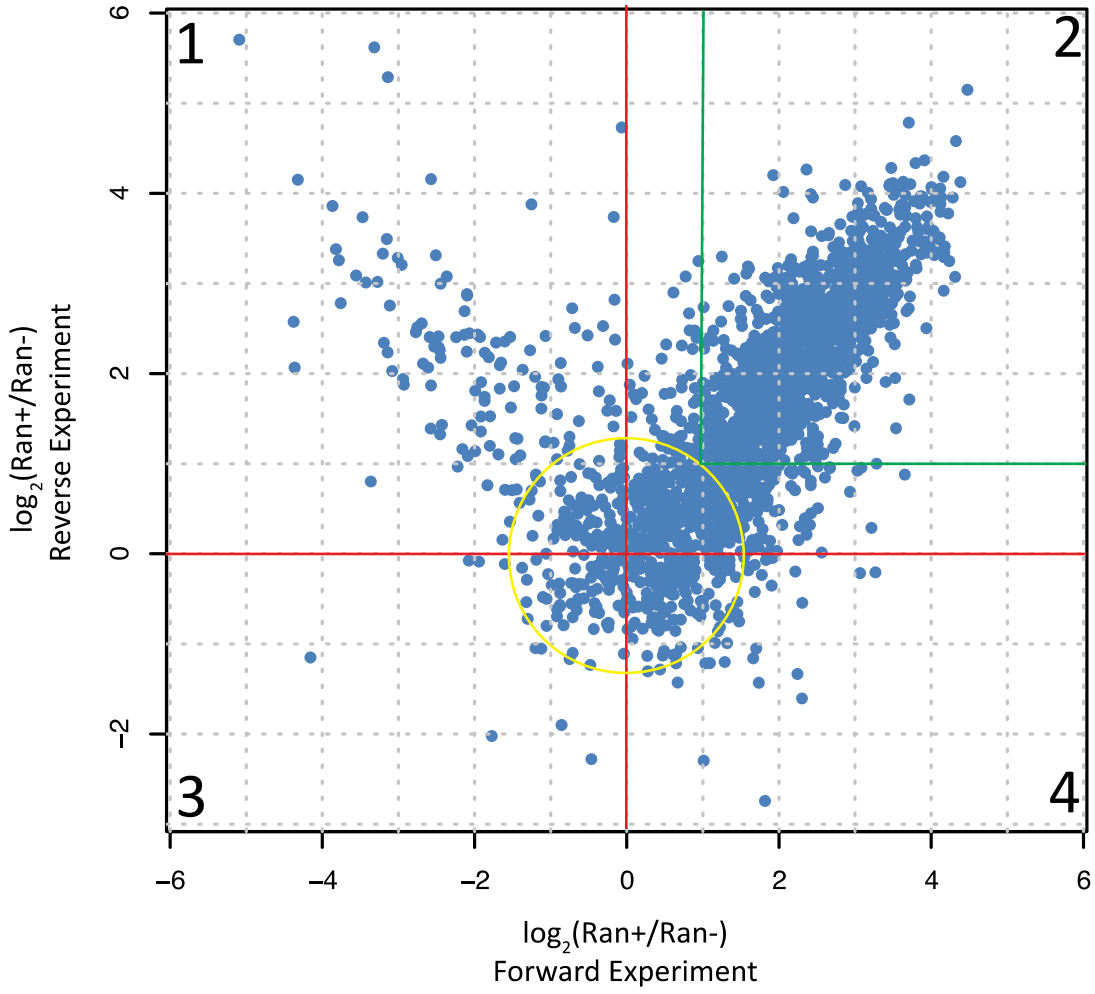

Figure 7-22 SILAC MS data with region markers

Ratios of proteins levels in CRM1 affinity chromatography with and without RanGTP were plotted. $\log_2$ of values were used to represent fold changes. Ratios from forward experiment were plotted on x-axis and ratios from reverse experiment were plotted on y-axis. x=0 and y=0 lines were shown in red. Each quartile divided by these 2 red lines are numbered 1 to 4. A central circle with the formula $x_2+y_2=2$ is drawn in yellow and the region between 2 green lines mark the proteins that show x and y values above 1.

Quartile 1 had data points that were reproduced in both experiments with an inverse correlation. Every sample of mass spectrometry is contaminated with many proteins from environment; a very common example is keratin. Since every contaminant is devoid of heavy labeled amino acids,

they contribute to the light sample. In forward experiment, ratios were calculated for heavy Ran+ sample divided by light Ran- sample. Light contaminants only contributed to the denominator, and since their heavy counterpart was missing, it gave a value less than 1, and $\log_2$ value was negative. For the reverse experiment, ratios were calculated for light Ran+ sample divided by heavy Ran- sample. In this case, light contaminants only contributed to the numerator, and since their heavy counterpart was missing, it gave a value more than 1, and $\log_2$ value was positive. Although most of these contaminants are filtered during data processing, there were still some that contributed. Also the proteins Ran and CRM1 were in this part of the data, because they were bacterially expressed and purified without any heavy amino acids. Another group of proteins were the ones that show inconsistent binding in 2 experiments with regards to RanGTP. This part of data was excluded from the analysis (Figure 7-22).

The more interesting portion of the data resided in 2nd quartile, where proteins were enriched on CRM1 beads in a RanGTP dependent manner in both experiments. Proteins that were at least 2 fold enriched in CRM1 affinity chromatography in the presence of RanGTP over CRM1 affinity chromatography in the absence of RanGTP were taken into consideration. The region on graph was marked with 2 green lines that pass through $\log_2 2$ values on x and y-axis (Figure 7-22). Out of 3070 proteins identified in reverse and forward experiments, 1263 proteins fell in between these two lines. Proteins in this region were not only enriched in Ran+ samples of both experiments. They also gave a very similar result in both experiments, which was visible on the graph since data was scatter along x=y line on quartile 2, x>1 and y>1 region. We considered this region as the promising part of our data.

To compare our data to literature, we used two different sources of previously described NES cargoes. First one was the NESdb; curated database of CRM1 cargoes from various species (Xu *et al.*, 2012a). Second one was a previous SILAC study based on changes in nuclear and cytoplasmic protein pools of HeLa cells upon Leptomycin B treatment (referred as LMB study) (Thakar *et al.*, 2013).

NESdb combines the previously published data for proteins that are exported by CRM1. Two protein lists were compiled from NESdb. First one was composed of human proteins of the NESdb. There were 120 proteins in this first list. Second list was composed of human proteins that had an ortholog from *Canis familiaris* (dog), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Gallus gallus* (chicken), *Xenopus laevis* (African clawed frog) and *Anaxyrus americanus* (American toad), which was described as CRM1 cargoes in the NESdb. The second list contained 54 proteins.

In total there were 174 proteins from NESdb. 59 out these proteins were among proteins identified in our SILAC MS analysis, and 50 of them had $\log_2$ ratios greater than 1 in both forward and reverse experiment (Figure 7-23).

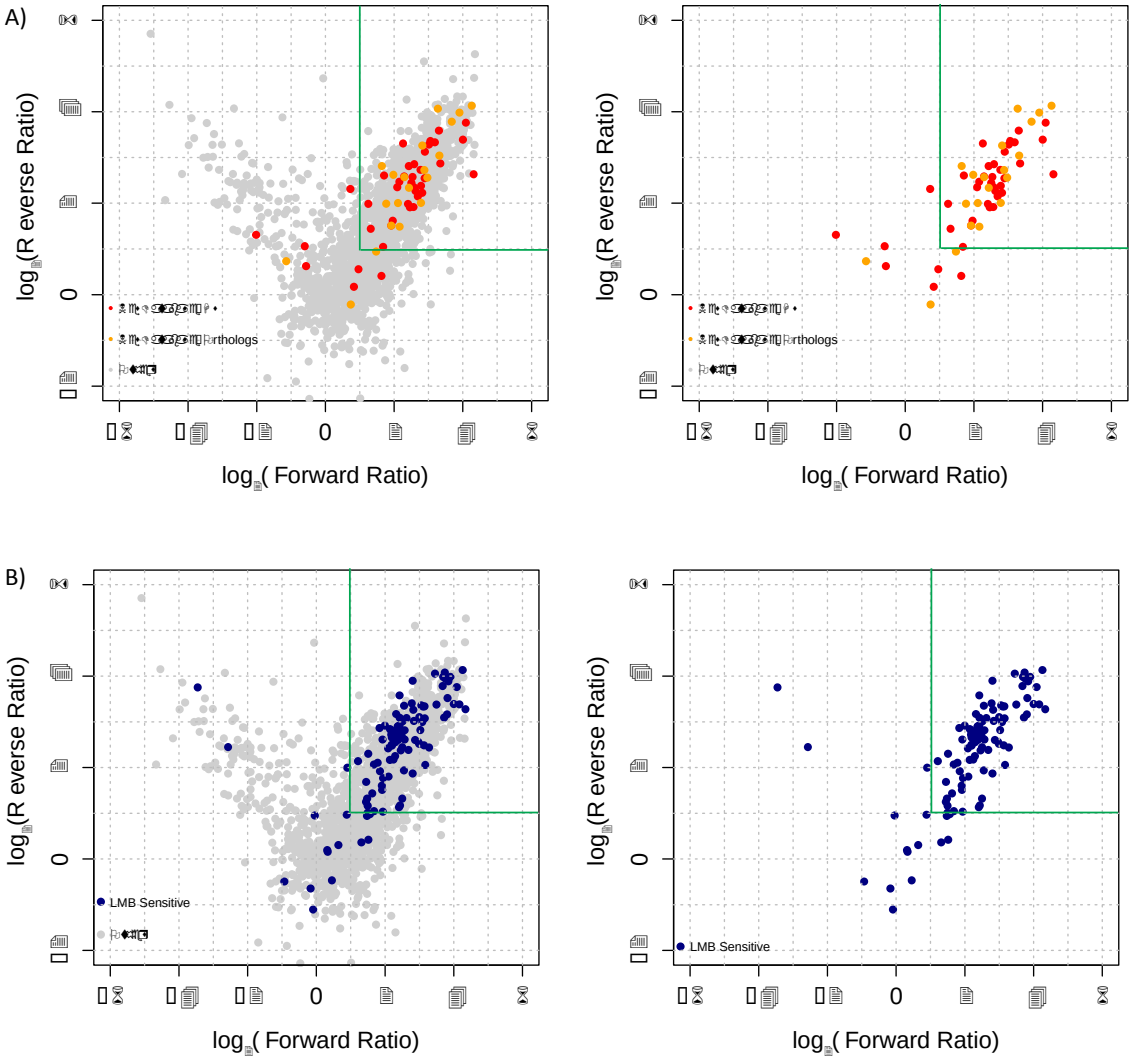

Figure 7-23 NESdb and LMB Sensitive Hits on SILAC MS data

A) Human proteins from NESdb (red) and human ortholog of proteins from NESdb (orange) are marked on our data set.
B) Proteins from LMB were marked blue on our data set.

We also compared our data to the previous LMB study. Leptomycin B treatment specifically inhibits CRM1 export activity by covalently modifying a cysteine residue in the CRM1 hydrophobic pocket. LMB study identified 138 proteins that either showed cytoplasmic depletion or nuclear enrichment upon Leptomycin B treatment. Out of these 138 proteins 104 were in our data set, and 90 of them had a $\log_2$ ratio greater than 1 in both forward and reverse experiments. Both comparisons showed that previously annotated CRM1 cargoes were enriched in our data.

We further analyzed the prominent part of our data for over- and under-represented protein groups. We used KEGG Pathways, KEGG Brite and Gene Ontology databases to compile primary protein groupings (Ashburner *et al.*, 2000; Kanehisa, 2013). We then hand curated these data to come up protein groups, which are over- or under-represented in the specifically CRM1/ RanGTP-bound fraction (Figure 7-24).

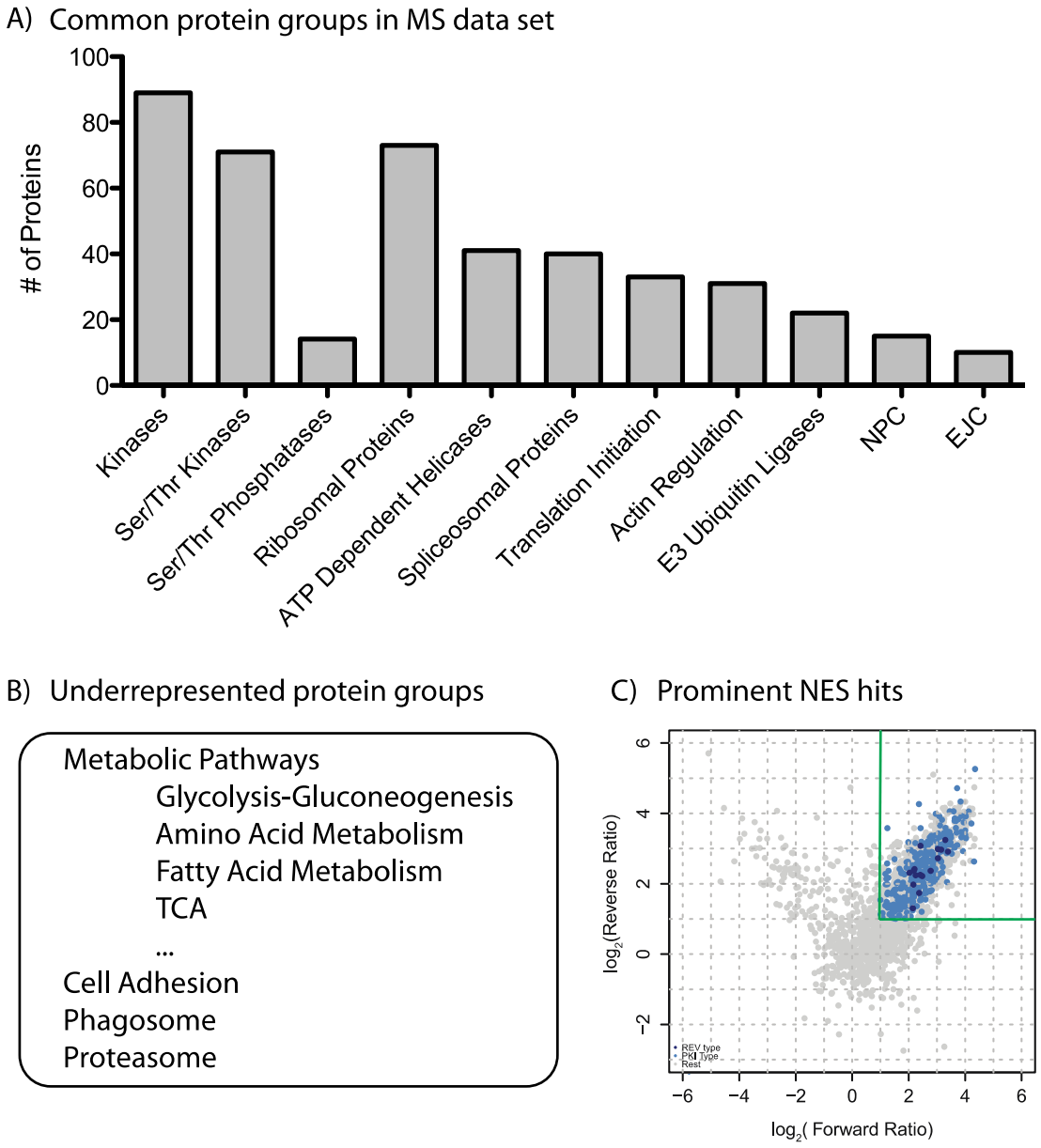

Figure 7-24 Protein groups that are over or under represented in MS data set

A) Number of proteins in MS data from the indicated groups. B) Pathways and activities that are underrepresented in the MS data. C) Prominent PKI-type and REV-type NES predictions in the MS data.

Most of the ribosomal proteins are found in our dataset. Besides them, we find serine threonine kinases, ATP dependent helicases, spliceosomal proteins, translation initiation factors, actin

regulators, and E3 ubiquitin ligases. We do not find proteins of metabolic pathways, cell adhesion, phagosome, and proteasome. The protein groups are not limited by these listed here.

We analyzed the prominent portion of our data for existence of NESs with our prediction algorithms. Out of 1263 proteins, 321 were predicted to have a PKI-type or REV-type NES. 16 of these proteins were predicted to have a REV-type NES with an aggregated disorder propensity higher than 1. 309 proteins were predicted to have a PKI-type NES with an NES score higher than 5000 and an aggregated disorder propensity higher than 1. 4 proteins were predicted to have both types. We found an NES in 25% of the MS data set. To assess its significance, we compared this to NES prediction in a CRM1 binder depleted pool of proteins. To construct this pool, we subtracted all the proteins found in the CRM1 chromatography from the proteins found in the total cytosolic extract. Out of 471 such proteins, 44 were predicted to have an NES with the same constraints. Only 9.3% of the CRM1 binder depleted pool of proteins was predicted to have a significant NES hit.

First of the most complete set of protein groups was the ribosomal proteins. Eukaryotic ribosomes are composed RNA-protein complexes, have 2 main subunits, 60S and 40S. 60S has 46 proteins and 40S has 46 proteins in the core structure (Ben-Shem *et al.*, 2011). 42 proteins of 60S ribosomal subunit, and 31 proteins of 40S ribosomal subunit were present in our data. Since ribosomal subunits are assembled in nucleus, they need to be transported to the cytoplasm. It has been shown that subunits are transported separately in a CRM1 dependent manner. (Thomas and Kutay, 2003). 60S subunit is exported via an adapter protein, NMD3 (Trotta *et al.*, 2003). NMD3 was part of our data set. The CRM1 dependent export mechanism of 40S subunit is not elucidated so far.

A second group of proteins were translation factors. In our data set we found 38 proteins from many initiation factors (Table 7-3). We also found 4 proteins from elongation factors, and 2 proteins from release factors (Table 7-4). We found all core initiation complexes including full members of core initiation complexes eIF2, eIF2B, eIF3. All initiation core complexes had at least one member with a predicted NES hit. eIF2β NES was identified in this study, and it is the member of eIF2 complex that binds to CRM1. We identified the NES on eIF2β, and the other members of eIF2 complex were not predicted to have an NES hit.

Complex	Protein Name	NES Prediction
Core initiation factors		
	eIF1	
	eIF1A	
eIF2	eIF2 alpha	
eIF2	eIF2 beta	*
eIF2	eIF2 gamma	
eIF2B	eIF2B alpha	
eIF2B	eIF2B beta	
eIF2B	eIF2B gamma	
eIF2B	eIF2B delta	
eIF2B	eIF2B epsilon	*
eIF3	eIF3a	*
eIF3	eIF3b	
eIF3	eIF3c	*
eIF3	eIF3d	
eIF3	eIF3e	
eIF3	eIF3f	
eIF3	eIF3g	*
eIF3	eIF3h	
eIF3	eIF3i	
eIF3	eIF3j	
eIF3	eIF3k	
eIF3	eIF3l	
eIF3	eIF3m	
	eIF4A-1	*
	eIF4E	
	eIF4G-1	
	eIF4B	*
	eIF4H	
	eIF5	
	eIF5B	*
Other initiation factors		
	eIF2A	*
	eIF2D	*
	eIF4E type 2	
	eIF6	
	eIF4G-2 (p97)	
	eIF4G-3	
	PABP	
	DHX29	*

Table 7-3 List of translation initiation factors from MS data

Core initiation factors are grouped according to classification by (Trotta *et al.*, 2003). A significant NES hit by the prediction algorithm is indicated by '*'.

Complex	Protein Name	NES Prediction
Elongation factors		
	eIF5A-1	
	eEF1A-1	
	eEF2	
	eEF1D	
Release factors		
	eRF1	
	eRF3A	

Table 7-4 List of translation elongation and release factors

Taken together, we were able to come up with a protein pool that was highly representative of known CRM1 cargoes and also contains many new candidate proteins and protein complexes. Some of these groups make perfect sense for nuclear exclusion, e.g., ribosomal subunits and translation factors, and some that needs further investigation to come up with the biological reasoning for their nuclear exclusion or nucleocytoplasmic shuttling.

8 DISCUSSION

Many routes in the cell have adapted cellular trafficking guided by linear localization sequences. Nucleocytoplasmic trafficking employs such sequences both for import and export. CRM1 is the NTR with the highest workload, responsible for recognition of NESs on many proteins of different functions and families.

8.1 A NEW PREDICTION ALGORITHM FOR CRM1 DEPENDENT NESs

Here we presented a new method of predicting CRM1 dependent nuclear export signals. Although this is not the first prediction algorithm for NESs, it has major differences to the previous NES prediction tools. So far all NES prediction tools focused on the pool of known NES sequences, and tried to come up with a consensus to cover them all. The methods to define this consensus ranged from basic alignments to neural networks. All follow the basic principle of fitting hydrophobic residues L, M, V, F, and I into a consensus that was deduced from the analysis of a NES pool. This approach is limited by a couple of pitfalls.

First, not all NESs found in the literature are true NESs. Since the very first definition of NES was mainly made up of leucine residues (Bogerd *et al.*, 1996), scientific community was biased while analyzing their proteins of interest for the existence of NESs. Many studies analyzed peptide sequences that fit into an NES consensus in an isolated context, performing binding assays with 10-15 residue long peptides. Although these binding assays can function, it is a poor diagnostic tool for assessment of the functional NES on the full-length protein. Since NESs have a high frequency of hydrophobic residues, there is a high chance for a candidate NES that functions in isolated context, to be buried in the protein structure. An important example of such analysis was done for actin. Actin has two NES like sequences on its primary structure, and these two sequences can direct a reporter protein to cytoplasm when fused to it (Wada *et al.*, 1998). However, the crystal structure of actin (PDB ID: 1ATN-A, Kabsch *et al.*, 1990) clearly shows that critical hydrophobic residues of these two "NESs" are deeply buried in the actin structure and thus inaccessible for CRM1 binding. Indeed, it was later shown that full-length actin does not bind to CRM1, but is exported by the dedicated NTR Exportin 6 (Stuven *et al.*, 2003). Likewise, actin is clearly excluded from the CRM1-dependent exportome analyzed in this thesis.

Further NES misannotation originates from sub-optimally designed studies using Leptomycin B (LMB). Leptomycin B fits into the hydrophobic pocket of CRM1 and covalently modifies a cysteine residue, thus blocks NES binding to this pocket (Kudo *et al.*, 1999a). The immediate effect of LMB is blocking of cargo export, but in an experimental setting it is not easy to identify the effect.

When LMB blocks CRM1, expected phenotype of a CRM1 cargo is entry into the nucleus. For small proteins this can be fast since the NPCs do not block passive diffusion of small proteins. Also shuttling proteins that have both NLS and NES accumulate in the nucleus when CRM1 is inhibited. But for some proteins it may take a very long time for them to diffuse into the nucleus, although they possess an NES. These proteins will fail in the LMB test although they are CRM1 cargoes. On the other hand, some proteins will show LMB sensitive localization, although they are not CRM1 cargoes. Prolonged incubations with LMB have secondary effects since CRM1 is responsible for exclusion of RanGAP and RanBP1, which mediate the RanGTP gradient. Also changes in localization of true CRM1 cargoes will have secondary effects, and will change localization of non-CRM1 cargo proteins. Since previous prediction tools first prepared a library of previously published CRM1 dependent NESs, they included false hits, and based their consensus on these mixed pool of true and false NES sequences.

Second, not all sequence features can be attributed to the NES function. The first NES prediction algorithm NetNES (la Cour *et al.*, 2004), trained a hidden Markov model with a true and a false set of NES sequences. At the end they came up with an NES Scoring based on the primary sequence of the query. This approach looks at all the different features with a single constraint. In fact it is possible to extract different features from the NES sequence and grade them separately. One obvious feature is the disorder tendency. For accessibility of the NES by CRM1, it should be kept solvent exposed, unless there are other mechanisms involved in conformational changes of cargoes. A previous NES prediction tool, NESsential (Fu *et al.*, 2011), considered disorder as a NES feature. Although they improved the precision of the NES prediction, there is more to extract from NES and disorder prediction. NESsential focuses on the disorder prediction for residues covering the NES region. N-terminus of PKI-type NESs is α-helical and thus creates a local dip in the disorder tendencies. To point out this feature, we considered disorder propensities of 3 regions, 6 amino acids before the NES hit, the NES hit itself, and 6 amino acids after the NES hit. The local dip in the disorder propensity of linear motifs features has been previously described (Fuxreiter *et al.*, 2007).

The history of NES consensus started with a very limited selection of residues that were allowed for the Φ positions. This initially was the result of the limited number of know NESs, and when the leucine rich NES consensus was defined, this created a bias towards leucine residues in Φ positions (Bogerd *et al.*, 1996). As number of identified cargoes increased over time, this definition also got broader to allow L, V, M, F and I in the Φ positions (la Cour *et al.*, 2004). When known NES sequences were aligned, only the most frequent amino acids made it to a statistical significance. In fact, this was a step where important information was lost due to averaging.

Previously, others groups described supra-physiological NESs, sequences that can bind to CRM1 with a very high affinity, even in the absence of RanGTP (Engelsma *et al.*, 2004). These supra-physiological binding should be achieved by fulfilling all the requirements for an NES extremely well. Features that contribute to NES-CRM1 interaction (e.g. residues in each Φ position, N-terminal α-helical propensity) can be pushed further to strengthen this interaction (Güttler *et al.*, 2010). When PKI NES is modified to have a stronger affinity for CRM1, it failed to dissociate from CRM1 and localized to the nuclear rim. At higher concentrations it even blocked nuclear exclusion of cherry fused to a positive control NES (Figure 7-16). Functional NESs are kept at a sub-optimal CRM1 affinity for a RanGTP regulated CRM1 binding.

Our prediction algorithm expanded the allowed sequences for Φ residues based on a systematic mutation study of these positions (Güttler *et al.*, 2010). Although some of these residues are not optimal for the position, we assumed that optimal residues in other positions could compensate for it. To represent this in our algorithm, we established an incremental scoring system. This allowed us to have a very flexible consensus sequence, but also a high precision with the scoring system. Since the algorithm did not depend on previously discovered NESs, it was not influenced by the false positives in the NES databases.

We observed the immediate outcome of this flexible consensus and scoring system on two different proteins, spRna1p and hseIF2β. Both proteins were known to interact with CRM1 but the responsible NES was not identified. In both cases and unusual alanine residue was in one of the Φ positions, and was overlooked by previous prediction algorithms. Xu et al. defined a new consensus by analyzing their curated NES database NESdb (Xu *et al.*, 2012b). This analysis revealed a broader consensus, but rare events were still excluded by the consensus NES definition. In their analysis alanine was found more frequently in Φ_1 and Φ_2 positions then in Φ_3 and Φ_4 positions. Therefore alanine was restricted to the Φ_1 and Φ_2 residues. That consensus fails to find the NES on spRna1p, which has an alanine in its Φ_3 position. We verified this alanine to be a Φ pocket residue by mutating the position to leucine and enhancing the interaction.

It is evident from previous studies that acidic residues contribute to the CRM1 NES interaction (Güttler *et al.*, 2010). Eukaryotic Linear Motifs Server (ELM) uses an NES consensus that seeks at least one acidic residue both at the N and C-terminus of NES hit (Dinkel *et al.*, 2012). This consensus misses any NES that starts with a Φ_1 position at the extreme N-terminus or ends with a Φ_4 position at extreme C-terminus of the protein sequence. An example would be the spRna1p NES that is at the very end of the protein sequence. Recent crystal structures of CRM1 with bound NES shed light on the contribution of negative residues in N-terminus of the NES. Upstream of the

CRM1 hydrophobic cleft where Φ_0 fits has a positively charged surrounding, and can be involved in electrostatic interactions with negatively charged residues on an NES (Figure 7-3).

One NES that we identified but cannot explain by our NES consensus is the NES from *Xenopus laevis* homolog of eIF2β (Figure 7-10). This NES has a threonine residue in its Φ_3 position. It might well be that threonine is allowed in this sequence context, but not in PKI NES sequence context. This remains to be tested. An alternative explanation could also be possible based on another mutation screen on super PKI, a modified version of PKI with supra-physiological binding affinity (Güttler *et al.*, 2010). Mutants of super PKI, that had one of the Φ positions mutated to alanine, were still able to bind to CRM1. It might mean that when NES features are optimal with 4 Φ positions, the other one can be dispensable. Since the mutation study was done with alanine only, we do not know the allowed sequence space for such a trade off.

Another advancement that came with the NES-CRM1 crystal structures was a new Φ position. This additional Φ position was preceding the already defined $\Phi_{1\text{-}4}$ positions, and the position was named Φ_0. This position is not necessary for the interaction, but can significantly influence the binding strength. To evaluate the effect of the Φ_0 position and surrounding negative charges, we excluded them from the strict consensus, allowing them to be optional, and gave extra score when they were present.

Crystal structure of REV NES with CRM1 revealed an unusual hydrophobic cleft fitting preference that we analyzed separately from the PKI-type NESs. Definition of REV-type included a strict Φ_0, Φ_2, Φ_3 and Φ_4 positions with common hydrophobic residues, and Φ_1, position with a proline residue.

Evaluation of our PKI-type NES prediction algorithm gave very promising results; the combination of NES score with disorder propensity revealed the true NES for the 11 known cargoes (Table 7-1). PKI-type definition covers many previously identified NES sequences and is in accordance with the previous NES consensus definitions. However, evaluation of REV-type NES matched only 7 NESs from NESdb. REV-type definition is more restrictive than the PKI-type and also was previously not recognized. Still, it is very important since it constitutes another class of NESs and cannot be covered by PKI-type definition. The first NES sequence that was described for *Saccharomyces cerevisiae* Rna1p is also REV-type NES (Figure 7-12). REV-type NES class is important for an NES definition with very high, if not complete coverage of all NESs.

8.2 MASS SPECTROMETRY ANALYSIS OF HELA CRM1 CARGOES

Having a powerful NES prediction tool in hand, we wanted to analyze a larger pool of proteins to discover novel NES harboring proteins and also new transport trends in the cell. The number of previously known CRM1 cargoes was already more than 100. To come up with an exhaustive list of proteins, we used a SILAC based quantitative mass spectrometry approach that can compare thousands of proteins from two different pools.

We were not the first ones to employ such a technique. A previous study made use of the CRM1 inhibitor Leptomycin B (LMB) to analyze pool of CRM1 binders (Thakar *et al.*, 2013). They analyzed total pools of nucleus and cytoplasm before and after 3 h of LMB treatment, to come up with proteins that change localization in response to CRM1 inhibition. There are couples of expected outcomes of this experiment. CRM1 cargoes might diffuse into the nucleus in the absence of functional CRM1, and outcome would be nuclear enrichment of such proteins. It is also possible to see cytoplasmic depletion of shuttling proteins with high turnover rate, since they will be trapped in the nucleus in the absence of CRM1, and the cytoplasmic pool would be degraded. In this study Thakar *et al*. identified 84 proteins that show cytoplasmic depletion, and 59 proteins that show nuclear accumulation. 5 proteins were in both groups. Data set contained many ribosomal proteins of 60S ribosomal subunit and 15 previously described cargoes.

This experimental system has some limitations in representing all CRM1 dependent export cargoes. CRM1 is responsible for establishment of RanGTP mediated transport system by keeping RanBP1 and RanGAP cytoplasmic. Upon LMB treatment RanBP1 becomes mainly nuclear in 30 min (Plafker and Macara, 2000a). This alone would create problems not only for CRM1 export but also for all RanGTP dependent export complexes. This effect is evident also in the data set that shows nuclear accumulation of NTRs importin α, importin 4, importin 8, transportin, and CRM1. The scope of other secondary effects of LMB was not addressed in the paper. Another drawback is the limited passive diffusion. Many potential CRM1 target may be part of larger cytosolic complexes, and thus have a very limited passive diffusion in the absence of CRM1. This experimental approach possesses another problem for identification of low abundant proteins. By analyzing whole cytoplasmic and nuclear fractions, the total complexity of the samples is kept very high and this complexity can mask identification of low abundant proteins or minor changes in protein localizations.

We used RanGTP dependent CRM1 affinity chromatography to enrich CRM1 binders from cytoplasmic HeLa extract. This way we were able to confine the protein pool complexity to CRM1 associated proteins. We supplied enough CRM1 molecules to limit competition, and to enrich

even low abundant proteins on our CRM1 streptavidin agarose beads in the presence of RanGTP. One of the key experimental advantages of using streptavidin agarose matrix was the significantly low background binding to HeLa cytoplasmic proteins. This low background made it possible to observe the drastic difference between CRM1 binders in the absence and presence of RanGTP. Previous groups also used such binding assays to identify import or export cargoes. Since the significant changes were limited to number of bands on SDS-PAGE, these bands were cut and analyzed by mass spectrometry (Mingot *et al.*, 2001). Our experimental results gave a higher complexity that wasn't possible to analyze on the level of distinct bands, thus we performed whole lane analysis. To be able to compare eluates of CRM1 affinity chromatography in the presence and absence of RanGTP, we used a SILAC based approach. This allowed us to process these two lanes at the same time and compare them with a single analysis. The binding experiment and the mass spectrometry analysis were repeated with the forward and reverse experiments.

Our experimental setup and analysis may also have some drawbacks. (i) Although we enriched for the RanGTP dependent CRM1 binders, we cannot rule out the possibility of losing very low abundant proteins. (ii) Many human proteins are either not expressed in HeLa cells or kept predominantly nuclear and did not exist in our cytoplasmic extracts. (iii) Some proteins might not yield any ionizable peptide by trypsin digestion. (iv) Some proteins might need modifications for functional NESs. (v) CRM1 chromatography was performed at low salt concentrations, and some proteins with very high CRM1 affinity might bind also in the absence of RanGTP, and fail to enrich. Snurportin 1 was an example of these proteins. Snurportin 1 has very strong CRM1 affinity, was bound also in the absence of RanGTP. (vi) Some NESs might be masked in cytoplasmic complexes.

We identified 1263 proteins that were at least two times enriched in the presence of RanGTP in both reverse and forward experiment. Many of the previously known CRM1 cargoes were part of our data. It also had a quite good coverage of the results from the LMB study.

The most prominent protein group in the data set is the ribosomal proteins with 72 hits. They are synthesized in the cytoplasm and then imported into the nucleus. In nucleoli they assemble with ribosomal RNAs into 40S and 60S ribosomal subunits. Nuclear export of these subunits was shown to depend on CRM1 (Thomas and Kutay, 2003). 60S subunit is exported via an adapter protein, NMD3 (Trotta *et al.*, 2003). We find 42 proteins of 60S ribosomal subunit, 31 proteins of 40S ribosomal subunit, and also NMD3 in our data set. So far no vertebrate adapter for 40S subunit was described. A shuttling protein Ltv1 that binds both Crm1 and 40S was described as the adapter in yeast (Seiser *et al.*, 2006). Our data set contains the human orthologs of this protein,

called protein LTV1 homolog. An NES was identified for yeast Ltv1, but it is highly unlikely to be the true NES, since it has an aspartic acid in Φ_4 position. Our prediction algorithm suggests a very C-terminal NES both for yeast and human proteins.

Another prominent group of proteins is the translation initiation factors. Separation of translation and transcription requires strict confinement of the key regulators, translation factors. Translation initiation factors eIF2, eIF2B, eIF3, eIF4A-1, eIF5 and eIF5B, elongation factors eEF1A, eEF1B, and eEF2, and termination factor eRF1 are kept strictly cytoplasmic (Bohnsack *et al.*, 2002). In our data set we find 38 proteins from many initiation factors, 3 proteins from elongation factors, and 2 proteins from release factors. We found and validated the NES on eIF2β, which can account for the 3 subunits of eIF2 complex in our data set. Another prominent initiation factor is the eIF3 complex. We found all of 13 members in our data set. eIF3G, eIF3C and eIF3A has one prominent NES hits each, and might be responsible for nuclear exclusion of the complex. eIF4A-1, eIF2A,eIF2B epsilon, eIF2D, eIF4B, and eIF5B are the other members of translation initiation complexes with a significant NES hit. We didn't find any significant hits in release and elongation factors. It needs further validation to see which translation initiation complexes are true CRM1 binders.

21 of 71 serine threonine protein kinases in our data set were predicted to have an NES. Two of them, dual specificity mitogen-activated protein kinase kinase 1 (MP2K1) and mitogen-activated protein kinase-activated protein kinase 2 (MAPK2), were previously shown to have NESs (Fukuda *et al.*, 1996; Engel *et al.*, 1998). MAPK2 is localized to the nucleus and upon stress induced phosphorylation; the NES is activated and exported to the cytoplasm. Since serine threonine protein kinases are involved in highly regulated processes like cell proliferation, programmed cell death, cell differentiation and embryonic development, their localization can also be part of their regulation as in the case of MAPK2. This may mean that some serine threonine protein kinases have a regulated NES, which might require other features than we assumed, and cannot be identified with the prediction algorithm. Also the experimental setup may fail to identify the NESs that require further modifications to become functional. Therefore analysis of NESs in regulated proteins requires much more attention.

There are many other groups of proteins that are part of our mass spectrometric data set, and need further classification into meaningful units, such as soluble complexes. Examples of such soluble complexes in our data are signal recognition particle (SRP), the human Augmin complex (HAUS), Ski complex, Arp2/3 complex and minichromosome maintenance protein complex (MCM).

9 OUTLOOK

The new prediction algorithm can come up with a fine selection of NES candidates, based on the Φ position specific scoring and disorder filtering. This prediction algorithm was based on NES crystal structures and mutation screen of PKI NES Φ positions. It is possible that a more complete mutation analysis of both PKI and REV-type NESs might reveal other aspects of the consensus definitions, and enlarge the repertoire of allowed amino acids in Φ positions. Inter-repeat sequences also play a significant role in stability of NES structure. N-terminus of PKI-type NESs has an α-helical structure, and amino acids with higher α-helical propensities may be preferred at this positions. Also effect of neighboring residues is not fully analyzed. It is known that negatively charged residues are preferred, however there might be other constraints on the allowed amino acids. We want to explore these preferences with a comprehensive NES mutation screen.

Our mass spectrometry data contains exhaustive number of proteins. We so far categorized the proteins into functional groups, but a more comprehensive categorization into soluble protein complexes is needed. With such categorization we will start investigation each soluble complex for the CRM1 binding member with our prediction algorithm. This way we can come up with a comprehensive list of NES. Protein groups without a predicted NES hit also constitute an interesting group of proteins; there might be some CRM1 cargoes with unusual NESs, or even different CRM1 interaction features. One example is translation release factor eRF1, which was shown to be Leptomycin B sensitive (Bohnsack *et al.*, 2002) and also was in our MS data set. There is no NES predicted for eRF1, and it remains to be seen what mechanism behind its leptomycin B sensitivity is.

Since our protocol for identification of CRM1 binders from a complex pool was highly efficient, we want to apply these experimental settings to different protein pools. By comparing CRM1 binder pools of synchronized cells from different stages, we might be able to find CRM1 dependent cell cycle regulators. This application can also performed in other model organism systems. Yeast is a very good candidate since there are already established SILAC protocols and proteomic localization studies.

10 MATERIAL AND METHODS

All described standard methods were performed on the basis of (Sambrook and Russell, 2001).

10.1 INSTRUMENTS

Instrument	Manufacturer
Sonifier 450	Branson, UK
Eppendorfbiophotometer	Eppendorf
Incubator/Climo-Shaker ISF1-X	Kuhner Shaker
TCS SP5 microscope	Leica, Mannheim
Thermo NanoDrop2000C	peqLab, Germany
Äkta-Purifier, Äkta-Explorer	Pharmacia, Upsala, Sweden
SensoQuest lab cycler	SensoQuest, Göttingen
miniDAWN™ TREOS®	WyattTechnology, Dernbach
DynaPro NanoStar™	WyattTechnology, Dernbach
1260 Infinity Quaternary LC System	Agilent Technologies, Waldbronn
Shodex ® RI-101	Showa Denko K.K.,Minato-ku, Japan
GenePulser	BioRad, Burlington, USA
Perfection V700 Photo Scanner	Epson
arium® pro UV	sartorius, Gottingen
SB3 rotator	Bibby Scientific, France
RVC 2-18 Rotational Vacuum Concentrator	Christ GmbH, Osterode am Harz

Table 10-1 Laboratory Equipment

Centrifuge	Rotor/Type	Manufacturer
Refrigerated tabletop centrifuge	5415R	Eppendorf
Tabletop centrifuge	5424	Eppendorf
RC6 plus centrifuge	F9, F10	Sorvall
WX Ultra centrifuge	T647.5, T125.0, Type 45 Ti	Sorvall
Discovery M120	S55A, S45A	Sorvall

Table 10-2 Centrifuges and Rotors

10.2 PREPARATION OF DNA CONSTRUCTS

All coding sequences are either acquired from previous lab plasmids, or amplified with specific primers from human cDNA library or yeast genomic DNA.

Coding Sequence	Source
mmCRM1	pTGA021
scCRM1	pSF879
spRna1p	pDG0044
$hsRan_{5\text{-}180}$ Q69L	pKG031
hseIF2β	pTGA404

Table 10-3 Sources of Coding Sequences

For preparing DNA constructs, PCR was performed with designed primers and then the product and the target vector were digested with compatible restriction enzymes with different flanking sequences on 5' and 3' to allow directional insertion. Digested products were checked for correct size on agarose gel electrophoresis, cut from the gel and purified. Vector and insert were ligated and transformed in electrocompetent *E.coli* cells. After preparation, all coding regions were sequenced with primers that anneal before and after the region of interest (Seqlab, Göttingen). All of these standard molecular biology methods were performed on the basis of (Sambrook and Russell, 2001). In the following pages, I tried to explain methods that are modified in detail.

10.2.1 Primer Design

Primers were designed for sub-cloning, introducing deletions or mutations to the DNA constructs. DNASTAR Lasergene SeqBuilder™ software PCR design feature was used for designing primers. When primers were constructed with flanking regions, melting temperature of annealing part was kept above 55°C, and with the flanking region it was kept above 65°C. High secondary structure propensities were avoided. Primers were ordered from Sigma-Aldrich Chemie GmbH (Steinheim, Germany) as desalted oligonucleotides.

10.2.2 Polymerase Chain Reaction (PCR)

PCR was performed for amplification of DNA fragments from templates with desired changes and appropriate restrictions enzyme sites for introducing them into a vector backbone (Mullis *et al.*, 1986) (Hutchison *et al.*, 1978). PCR Enzyme PfuS triple mix components were recombinantly expressed in *E.coli* and purified by Steffen Frey from our lab, diluted to the final mix concentrations in PfuS buffer (200 mM Tris/HCl pH9, 250 mM KCl, 15 mM $MgSO_4$, 100 mM $(NH_4)_2SO_4$, 1% Tween-20, 1 mg/ml BSA).

Protein	Expression vector	Concentration
PfuS	pSF302	100 ng/µl
Pyrococcus abyssi pyrophosphatase	pSF336	15 ng/µl
Pyrococcus abyssi dUTPase	pSF337	2.5 ng/µl

Table 10-4 PfuS Triple Mix Components

PfuS stands for an improved version of thermostable proofreading Pfu polymerase from *Pyrococcus furiosus.* By fusion of Sac7D DNA binding module from *Sulfolobus acidocaldarius*, the enzyme gained a 10-fold increase in processivity (Yang and Wang, 2004).

A typical 100µl PCR reaction was performed with 100 ng of template DNA, 1µl of PfuS triple mix, 2µl DMSO, 8 µl of dNTP-mixture (2.5 mM of each dNTP), 20 µl of 5X Phusion HF Buffer (New England Biolabs, Ipswich, MA, USA), 1 µl reverse and forward primer (100 µM each), and volume was completed with ddH_2O. A SensoQuest lab cycler (Göttingen) was used for PCR reactions. A typical example for a PCR reaction protocol is the following:

Step No	Step Name	Temperature (°C)	Length	Repeat
1	Initial Denaturation	98.5	5’	1
2	Denaturation	98.5	30”	3
3	Primer annealing	55-60	30”	
4	Extension	68	Variable	
5	Denaturation	98.5	30”	30
6	Primer annealing	65-70	30”	
7	Extension	68	Variable	
8	Final extension	68	10’	1

Table 10-5 PCR Reaction Steps

The first three cycles have low annealing temperature to incorporate the flanking regions of primers into the template, and when it is completed, annealing temperature is raised and cycle is continued for 30 rounds. Extension time depends on the length of the PCR product; PCR product length divided by polymerase speed 2kb/min was used as a standard way to determine this time.

10.2.3 Mutagenesis PCR

10.2.3.1 BsaI Mediated Mutagenesis

BsaI is a restriction enzyme that cleaves outside of its recognition sequence and creates recognition sequence independent overhangs. This has been employed in challenging cloning

projects (Engler *et al.*, 2008). As depicted in Figure 10-1, two PCR products were created with two primers each, P1 and P2, P3 and P4. P1 carried a regular restriction site for the integration of the 3' of the construct and P4 carried another one for the integration of the 5' to the target vector. On P2 and P3, BsaI was introduced before and after the site of mutagenesis, with opposite orientation. When both products were cleaved with BsaI, compatible overhangs with mutation were created. It was still highly efficient to ligate two digested PCR products and digested vector backbone in a single ligation reaction.

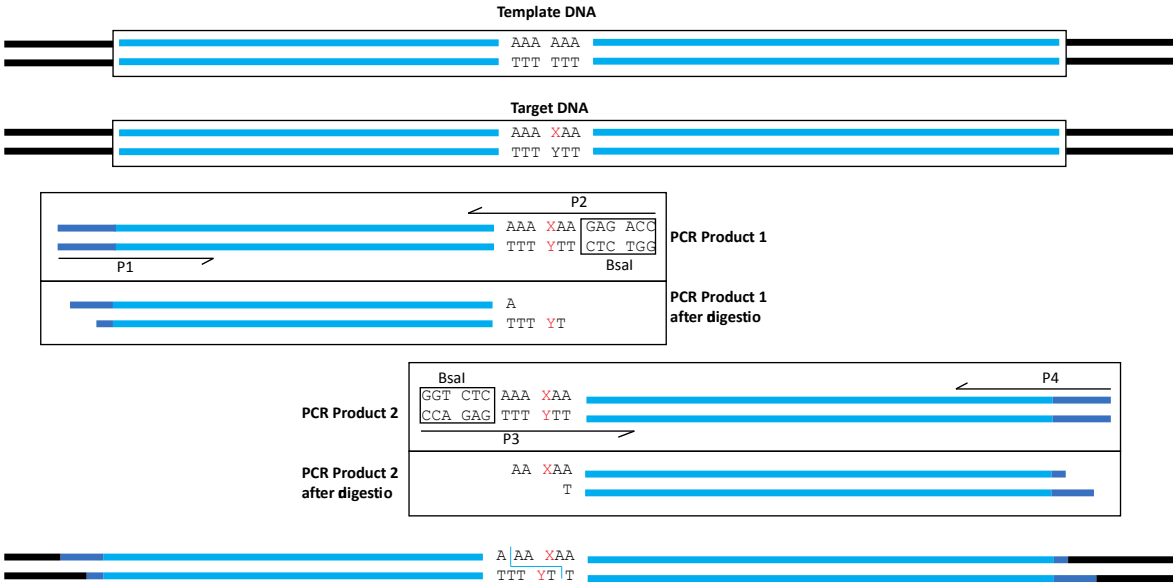

Figure 10-1 BsaI Mediated Mutagenesis

Since most of the existing constructs were already cloned via primers (P1 and P4), only two additional primers (P2 and P3) for each mutagenesis were needed.

When a BsaI site exists in the sequence of product of interest, it can be replaced by BbsI, and when BbsI also exists in the sequence, it is not possible to use this technique. For such cases following method is applied.

10.2.3.2 Blunt End Ligation Mutagenesis

Desired change in the sequence was coded on one of the primers, and two primers were adjacent to each other in the reverse directions, pointing their 3'OH ends away from each other. With these two primers, the whole plasmid can be amplified with one end bearing the mutation. Since the PfuS polymerase leaves blunt ends, ligation of this linear PCR product yields the desired DNA construct. After PCR, 1µl DpnI was added to the PCR reaction to digest the methylated template DNA. The newly synthesized PCR product was not methylated, thus not digested. After DpnI

digestion, PCR product was purified over agarose gel and ligated. Since blunt end ligation is not as efficient as stick end ligation, this method was always the second choice.

10.2.4 DNA Cleavage with Restriction Enzymes

Restriction enzymes were bought from New England Biolabs (NEB, Ipswich, MA, USA) and used as recommended by NEB. When possible, either high fidelity enzymes or a selection of restriction enzymes was preferred (Table 10-6).

Enzyme	Recognition Sequence and Cleavage Position	Enzyme	Recognition Sequence and Cleavage Position
Acc65I	G/GTACC	NheI	G/CTAGC
AgeI	A/CCGGT	SpeI	A/CTAGT
BamHI	G/GATCC	EcoRI	G/AATTC
BspEI	T/CCGGA	EagI	C/GGCCG
HindIII	A/AGCTT	BsaI	GGTCTCN/NNNNN

Table 10-6 Preferred Restriction Enzymes

10.2.5 DNA Gel Electrophoresis

50x TAE
242 g Tris Base
57.1 ml Acetic acid
100 ml 0,5M EDTA pH 8,0
ddH_2O to 1 Litre

DNA-ladder
50 ng/µl 1kb-Ladder (Thermo)
in Orange Sample Buffer (Gibco)

Orange G Sample Buffer
10 mM Tris/HCl pH 8.0
10 mM EDTA pH 8.0
50 % (w/v) Glycerol
25 % (w/v) Orange G

DNA fragments were separated as described in (Sambrook and Russell, 2001) on agarose gels made of 1 % agarose in TAE buffer. To visualize the DNA fragments 0.05 µg/ml ethidium bromide was added to the liquid agarose. DNA samples were combined with 1/10 volume Orange G sample buffer. After the run DNA bands were visualized on a UV Table (Benda Laborgeraete, Wiesloch), and excised.

10.2.6 DNA Extraction From Agarose Gels

For the purification of DNA fragments from excised agarose bands or from a solution Zymoclean Gel DNA recovery kit (Zymo Research, Freiburg) was used according to the manufacturer's instructions.

10.2.7 Determination of DNA Concentration

The concentration of DNA solutions was determined via the extinction at 260 nm (E_{260}), with E_{260} = 1.0 corresponding to 50 µg/ml double-stranded DNA (Sambrook and Russell, 2001). Measurements were done using ND-2000C spectrophotometer.

10.2.8 Ligation of DNA Fragments into Vectors

Vectors were treated with Fast Alkaline Phosphatase (FastAP, Fermentas) for 30 min 37°C in order to remove 5' phosphate groups and preventing the re-ligation of the vector. Digested and purified insert and vector fragments were ligated by T4 DNA ligase (100ng/µl; expressed in our lab by Steffen Frey from vector TB018) in a 10 µl volume at RT for 1 h in 1x ligase buffer (10x ligase buffer: 500mM Tris pH7.5, 100mM $MgCl_2$, 100mM DTT, 10mM ATP, 250µg/ml BSA). 50 ng of vector DNA was incubated without insert (re-ligation control), and with three fold molar excess of insert DNA. 1 µl of the ligation reaction was transformed into electrocompetent *E.coli* cells.

10.2.9 Electroporation of *E. coli* Cells

Electrocompetent cells were prepared by Gabriele Kopp according to the protocol form (Sambrook and Russell, 2001). To reach optimal transformation efficiency, aliquots of frozen electrocompetent *E. coli* cells were thawed slowly and kept on ice until electroporation. 45 µl electro-competent *E. coli* cells and 1 µl of ligation reaction were combined in an electroporation cuvette (165-2086; Bio-Rad, Hercules, CA, USA). Electroporation was performed using the GenePulser (Bio-Rad, Hercules, CA, USA) according to manufacturer's recommendations. Cells were recovered with 2YT medium (16 g Tryptone, 10 g Yeast extract, 5 g NaCl, in 1l ddH_2O) without antibiotics for 1h at 37 °C and 200 µl of cells were plated on LB agar (10 g Tryptone, 5 g Yeast extract, 10 g NaCl, 15 g Agar in 1l ddH_2O) containing the appropriate antibiotics for selection and incubated o/n at 37 °C.

Antibiotics were used with following concentrations for the selection of transformants; 100 µg/ml Ampicillin, 34 µg/ml Chloramphenicol, 25 µg/ml Kanamycin, and 50 µg/ml Spectinomycin.

10.2.10 *E. coli* Strains

BLR(DE3) (69053-3, Novagen), NEB Express I^q (C3037, New England Biolabs) cells were used for protein expression. For CRM1 expressions, BLR cells performed better in terms of protein yield and purity. For other proteins, NEB Express I^q cells were used. NEB10-beta (C3019; New England Biolabs) cells were used for cloning.

10.2.11 DNA Purification From *E. coli* Cultures

Cell cultures were started from single colonies in LB medium (10 g Tryptone, 5 g Yeast extract, 10 g NaCl in 1l ddH_2O) with appropriate antibiotics. Small-scale plasmid DNA preparations ("mini-preps") were started with 8 ml medium; large-scale plasmid DNA preparations ("midi-preps") were started with 250 ml medium. Mini and Midi preps are processed using the two kits NucleoSpin Plasmid and NucleoBond PC100 (both Macherey-Nagel, Düren, Germany).

10.2.12 Bacterial Expression Constructs

Construct ID	Construct Content
pKoKNES005	H10ZZT-spRna1p373_386
pKoKNES006	H10ZZT-spRna1p2-386
pKoKNES007	H10ZZT-spRna1p1-383
pKoKNES069	H14ZZbrSumo-spRna1p2-386
pKoKNES070	H14ZZSumo-spRna1p2-383
pKoKNES072	H14ZZSumo-spRna1pFull A384L
pKoKNES073	H14ZZSumo-spRna1pFull I386S
pKoKNES090	H14ZZSumo-scCrm1
pKoKNES100	H14ZZSumo-mmCrm1
pKoKNES103	H14Sumo-Avi-mmCRM1
pKoKNES114	H10ZZT-eIF2bNES
pKoKNES115	H10ZZT-eIF2bNES O1L
pKoKNES116	H10ZZT-eIF2bNES Xenla
pKoKNES126	H14AviSumo-hseIF2beta
pKoKNES127	H14AviSumo-hseIF2beta I103S
pKoKNES132	H14Sumo-Avi-scCRM1

Table 10-7 Bacterial Expression Constructs

H10 and H14 stand for 10 or 14 histidine residues used as N terminal tag. T stands for TEV site. Avi stands for the Avi tag that is recognized by BirA and covalently modified with biotin.

10.3 PROTEIN EXPRESSION AND PURIFICATION

10.3.1 Native Protein Expression and Purification

All proteins were expressed in appropriate *E. coli* strains (10.2.10). Optimal expression conditions were determined for each protein individually. The following protocol was used as the common method of expression and purification, and when it was not efficient enough, expression and purification conditions were further optimized by using different E.coli strains, N-terminal tags and resuspension buffers.

10.3.1.1 Common Purification Conditions

Frozen cells were thawed in hand warm water. Although freeze-thaw breaks cell walls, it is not enough since the DNA needs to be sheered by sonication for complete solubilization of the expressed proteins. Sonication was performed with Branson Sonifier settings 40% duty cycle and 10 output power in ice bucket for 2 minutes to compensate for the heat produced by the sonicator for 25 ml of resuspended cells. If the volume was larger, sonication was performed with cycles of 2min sonication and 1 min incubation in ice. The lysate was ultracentrifuged for 2h in a T1250 rotor (Table 10-2) at 38,000 rpm to remove cell debris and large aggregates.

We expressed the proteins with an N-terminal histidine tag that consist of 10-14 histidine residues that can be used for affinity purification with Ni^{2+} immobilized beads. By this approach, histidine-tagged proteins can efficiently be purified from complex protein mixtures. The matrix (prepared by Dirk Görlich) was equilibrated with RS1 buffer. The amount of matrix added always depended on the level of protein expression. Ni^{2+}-matrix was incubated with the lysate for 2h at 4°C under rotation in the presence of 15mM imidazole in order to decrease the background binding to Ni^{2+}-matrix from the bacterial lysate. .

The matrix was then let to settle, and after removing the supernatant, matrix was resuspended with RS1 buffer and applied to gravity flow column (volume approx. 10x matrix bed volumes). The resin was washed thoroughly with RS1 buffer containing 25mM imidazole to remove low affinity binders.

10.3.1.2 Elution of Proteins from Ni^{2+} Matrix

Elution of proteins from the Ni^{2+}-matrix depended on the N terminal tag, and the purpose of the purification. If the protein was needed with the tag (e.g., enhancing solubility, immobilization on other matrices) it was eluted with imidazole that competes with histidine residues for Ni^{2+} binding. After the washing step, RS1 buffer containing 0.5 M imidazole was added to the matrix in steps of 1/3 of matrix volume with 2 min incubation, and each step is collected as a different fraction. These fractions were measured for their A_{280} values, and peak fractions were pooled.

In cases where histidine tag was not further needed, proteins were eluted from the resin by digestion of the respective protease cleavage site between the His tag and the protein (e.g. sumo protease digestion for the His – Sumo tagged proteins). After the washing step, buffer in the resin column was quickly exchanged with RS1 buffer containing 5 mM imidazole and 25 nM untagged sumo protease (Expressed and purified by Steffen Frey from plasmid TB005). Resin was incubated for 1h 4°C standing. Elution was done by slowly adding RS1 buffer (1 matrix volume) from the top and collecting the eluate in a single fraction. Elution contained the untagged protein in high

concentration and purity with an insignificant contamination (1:1000 molar ratio) of the used protease.

Sucrose was added to the eluate to a final concentration of 250 mM and the proteins were snap-frozen in liquid nitrogen in aliquots to prevent repeated freeze-thawing, and stored at -80°C. Samples of the uninduced and induced cells, post-sonication and -ultracentrifugation, flow-through of the Ni^{2+}-matrix and the eluted proteins were analyzed on a SDS-PAGE and visualized by Coomassie staining.

10.3.1.3 Expression of Proteins with Biotinytlation

Binding assays on streptavidin-agarose beads require biotinylated proteins as bait. For biotinylation, an N-terminal Avi-tag was included in the expression construct. Avi tag is a 15 amino acid long stretch (GLNDIFEAQKIEWHE), and in the expression construct it is flanked by flexible amino acid stretches. This tag can be recognized by biotin holoenzyme synthetase BirA and a biotin moiety is covalently conjugated to the avi-tag (Beckett *et al.*, 1999). Protein expression vector was co-transformed with BirA expression vector (TB022, prepared by Steffen Frey). At the time of induction with IPTG, also 20 µg/ml biotin (10 mg/ml pH 7.0 stock) is added to the culture. With this method *in vivo* biotinylation was achieved with >99% efficiency.

10.3.2 Determination of Protein Concentrations

Protein concentrations were determined by conversion of the A_{280} value with the calculated coefficients. The A280 value is measured with ND-2000C spectrophotometer that was blanked with RS1 buffer containing 5 mM imidazole and 25 nM untagged sumo protease.

A script written in Python Programming Language with Biopython Package (Cock *et al.*, 2009) was used to extract the protein sequence from Lasergene SeqBuilder™ files (vector maps), check for sumo existence, and calculate the molecular weight and A_{280} absorption coefficient based on the Equation 10-1 where n is the number of indicated amino acid in the sequence.

$$\boldsymbol{\varepsilon = (nW\ x\ 5500) + (nY\ x\ 1490) + (nC\ x\ 125)}$$

Equation 10-1 Absorption coefficient based on amino acid composition (Pace *et al.*, 1995)

Values were calculated both for the full-length and sumo protease cleaved versions. With these values in hand, it is possible to calculate both the mass and molar concentrations.

10.3.3 SDS-PAGE

The method of discontinuous sodiumdodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to standard protocols (Sambrook and Russell, 2001) that provide an up to date version of the original description (Laemmli, 1970). The composition of the gradient

SDS-polyacrylamide gels prepared by Gabriele Kopp and Jürgen Schünemann is described below (Table 10-8). Equipment for the protocol (e.g., Glass plate sets, combs, electrophoresis chambers) had been built by the workshop of the MPI-BPC. Gels were run at 50 mA constant current until the bromophenol blue dye present in the sample buffer reached the bottom of the gel. Subsequently, proteins were fixed and stained by heating the gel in 3% acetic acid and 1:100 dilution of the Coomassie stock solution (2 % (w/v) Coomassie Brilliant Blue G250 in 50 % Ethanol). Gels were destained in water and documented using the EPSON scanner.

	'Heavy' Gel 16%	'Light' Gel 7.5%	Stacking Gel 4.5%
	200 ml	200 ml	100 ml
2M Tris pH 8,8	40 ml	40 ml	-------
0,5 M Tris pH 6,8	-------	--------	15 ml
H_2O	32 ml	107 ml	68 ml
2M Sucrose	10 ml	-------	--------
Glycerol (87%)	8 ml	-------	--------
10 % SDS	2 ml	2 ml	1 ml
Rotiphorese Gel 30	108 ml	51 ml	15 ml
TEMED	120 µl	120 µl	100 µl
APS 10%	2 x 580µl	2 x 580µl	1 ml

Table 10-8 Gradient Gel Solutions

10.4 BINDING ASSAYS WITH CRM1

10.4.1 Binding Assays with Purified Components

Binding assays with purified proteins were performed to test RanGTP dependent CRM1 interaction of candidate cargos or peptides. Different methods were used to pull down the complex. Regardless of the how the complex was pulled down, it was formed as described below.

All proteins were in aliquots and stored at -80°C. Proteins used in the assays were thawed once and was not used again with another freeze thaw cycle for other binding experiments. A 10X stock of binding buffer was prepared and used in all binding assays. Binding buffer was 50 mM Tris/HCl,2 mM $Mg(OAc)_2$ and prepared by 1:10 dilution of the 10X stock with ultrapure water (arium® pro UV, Sartorius, Gottingen), freshly thawed DTT was added to a final concentration of 5 mM, and the buffer was filtered through 0.2 µM filters (Whatman GmbH, Dassel). For binding reaction and other steps of the assay, Mobicols (MoBiTec, Göttingen) were used as the reaction chamber. Mobicols are 700 µl tubes with a conical bottom that can be plugged with filters designed for mobicols (35 µm pore size, MoBiTec, M513515).

Binding reaction was done in 500 µl total volume in Mobicols. CRM1 concentration was 2 µM. Since CRM1 stock was 50 µM with 250 mM NaCl, CRM1 contribution to the final salt was 10 mM NaCl. For samples with RanGTP, 3 µM $Ran_{5\text{-}180}$ Q69L GTP was used from a stock of 50 µM $Ran_{5\text{-}180}$ Q69L GTP with 400 mM NaCl. For each RanGTP sample 30 µl of stock RanGTP was used with a final contribution of 24 mM NaCl. For samples without RanGTP the same volume of RanGTP buffer was added to have same salt contribution. Candidate cargoes and peptides were used at a final concentration of 2 µM. Since purification of different proteins required different buffers and had different final protein yields, their contribution to final buffer conditions were calculated for each binding experiment. Final salt concentration was adjusted with a high salt buffer composed of 50mM NaCl, 400 mM NaCl, 2 mM $Mg(OAc)_2$ 5 mM DTT.

When components were brought together in mobicols with a bottom plug and a screw cap, they were incubated in cold room at 4°C on SB3 rotator (Bibby Scientific, France) with a speed of 10 rpm for 2 hours. After 2 hours the respective affinity matrix was added.

10.4.1.1 Binding Assays with ZZ-affibody Beads

When candidate cargoes are expressed with an N-terminal ZZ tag, this tag can be used to bind them to ZZ affibody. Z in ZZ stands for the synthetic IgG Fc region binding domain from the B domain of the *Staphylococcus aureus* protein A. Affibody molecules are small proteins, engineered for specific protein interactions. An affibody made for Z domain binding (Wahlberg *et*

al., 2003) was expressed and immobilized to functionalized silica beads (prepared by Dirk Görlich) by Steffen Frey. 5 mg of ZZ affibody was immobilized on 1ml of functionalized silica beads. For each binding assay 20 µl ZZ-affibody beads were used, and this amount was enough to pull the 1 nmol of ZZ tagged candidate protein or peptide. When tested with ZZ-PKI NES peptide, all immobilized protein was competent in RanGTP dependent CRM1 binding.

Stocks of the ZZ-affibody beads were kept in 4.1 M ammonium sulfate at 4°C. 10% excess of the required amount of the beads were removed with pipette using a cut pipette tip, to not to harm the beads. Beads were placed in a mobicol, and washed 5 times with 500 µl binding buffer. After addition of buffer, mobicol was placed in a 2 ml Eppendorf tube and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge for 30 sec. At this low speed of centrifugation the sepharose-beads remained intact.

Beads were resuspended in 1:1 volume of binding buffer and 40 µl of suspension was pipetted in the binding reaction. Binding reaction again was incubated in cold room at 4°C for 1 hour. After incubation, mobicols were unplugged and placed in 2 ml Eppendorf tubes, and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge. Flow-through was collected and beads were washed with 500 µl binding buffer by 1000 rpm centrifugation for 30 sec at 4°C 2 times. To get rid of the buffer that remained in the bead volume, a very short (5-10 sec) centrifugation at 3000 rpm was performed. Mobicols were placed in 1.5 ml Eppendorf and 50 µl SDS sample buffer was added. Mobicols were kept at room temperature for 5 min and centrifuged at 1000 rpm for 1 min at room temperature tabletop centrifuge. Another 50 µl SDS sample buffer was added and the mobicols were centrifuged at 1000 rpm for 1 min at room temperature tabletop centrifuge. At the end 100 µl elution was collected in a single 1.5 ml Eppendorf, and 10 µl of each elution was analyzed on SDS-PAGE.

10.4.1.2 Binding Assays with Streptavidin-agarose Beads

For some binding assays either the candidate cargo or the CRM1 protein was biotinylated. Biotinylation was performed in vivo as explained in section 10.3.1.3. Biotin has a very high affinity for streptavidin homo tetramers (K_d in the order of 10^{-14}). For pull downs of biotinylated proteins streptavidin-agarose beads (Sigma Aldrich) were used. 10% excess of the required amount of the beads was removed with pipette using a cut pipette tip, in order not to harm the beads. Beads were placed in a mobicol, and washed 5 times with 500 µl binding buffer. After addition of buffer, mobicol was placed in a 2 ml Eppendorf tube and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge for 30 sec. At this low speed of centrifugation the sepharose-beads remained intact.

Streptavidin agarose beads were resuspended in 1:1 volume of binding buffer and 40 µl of suspension was pipetted in the binding reaction. Binding reaction was incubated in cold room at 4°C for 1 hour. After incubation, mobicols were unplugged and placed in 2 ml Eppendorf tubes, and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge. Flow-through was collected and beads were washed with 500 µl binding buffer by 1000 rpm centrifugation for 30 sec at 4°C 2 times. To get rid of the buffer that remained in the bead volume, a very short (5-10 sec) centrifugation at 2000 rpm was performed. Mobicols were placed in 1.5 ml Eppendorf tubes and 50 µl SDS sample buffer was added. Tubes were placed on thermo shaker at 37°C, and after an initial 1050 rpm shake for 5 sec, they were incubated for 5 min at 350 rpm shaking. If the initial high speed shaking was not done, bead would fail to mix with the SDS sample buffer and settle at the bottom. Mobicols in 1.5 ml Eppendorf tubes were centrifuged at 1000 rpm for 1 min at room temperature tabletop centrifuge. SDS sample buffer was kept at 37°C, and another 50 µl SDS sample buffer was added and centrifuged at 1000 rpm for 1 min at room temperature tabletop centrifuge. At the end 100 µl elution was collected, and 10 µl of each elution was analyzed on SDS-PAGE.

Streptavidin-biotin interaction is very strong and needs denaturing conditions to fully dissociate biotin from streptavidin. With this elution method, all of the prey material was retrieved, but only a partial dissociation of biotin-cargo or biotin-CRM1 was possible. To check immobilized protein levels and control if any prey was left on the beads, 20 µl streptavidin agarose beads were boiled in 100 µl SDS sample buffer and 10 µl was analyzed on SDS-PAGE.

10.5 PULL DOWN FROM CYTOPLASMIC EXTRACT WITH CRM1

10.5.1 Preparation of Cytoplasmic Extracts

Cytoplasmic HeLa extracts were kindly provided by Lührmann Lab, Department of Cellular Biochemistry, MPI-BPC, and prepared with the following protocol modified from (Wahlberg *et al.*, 2003). Monolayer of HeLa cells at 80% confluence was harvested by trypsinization. Cells are pelleted at 2000rpm for 5 min and washed 3 times with phosphate-buffered Saline (PBS: 130 mM NaCl, 20 mM K_2HPO_4/KH_2PO_4, pH 7.4) for 10 min each. Cells were pelleted again and cell pellet was weighed. Packed cell volume (ml) was calculated by multiplying cell weight in gr by 0.96. Total cell number was calculated by multiplying packed cell volume by 0.03 x 10^{10}. Cells were resuspended in 1.25 times the volume of packed cell volume of MC buffer (10 mM HEPES/KOH pH 7.6, 10 mM KOAc, 0.5 mM MgOAc, 5 mM DTT, 1x complete EDTA free proteinase inhibitor). Suspension was incubated 5 min on ice at 4°C and dounced 18 times using cell homogenizer.

Homogenate was pelleted in Corex tubes at 13000 rcf in SS34 rotor for 5 min. Supernatant was taken and frozen in liquid nitrogen, stored at -80°C.

This extracts were subjected to ultracentrifugation at 42000 rpm in 55A rotor at 4°C for 1h. Supernatant was collected, aliquoted, frozen in liquid nitrogen, and stored at -80°C.

10.5.2 Preparation of Cytoplasmic SILAC HeLa Extracts

Cytoplasmic SILAC HeLa extract were kindly supplied by Miroslav Nikolov (Mass Spectrometry Research Group, MPI-BPC). HeLa S3 cells were grown in lysine- and arginine-deficient Dulbecco's modified Eagle's medium supplemented with 10% dialyzed fetal bovine serum (PAA, Pasching, Austria). One cell population was supplemented with natural L-lysine and L-arginine (Sigma, Munich, Germany) and another with heavy isotope labeled 4,4,5,5-d_4-L-lysine and $^{13}C_6$-L-arginine (Euriso-Top, Saint-Aubin Cedex, France) generating mass shifts of 4 and 6 Da, respectively. Cells were grown for at least six passages at smaller volumes and then expanded to 2 l in spinner flasks (0.5–1.0 x 10^6 cells/ml) (Ong and Mann, 2006). The cells were then transferred to a 5 l fermenter (Applikon, Schiedam, Netherlands) and grown under standard conditions (2.5–5.0 x 10^6 cells/ml). Harvested cells were used to prepare cell extracts. Cells are pelleted at 2000rpm for 5 min and washed 3 times with phosphate-buffered Saline (PBS: 130 mM NaCl, 20 mM K_2HPO_4/KH_2PO_4, pH 7.4) for 10 min each. Cells were pelleted again and cell pellet was weighted. Packed cell volume (ml) was calculated by multiplying cell weight in gr by 0.96. Total Cell number was calculated by multiplying packed cell volume by 0.03 x 10^{10}. Cells were resuspended in 1.25 times the volume of packed cell volume of MC buffer (10 mM HEPES/KOH pH 7.6, 10 mM KOAc, 0.5 mM MgOAc, 5 mM DTT, 1x complete EDTA free proteinase inhibitor). Suspension was incubated 5 min on ice at 4°C and dounced 18 times using cell homogenizer. Homogenate was pelleted in Corex tubes at 13000 rcf in SS34 rotor for 5 min. Supernatant was taken and frozen in liquid nitrogen, stored at -80°C.

This extracts were subjected to ultracentrifugation at 42000 rpm in S55A rotor at 4°C for 1h. Supernatant was collected, aliquoted, frozen in liquid nitrogen, and stored at -80°C.

10.5.3 CRM1 Affinity Chromatography with Cytoplasmic HeLa Extracts

CRM1 Affinity Chromatography with Cytoplasmic HeLa Extracts was done to enrich the RanGTP dependent CRM1 binders of cytoplasmic HeLa proteins. This method was first used with cytoplasmic HeLa extract from Lührmann Lab, Department of Cellular Biochemistry, MPI-BPC to test and optimize the assay conditions. For the mass spectrometry analysis, we used this method with the cytoplasmic SILAC HeLa extracts from Miroslav Nikolov (Mass Spectrometry Research Group, MPI-BPC).

All proteins and extracts were prepared as previously described and stored in aliquots -80°C. Proteins and extracts used in the assays were thawed once and was not used again with another freeze thaw cycle for other assay. A 10X stock of binding buffer was prepared and used in all binding assays. Binding buffer was 50 mM Tris/HCl,2 mM $Mg(OAc)_2$ and prepared by 1:10 dilution of the 10X stock with ultrapure water (arium® pro UV, sartorius, Gottingen), freshly thawed DTT was added to a final concentration of 5 mM, and the buffer was filtered through 0.2 µM filters (Whatman GmbH, Dassel). For binding reaction and other steps of the assay, Mobicols (MoBiTec, Göttingen) were used as the reaction chamber.

CRM1 immobilization was done on streptavidin-agarose beads. For each reaction 20 µl of streptavidin-agarose beads were used. 10% excess of the total required amount of beads was taken with pipette using a cut pipette tip. Beads were placed in a mobicol, and washed 5 times with 500 µl binding buffer. After addition of buffer, mobicol was placed in a 2 ml Eppendorf tube and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge for 30 sec. At this low speed of centrifugation the sepharose-beads remained intact.

For each reaction 0.5 nmol of biotin-CRM1 was immobilized on 20 µl of streptavidin-agarose beads. Total calculated amount of biotinylated CRM1 was added to mobicol with the beads, volume was completed to 500 µl with a buffer containing 50 mM Tris/HCl, 100 mM NaCl, 2 mM $Mg(OAc)_2$, 5 mM DTT. For complete immobilization of biotinylated CRM1 on streptavidin-agarose beads, mobicol was kept in cold room for 1 hour on SB3 rotator (Bibby Scientific, France) at 10rpm. After incubation, beads in mobicols were washed 5 times with 500 µl of buffer containing 50 mM Tris/HCl, 25 mM NaCl, 50 µM biotin, 2 mM $Mg(OAc)_2$, 5 mM DTT. After the addition of buffer, mobicols were kept for 1 min on ice and, placed in a 2 ml Eppendorf tube and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge for 30 sec. Beads were resuspended 1:1 volume of binding buffer and 40 µl of suspension was pipetted in a new mobicol for each reaction. Aliquots of light and heavy extracts were thawed on ice, and centrifuged for 15min at 4°C in S45A rotor at 37000 rpm.

CRM1 affinity chromatography was done in 500 µl total volume in mobicols, with 20 µl of CRM1 immobilized streptavidin agarose beads. For reactions with RanGTP, 2 µM $Ran_{5\text{-}180}$ Q69L GTP was used from a stock of 50 µM $Ran_{5\text{-}180}$ Q69L GTP with 400 mM NaCl. For each RanGTP sample 20 µl of stock RanGTP was used with a final contribution of 20 mM NaCl. For samples without RanGTP the same volume of RanGTP buffer was added to have same salt contribution. 100 µl of centrifuged extract was added, final volume was brought to 500 µl with binding buffer and final salt concentration was adjusted to 25mM NaCl.

When components were brought together in mobicols with a bottom plug and a screw cap, they were incubated in cold room at 4°C on SB3 rotator at 10 rpm for 3 hours. After incubation, mobicols were unplugged and placed in 2 ml Eppendorf tubes, and centrifuged at 1000 rpm at 4°C in a refrigerated table top centrifuge. Flow-through was collected and beads were washed with 500 µl binding buffer with 25mM NaCl by 1000 rpm centrifugation for 30 sec at 4°C 2 times. To get rid of the buffer that remained in the bead volume, a very short (5-10 sec) centrifugation at 2000 rpm was performed. Mobicols were placed in 1.5 ml Eppendorf tubes and 50 µl SDS sample buffer was added. Eppendorf tubes were placed on thermo shaker at 37°C, and after an initial 1050rpm shake for 5 sec, they were incubated for 5 min at 350 rpm shaking. Mobicols in 1.5 ml Eppendorf tubes were centrifuged at 1000 rpm for 1 min at room temperature tabletop centrifuge. SDS sample buffer was kept at 37°C, and another 50 µl SDS sample buffer was added and centrifuged at 1000 rpm for 1 min at room temperature tabletop centrifuge. At the end 100 µl elution was collected, and 10 µl of each elution was analyzed on SDS-PAGE.

10.5.4 Mass Spectrometry Analysis of Elution Fractions

For forward experiment, 20ul of elutions from CRM1 beads with light HeLa extract without RanGTP, and heavy HeLa extract with RanGTP were put together. For reverse experiment, 20 µl of elutions from CRM1 beads with heavy HeLa extract without RanGTP, and light HeLa extract with RanGTP were put together. Both combined elutions were concentrated to 20 µl on rotational vacuum concentrator. Also 1:1 mix of input light and heavy extracts dissolved in SDS-sample buffer. Concentrated samples of forward and reverse experiment, and 1:1 mix of extracts were separated on 4–12% gradient SDS-PAGE gels (NuPAGE, Invitrogen, Carlsbad, CA) and stained with Colloidal Coomassie Blue (0.08 % (w/v) Coomassie Brilliant Blue G250, 1.6 % (v/v) ortho-phosphoric acid, 8 % (w/v) Ammonium sulphate, 20 % (v/v) Methanol) (Neuhoff *et al.*, 1988). Each gel lane was cut into 12 equal gel slices and proteins therein were in-gel digested with trypsin (Promega, Madison,WI) as described in (Neuhoff *et al.*, 1988).

The rest of the protocol was performed by Samir Karaca (Mass Spectrometry Research Group, MPI-BPC). Extracted peptides were loaded into an in-house packed C18 trap column (1.5 cm, 360 µm outer diameter, 150 µm inner diameter, Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) at a flow rate of 10 µl/min. Retained peptides were eluted and separated on an analytical C18 capillary column (15 cm, 360 µm outer diameter, 75 µm inner diameter, Reprosil-Pur 120 Å, 5 µm, C18-AQ, Dr. Maisch GmbH, Germany) at a flow rate of 300 nl/min with a gradient from 5% to 38% acetonitrile in 0.1% formic acid for 50 min using an Agilent 1100 nano-flow LC system (Agilent Technologies, Santa Clara, CA) coupled to an LTQ-Orbitrap Velos hybrid mass spectrometer (Thermo Electron, Bremen, Germany). The LTQ-Orbitrap Velos

was operated in data-dependent mode and survey scans were acquired in the Or- bitrap (m/z 350–1600) with a resolution of 30,000 at *m/z* 400 with a target value of 1 x e106. Up to 15 of the most intense ions with charges ≥ +2 from the survey scan were sequentially isolated for collision-induced dissociation with normalized collision energy of 37. Dynamic exclusion was set to 60 s to avoid repeating the sequencing of peptides. Each sample was analyzed in two technical replicates.

10.5.5 Data and Bioinformatics Analysis

Raw MS files from the LTQ- Orbitrap Velos were analyzed using MaxQuant software (version 1.3.0.5)(35) with Andromeda search engine. Peak lists generated by MaxQuant software were searched against the Uniprot Human protein database (downloaded on 10 July 2013, containing 88,354 entries) supplemented with common contaminants (e.g. keratins, serum albumin) and concatenated with the reverse sequences of all entries. MaxQuant search parameters were as follows: carbamidomethylation of cysteine was set as a fixed modification, whereas oxidation of methionine and N-terminal protein acetylation were set as variable modifications; tryptic specificity with no proline restriction and up to two missed cleavages was used. The MS survey scan mass tolerance was 6 ppm and for MS/MS 0.5 Da. Only peptides with a minimal length of five amino acids were considered for identification. The false discovery rate was set to 1% at both the peptide and the protein level. “Re-quantify ” was enabled, and “keep low scoring versions of identified peptides ” was disabled. Quantification of SILAC pairs was performed with a minimum ratio count of two by considering unique and razor peptides. To generate results with a high confidence interval, two biological replicates were performed, and each biological replicate was analyzed twice. To avoid false positives due to the experimental workflow, label-swap experiments were performed. Proteins behaving adversely in forward and reverse labeling experiments were excluded from the analysis.

Tab-delimited text file output from MaxQuant (proteinGroups.txt) was imported in R statistical Environment without pre-processing. All “Reverse” and “Contaminant” entries were excluded from further analysis. Non-normalized enrichment ratios in both label-swap experiments are represented in log2 scale.

The enrichment analysis for GO MF, BP and CC (Ashburner *et al.*, 2000) were done for significant list with respect to HeLa proteome (Nagaraj *et al.*, 2011) by the “conditional hypergeometric test” available in the GOstats package (Falcon and Gentleman, 2007) in the R statistical environment (R Development Team, 2012). KEGG pathway (Kanehisa and Goto, 2000) enrichment analysis was done in the same way, except that the hypergeometric test was used and the reference set was complete human KEGG annotations.

10.6 TRANSIENT HELA CELL TRANSFECTIONS

GFP fusions of spRNA1 and PKI versions were prepared with a modified pEGFP-C1 (Invitrogen, Carlsbad, CA) vector. Each construct was co-transfected with vector coding for mCherry fusion of eIF2β65-114, that has an NLS and NES. This construct from Chandini Kadian was used as the positive control.

Construct ID	Construct Content
pKoKeu004	eGFP-superPKI NES 3
pKoKeu005	eGFP-PKI NES
pKoKeu006	eGFP-PKINESp4A
pKoKeu008	eGFP-spRna1p
pKoKeu011	eGFP-spRna1p2-376
pKoKeu035	eGFP-SV40NLS-spRna1p
pKoKeu038	eGFP-SV40NLS-spRna1p2-376
pCK118	mCherry-eIF2beta65-114

Table 10-9 Eukaryotic transfection constructs

1x 10^4 HeLa cells were plated on each coverslip in a 24 well plate (1.88 cm^2 growth area/ coverslip). Next day, when cells reach a confluency of 50-80%, transfection was performed. Transfection was done with 0.1 µg of each construct with FuGENE6 (Promega, Madison,WI) reagent. Transfections were done by Heinz-Jürgen Dehne with the provided guidelines from the producer. After 24 h, cells on coverslips were washed 2 times with 3 ml of PBS, and cells on coverslips were fixed with 5 min incubation in 4% paraformaldehyde. Excess paraformaldehyde was removed with another PBS wash including DAPI. 2.5 µl Vectashield (Vector Laboratories, CA, USA) was placed on glass slides, and coverslips were placed on the glass slide with cells facing the Vectashield drop. This way, the cells were in between the coverslips and the glass slides. Confocal microscopy images were recorded using a Leica TCS SP5 Laser scanning microscope equipped with a HC PL APO 20x glycerol objective (Leica GmbH, Mannheim).

11 ABBREVIATIONS

A280	Absorbance at λ = 280 nm
ADP	Adenosine 5'-diphosphate
ATP	Adenosine 5'-triphosphate
C-terminus	Carboxy-terminus
cDNA	Complementary deoxyribonucleic acid
CRM1	Chromosomal region maintenance 1 (Exportin 1/ Xpo1p)
ct	*Chaetomium thermophilum*
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
eIF	Eukaryotic translation initiation factor
Exp	Exportin
FG repeat	Phenylalanine-glycine repeat
GAP	GTPase-activating protein
GDP	Guanosine 5'-diphosphate
GFP	Green fluorescent protein
GTP	Guanosine 5'-triphosphate
GTPase	GTP hydrolase
HEAT repeat	Class of protein repeats (Huntingtin, Elongation factor 3, Protein phosphatase 2A, TOR1)
HIV	Human immunodeficiency virus
hnRNP	Heterogeneous nuclear ribonucleoprotein
hs	*Homo sapiens*
Imp	Importin
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilo Dalton
LB	Luria-Bertani (lysogeny broth, medium)
LMB	Leptomycin B
mCherry	Monomeric Cherry (a red-fluorescent protein)
MDa	Mega Dalton
mm	*Mus musculus*
mRFP	Monomeric red fluorescent protein
N-terminus	Amino-terminus (start of a protein)
NE	Nuclear envelope
NES	"Leucine-rich" nuclear export signal
NLS	Nuclear localization signal
NPC	Nuclear pore complex
NTF2	Nuclear transport factor 2
NTR	Nuclear transport receptor
Nup	Nucleoporin (NPC protein)
OD_{600}	Optical density λ = 600 nm
PBS	Phosphate-buffered saline
PKA	Protein kinase A (cAMP-dependent protein kinase)

PKI	Protein kinase A inhibitor
PMSF	Phenylmethylsulfonyl fluoride
Ran	Ras-related nuclear antigen
RanBP	Ran-binding protein
RanGAP	RanGTPase-activating protein
RanGEF	Ran guanine nucleotide exchange factor
Ras	Rat sarcoma
RCC1	Regulator of chromosome condensation 1 (see also "RanGEF")
rpm	Rounds per minute
sc	*Schizosaccharomyces pombe*
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sp	*Saccharomyces cerevisiae*
SPN1	Snurportin 1
SV40	Simian virus 40
TEV	Tobacco etch virus
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
z	(in "zz") IgG-binding domain of the Staphylococcal protein A

12 APPENDIX

12.1 PYTHON SCRIPTS

12.1.1 PKI-type NES Prediction

```
#Import used Libraries
import os
from Bio import SeqIO
from Bio import Seq

input='/users/koray/desktop/NESprediction/Trial.fasta'   #input fasta file
#Parse Input File into Sequences
handle = open(input)
for record in SeqIO.parse(handle,'fasta'):
    #Put single sequence into a temp file
    liste=[]
    liste.append(record)
    handle2=open('temp.fasta','w')
    SeqIO.write(liste,handle2,'fasta')
    handle2.close()

#Get Disorder For a Single Protein
    p = Popen('./iupred temp.fasta short', shell=True, stdin=PIPE, stdout=PIPE, stderr=PIPE, close_fds=True)
    (stdin, stdout, stderr) =(p.stdin, p.stdout, p.stderr)
    results = stdout.readlines()
    x=0
    ListDis=[]
    for i in results:
        if i[0] =='#':
            continue
        a,b = i.split('    ')
        b,c=b.split('\n')
        x=x+1
        ListDis.append((x,b))
    stdout.close()

#Get Domains from SMART Database
    p = Popen('perl Smart_batch.pl --inputFile temp.fasta --outputDirectory .../NESprediction/ --includePfam',
shell=True, stdin=None, stdout=None, stderr=None)
    p.wait()
    handle3= open('tempDomain.txt','r')
    domains= handle3.read()
    handle3.close()
    domainList = domains.split('\n\n')

    #process the result file, remove the tags
    del domainList[0:2]
    del domainList[-1]
    ListFeatures=[]

    #put values in to list format with filters for low complexity and coiled coils and non overlap (as it is on smart)
    for i in domainList:
        ListFe=[]
        ListeF=i.split('\n')
        for x in ListeF:
            a,b = x.split('=')
            ListFe.append(b)
        if ListFe[0]!='low_complexity_region' and ListFe[0]!='coiled_coil_region' and ListFe[5][-2:]=='OK':
```

```
            ListFeatures.append(ListFe)

#Find Hits PKI-type Hits
    seq=str(record.seq)
    seq= seq.upper()
    Seq = 'XXXX'+seq
    L=len(Seq)
    l=len(seq)
    print str(record.id)

    Hits =[]
    #an iteration to make sure that we also get overlapping hits
    for x in range(L-14):
        piece =Seq[x:x+15]
        res=re.search(r'....[LIVMFWAY][^P][^P][^P][FMLIVYW][^P](2,3)[LMIVFWAY][^P][LIMVFPWY]',piece)
        if res!= None:
            hit = res.group()
            a,b = res.span()
            a,b = a+x-3,b+x-4

            Hits.append(( a,b,hit))

    Hits = list(set(Hits))#remove doubles
    Hits = sorted(Hits)
    hits=[]
    for a,b,c in Hits:
        hits.append([a,b,c])

#Filter With Domains
    for x in hits[:]:
        done =False
        for y in ListFeatures:
            if x[0]+4<int(y[2]) and x[1]>int(y[1]): #a+4 because of phi0 residue ?????
                x.append('in '+y[0])
                done = True
                break
        if done:
            continue
        else:
            x.append('n.i.d.')

#Get Averaged Disorder Values for 3 regions
    hitsGraded=[]
    for x in hits:
        a,b,c,d = x[0],x[1],x[2],x[3] #start end seq,domain
        dis=0
        disB=0
        disA=0
        if a<1:
            for x,y in ListDis[0:b-1]:
                dis =dis+float(y)
            dis=dis/(b-1)
        else:
            for x,y in ListDis[a-1:b-1]:
                dis =dis+float(y)
            dis=dis/(b-a)
        #Getting disorder average of before and after
        if a<2:
            disB=00.00
        elif a<7:
            for x,y in ListDis[0:a-1]:
                disB =disB+float(y)
            disB=disB/(a-1)
```

```
        else:
            for x,y in ListDis[a-7:a-1]:
                disB =disB+float(y)
            disB=disB/(6)

        if l==b:
            disA=00.00
        elif l-b<6:
            for x,y in ListDis[b:]:
                disA =disA+float(y)
            disA=disA/(l-b)
        else:
            for x,y in ListDis[b-1:b+5]:
                disA =disA+float(y)
            disA=disA/(6)

        hitsGraded.append([disB,dis,disA,c,a,b,d])  #disorder of seq,disorder of 6 aa before, disorder of 6aa
after,sequence,start,end,domain

#Calculate NESScore

    for x in hitsGraded:
        i = x[3]
        L = len(i)
        o0 =i[2-1]
        o1 = i[5-1]
        o2 = i[9-1]
        o3 = i[-3]
        o4 = i[-1]

        #Scoring for o0 and its neighbors
        if o0 in 'IVML':
            s0 = 0.5
        elif o0 is 'FAYW':
            s0= 0.25
        else:
            s0=0

        if i[0] in 'DE':
            s0 = s0 + 0.5
        if i[2] in 'DE':
            s0 = s0 + 0.5
        if i[3] in 'DE':
            s0 = s0 + 0.5
        if s0<1:
            s0=1
        else:
            s0=float(s0)
        #Scoring for o1 strength
        if o1 is 'L':
            s1 = 10
        elif o1 is 'I':
            s1= 8
        elif o1 in 'VM':
            s1 = 6
        elif o1 is 'F':
            s1=4
        elif o1 is 'A':
            s1=2
        elif o1 in 'WY':
            s1=1
        #Scoring for o2 strength
        if o2 in 'FML':
```

```
        s2 = 10
    elif o2 in 'IV':
        s2 =8
    elif o2 is 'Y':
        s2=3
    elif o2 is 'W':
        s2=1
    #Scoring for o3 strength
    if o3 in 'LM':
        s3 = 10
    elif o3 is 'I':
        s3= 9
    elif o3 in 'V':
        s3 =7
    elif o3 in 'F':
        s3=3
    elif o3 in 'WA':
        s3=2
    elif o3 in 'YT':
        s3=1
    #Scoring for o4 strength
    if o4 in 'L':
        s4 = 10
    if o4 in 'I':
        s4 = 9
    elif o4 is 'M':
        s4= 8
    elif o4 in 'V':
        s4 =7
    elif o4 in 'F':
        s4 =5
    elif o4 in 'PYW':
        s4=1
    Score = int(s0*s1*s2*s3*s4)
    x.append(Score)

    #Simplify the Disorder values
    if x[0]<0.25:
        DisIn=1
    elif x[0]<0.5:
        DisIn=2
    else:
        DisIn=3
    if x[1]<0.25:
        DisSeq=1
    elif x[1]<0.5:
        DisSeq=2
    else:
        DisSeq=3
    if x[2]<0.25:
        DisOut=1
    elif x[2]<0.5:
        DisOut=2
    else:
        DisOut=3
    x[0]=DisIn
    x[1]=DisSeq
    x[2]=DisOut
    Pri=''
    for em in x:
        Pri= Pri + str(em)+'\t'
    print Pri, '\n\n'
```

12.1.2 REV-type NES Prediction

```
#Import used libraries
import os
from Bio import SeqIO
from Bio import Seq

input='/users/koray/desktop/NESprediction/Trial.fasta'   #input fasta file

#Parse Input File into Sequences
handle = open(input)
for record in SeqIO.parse(handle,'fasta'):
    #Put single sequence into a temp file
    liste=[]
    liste.append(record)
    handle2=open('temp.fasta','w')
    SeqIO.write(liste,handle2,'fasta')
    handle2.close()

#Get Disorder For a Single Protein
    p = Popen('./iupred temp.fasta short', shell=True, stdin=PIPE, stdout=PIPE, stderr=PIPE, close_fds=True)
    (stdin, stdout, stderr) =(p.stdin, p.stdout, p.stderr)
    results = stdout.readlines()
    x=0
    ListDis=[]
    for i in results:
        if i[0] =='#':
            continue
        a,b = i.split('    ')
        b,c=b.split('\n')
        x=x+1
        ListDis.append((x,b))
    stdout.close()

#Get Domains from SMART Database
    p = Popen('perl Smart_batch.pl --inputFile temp.fasta --outputDirectory .../NESprediction/ --includePfam',
shell=True, stdin=None, stdout=None, stderr=None)
    p.wait()
    handle3= open('tempDomain.txt','r')
    domains= handle3.read()
    handle3.close()
    domainList = domains.split('\n\n')

    #process the result file, remove the tags
    del domainList[0:2]
    del domainList[-1]
    ListFeatures=[]

    #put values in to list format with filters for low complexity and coiled coils and non overlap (as it is on smart)
    for i in domainList:
        ListFe=[]
        ListeF=i.split('\n')
        for x in ListeF:
            a,b = x.split('=')
            ListFe.append(b)
        if ListFe[0]!='low_complexity_region' and ListFe[0]!='coiled_coil_region' and ListFe[5][-2:]=='OK':
            ListFeatures.append(ListFe)

#Find Hits REV-type Hits
    seq=str(record.seq)
    seq= seq.upper()
    Seq = 'XXXX'+seq
    L=len(Seq)
```

```
    l=len(seq)
    NAME= str(record.id)

    Hits =[]
    #an iteration to make sure that we also get overlapping hits
    for x in range(L-9):
        piece =Seq[x:x+9]
        res=re.search(r'[LIVM][P].[LIVMF]..[LMIV].[LIMVF]',piece)

        if res!= None:
            hit = res.group()
            a,b = res.span()
            a,b = a+x-3,b+x-4

            Hits.append(( a,b,hit))


    Hits = list(set(Hits))#remove doubles
    Hits = sorted(Hits)
    hits=[]
    for a,b,c in Hits:
        hits.append([a,b,c])

    if Hits == []:
        continue

#Filter With Domains
    for x in hits[:]:
        done =False
        for y in ListFeatures:
            if x[0]+4<int(y[2]) and x[1]>int(y[1]): #a+4 because of phi0 residue ?????
                x.append('in '+y[0])
                done = True
                break
        if done:
            continue
        else:
            x.append('n.i.d.')

#Get Averaged Disorder Values for 3 regions
    hitsGraded=[]
    for x in hits:
        a,b,c,d = x[0],x[1],x[2],x[3] #start end seq,domain
        dis=0
        disB=0
        disA=0
        if a<1:
            for x,y in ListDis[0:b-1]:
                dis =dis+float(y)
            dis=dis/(b-1)
        else:
            for x,y in ListDis[a-1:b-1]:
                dis =dis+float(y)
            dis=dis/(b-a)
        #Getting disorder average of before and after
        if a<2:
            disB=00.00
        elif a<7:
            for x,y in ListDis[0:a-1]:
                disB =disB+float(y)
            disB=disB/(a-1)
        else:
            for x,y in ListDis[a-7:a-1]:
```

```
            disB =disB+float(y)
        disB=disB/(6)

    if l==b:
        disA=00.00
    elif l-b<6:
        for x,y in ListDis[b:]:
            disA =disA+float(y)
        disA=disA/(l-b)
    else:
        for x,y in ListDis[b-1:b+5]:
            disA =disA+float(y)
        disA=disA/(6)

    hitsGraded.append([disB,dis,disA,c,a,b,d])  #disorder of seq,disorder of 6 aa before, disorder of 6aa after,sequence,start,end,domain

#Calculate NESScore

  for x in hitsGraded:
    x.append('REV-Type')

    #Simplify the Disorder values
    if x[0]<0.25:
        DisIn=1
    elif x[0]<0.5:
        DisIn=2
    else:
        DisIn=3

    if x[1]<0.25:
        DisSeq=1
    elif x[1]<0.5:
        DisSeq=2
    else:
        DisSeq=3

    if x[2]<0.25:
        DisOut=1
    elif x[2]<0.5:
        DisOut=2
    else:
        DisOut=3

    x[0]=DisIn
    x[1]=DisSeq
    x[2]=DisOut
    Pri=''
    for em in x:
        Pri= Pri + str(em)+'\t'
    print Pri

  print '\n\n'
```

12.2 PROTEIN IDENTIFIERS OF THE DATA SETS

12.2.1 NESdb Proteins

O95149	O14746	O15360	Q99612	Q8N720	Q9BY84
Q96D46	P24385	Q8N668	Q16828	P15336	P14373
P61925	P35869	P11388	P03372	P38936	P25054
P04637	P46108	O15392	P51587	P67775	O00221
P43487	Q01094	Q9UNH5	P05230	Q6UB99	Q9UMX3
P38398	P56693	Q01658	O75832	Q96JZ2	Q16236
P14635	Q9UQL6	P48436	Q9NQS1	Q9HC62	Q9NYF0
Q00987	P56524	Q9HAP2	O94916	Q96T21	P55265
Q13485	Q16254	Q9BYM8	O43196	Q86TB9	Q7RTN6
O60716	Q13043	Q14457	Q9H4D5	Q96RS0	P30740
Q06787	Q9GZX7	P46777	Q9BRK4	Q9H9S0	Q15172
P42566	Q05397	Q9UGR2	Q96K30	Q9HBL8	O95613
Q04206	Q13568	Q93052	O15457	O43707	O00311
P42224	O95644	O96018	P40692	Q9Y3I1	Q17RY0
Q14872	P06748	Q96C86	Q14145	P78545	Q8IXJ6
Q9Y572	Q02880	O15519	O15265	Q9NRA8	P28289
P25963	Q99653	P30291	Q14140	O96013	Q9HCE7
Q14653	P41970	Q9BZB8	Q13148	Q9P286	O00255
O15350	Q14494	Q9Y3M2	Q00535	Q00653	P49023
Q15797	Q13490	Q9BVS4	Q92688	P42858	Q8IVI9-3

Table 12-1 List of Uniprot IDs of human proteins of NESdb

P14635	P26651	Q9UHL0	Q15654	P40424	P17931
Q02750	P41743	P41134	Q9UD71	Q9NP71	O00327
Q92664	Q00994	Q02363	Q14012	Q03052	O00571
P49137	P40763	O60934	P54646	P06241	Q92598
P00519	P26196	Q7Z2W4	Q9UBK2	Q9Y261	Q92905
Q15942	Q16236	Q13887	Q13574	O15530	O95863
O15534	P41238	Q6NXT1	Q8IW41	Q9H3M7	O14867
O15055	Q9NSV4	Q9NUL3	Q9NRF2	Q6PIJ6	O00401
P51178	Q8NHY2	Q99593	O75553	Q9UDY2	

Table 12-2 List of Uniprot IDs of human ortholog proteins of NESdb

12.2.2 LMB Study Proteins

Q13501	B3KS26	Q9NVU7	Q71V88	B4DY85	B4DFI6
C9JDG0	Q53EY3	B4DH66	B3KPH9	B4DGH7	P62917
Q9NVN8	B5BTY4	Q9Y2U5	B4DHJ3	O43818	Q76N54
Q13895	Q13867	B4E303	Q53GL4	Q99873	Q5VVD0
Q00653	Q12769	Q96CT7	Q53RG0	B7Z6Z3	P46776
Q12800	Q9BRP8	E7EQZ4	Q92973	O00488	B2R4K7
Q2NL82	O00629	A8K333	A8K4W0	Q9BQE6	O95036
B3KT11	Q9BZ95	O15397	Q92905	B4DZW4	Q9BYF5
B4DTT8	Q6NVW7	Q9UBB6	A8K0B9	Q2TAC2	Q1JQ76
A8K066	Q96EY4	Q9BRJ6	Q9H1A4	B2RBA0	Q6IBH6
D2CFK9	Q9UHA3	Q8N6M0	Q9BV44	Q2TA84	Q53Z07
Q9H7E9	Q8NCK5	B4DTL6	Q8TEX9	B4DM84	P61513
Q96QL0	B2R823	Q9H1C7	Q8IYV2	Q9UPW5	P62888
P56537	B3KM71	B4DR01	Q9UI30	Q9NPD3	Q53H34
Q5SRE5	Q13823	P13861	Q53FF6	Q9UJX2	F8W727
Q04206	Q9H8Y5	Q56VW8	Q6NW29	Q8N1F7	P18077
C9JEH3	Q8TCG1	Q5BKZ2	B3KNI0	Q9UJZ1	B2R4C1
Q92529	B5BUK7	B3KTA3	A8K4M5	Q9Y2L1	Q6IAX2
B3KMR5	Q9BRP1	Q9HBB9	Q96KJ8	B7Z5J8	P49207
Q9NVB0	B4DII5	Q9H3S7	Q76MU0	P83881	Q59EL2
A8K2F9	Q53F09	B7Z871	Q5T1Z8	Q0QEW2	Q567U8
Q53FV3	Q7L5N1	Q99627	Q06210	Q59GS5	Q6VY07
Q6Y7W6	A8K3Q9	A8K070			

Table 12-3 List of Uniprot IDs of proteins from LMB Study

12.2.3 1265 Significant Hits of SILAC MS Data

P35869	Q9UMR2	Q12774	P62424	P35269	Q8WW12
Q14145	P55039	P16989	P83731	Q9Y606	P53041
Q86TB9	Q9C0B9	Q99570	P19525	P09012	P02545
Q93052	P11274	P10398	Q9BRZ2	Q96B26	P49736
Q17RY0	Q96GV9	Q9Y3U8	Q9BXP5	P47755	Q9NP77

Q2TAM5	F5GWA6	O15355	Q8NHQ9	Q6P6C2	P26358
Q96D46	Q15628	P53618	P62241	Q9UNQ2	Q14566
Q8IXJ6	O60678	Q9NWT1	Q8TBB5	Q8N163	J3KTL2
Q9BVS4	Q9UHB9	P19784	P13489	Q9H9A5	O43390
Q14653	Q7Z739	Q96HC4	Q13627	Q12986	Q9Y2W1
P25963	Q92572	Q13347	A6NIH7	Q01844	Q13123
Q13043	Q6YHU6	Q8TED0	Q5T1V6	Q8WX92	E9PAV3
P30291	Q9BQA1	O15116	P67809	Q9Y224	O14979
P25054	Q8IWV7	P46934	P62081	P09661	P62913
Q99717	Q96PU5	P52434	Q96N67	O75794	P23246
Q9NRA8	O43933	P39019	J9JIC5	Q00688	Q9UKM9
Q9NQS1	Q7LBC6	B3KYA7	Q9NZL4	F8VVA7	J3KN67
Q92688	Q96QG7	P55209	Q2M389	E9PC97	Q27J81
P67775	P25398	P29597	O43148	B1AHD1	Q15691
P42566	P60228	Q9BW66	Q69YN2	Q9H9A6	Q9H307
O96013	Q9BQL6	Q7Z7L7	Q96ME1	Q9Y2A7	O43395
P46777	O43684	P62249	P30050	Q3KQU3	O95239
P42224	O95271	Q9NRF8	A8MSH5	K4DI95	P62995
Q13485	P49757	Q9Y295	O00458	O75436	P30101
O60716	Q9Y3S1	O60942	Q9UBK8	H7BXP1	O60832
P06748	P05783	P19388	Q8ND56	Q9BX40	Q99459
Q13148	Q9NW82	Q8NCE2	G3XAI2	Q9UPQ0	P67936
P14635	E9PBB4	O95218	J3KNK4	Q14CX7	P13797
E7ENU4	B7ZKM8	O94887	P07355	Q15056	P08621
P43487	Q92900	P53621	Q9UPR3	Q9GZS3	Q07666
O43196	P61160	Q9UBF2	O15145	P12956	P31946
O96018	Q15370	Q9UNF1	Q92620	Q08AD1	Q15233
Q9HAP2	Q5VTI5	Q6R327	Q9NSD9	Q13136	A8MXP9
Q6NXT1	Q8NHG8	P17812	Q8TAF3	P42285	B2RNG4
C9JDG0	Q6PL24	Q9HAU5	P62314	E9PRY8	Q92997
Q15654	Q9ULJ7	Q9Y4H2	Q9UK41	Q14103	B5MCU0
O00571	P27361	O15514	Q15061	B7ZKT7	Q8NFZ5
Q15942	O15234	Q8NE71	Q9P265	P78346	P07197

O15530	Q15436	P05388	Q9NR12	O60306	E7EN19
Q02750	P49674	Q6P1N0	Q9Y6W5	O95747	E9PAL6
P41743	Q9Y597	A5YKK6	P46783	P49588	E9PC69
P49137	Q676U5	Q504Q3	Q9HB21	Q9H6H4	E9PF99
Q7Z2W4	Q13492	P49792	Q15054	P09651	E9PR30
P26196	Q9UDY8	P23193	Q9P2Y5	Q13247	F5GZ78
Q9UDY2	P49790	P59998	Q5VYS8	Q13151	F5H604
Q9NUL3	Q9P107	Q13371	Q96SB4	Q8NCA5	H0Y4E8
P14635	P68400	P62851	F5H2R7	P02792	H0YEF7
O00401	Q5VW36	Q7Z460	O43583	Q7L2E3	H0YKD8
Q13574	Q52LW3	Q5T5C7	Q9Y4E8	Q6UXN9	H3BQK9
O15534	Q6PD74	Q96PY6	O43823	P22087	P62068
Q5D1E8	Q6P2P2	Q9UJY4	Q6PGP7	Q96B97	J3KPM3
Q9UI10	Q9HC44	J3KP97	Q68DQ2	P18754	K7EM18
A0AV96	Q14558	Q9Y6V7	H0Y9Z5	Q9NZI8	Q8TBC3
Q9H6K1	Q15057	P62269	Q9UFF9	Q70J99	O00507
P47736	P60866	P62847	Q86X10	Q8TDJ6	O14907
B4DZH6	Q92600	Q5T447	O15066	P30622	O15063
Q9ULI2	I3L097	B4DXJ1	Q9BZF9	P63104	O15553
Q8TF05	P46060	Q9BY44	Q8ND04	B1AK88	O43900
Q00537	Q9H4B6	O43865	O60318	M0QZR4	O60239
Q13546	Q9UPM8	O94927	Q93073	Q9Y5S9	P0CAP2
Q9Y4K1	P22681	O00443	Q13610	O75116	P15170
Q15434	Q13615	Q5VZE5	Q9H074	Q14847	P36954
Q8IXW5	P24928	Q9BVM2	P17612	P35241	P41279
Q9Y484	Q8IW35	Q53EZ4	B4DYX9	Q5W0V3	Q9H1A6
Q00653	O14617	Q6ZS17	P26373	Q13428	P62273
Q8WU79	Q13496	Q4J6C6	Q96P48	P06730	P62304
Q99567	P61962	Q96HN2	P61981	O00267	P62312
Q9BQK8	Q9BUB5	P49840	P19338	P47914	P62857
Q5U5Q3	P53992	P62263	Q9UNS2	Q9UJU6	P62875
Q6P2H3	P62487	Q9HD67	Q7Z2T5	Q15185	P62891
Q13144	P78337	Q9UKZ1	Q8NC51	Q6GYQ0	P63167

Q8WUX9	J3KNL6	P29353	Q2TAY7	O95999	P63220
O75843	Q9UGP4	Q99613	B4DR52	Q8TD19	D6R919
Q2TAZ0	O43164	Q9H6S0	P08670	Q9UBU9	P84090
Q9H6D7	Q15652	P30153	Q9NW08	Q9Y5S2	Q13772
Q8N122	Q8N8R7	E9PM46	P47813	P15311	Q14679
C9JC87	Q13190	Q8TC07	F5HFY4	P05386	F5GY24
Q8WXG6	Q8N6H7	J3KNR0	Q8TD16	Q93009	Q15051
Q5JTW2	Q9H9H4	Q5VT06	Q8IUD2	Q15637	Q15366
Q96K76	P52630	P62280	Q14527	Q96II8	Q16204
Q9UNY4	Q9UMZ2	Q9H267	O60333	O43149	Q5JTC6
P09132	O14744	Q658Y4	P52907	O14776	Q5T2D3
Q8NDI1	P37198	Q9H0W8	Q9NSI2	P13984	Q5TC82
O00203	G5E948	P25098	O00178	Q9BUJ2	Q5XG87
Q96B36	Q14671	Q9Y230	P0C0S8	Q14978	Q6P3S1
Q8IX03	Q86XL3	Q9Y6K9	P23921	Q08945	Q86UY5
Q96CN4	Q9BPY3	P19387	Q969S3	P51114	Q8N2Y8
P49458	Q9BYV8	Q96RG2	G3V0I6	P52272	Q8N4B1
P35658	B4DKT0	Q7Z4H7	Q92625	Q04917	Q8NDG6
O95684	P41091	Q96QR8	Q12768	P52701	Q8NDH2
Q5T7W7	Q9BSJ2	Q8NEC7	Q96HR8	Q9C0C2	Q8WVF1
Q93008	Q9BX10	P41214	Q9UBQ5	O60294	Q8WZ19
Q8IYB5	Q9Y6K5	Q96F86	Q15650	P17844	Q96A73
Q8NDV7	O94855	Q5TKA1	O75937	P31946	Q96E09
A7E2V4	Q14157	Q13617	Q86VP6	Q86UK7	Q96FQ6
B4E0T2	Q9BT25	P61966	Q7L7X3	Q9Y3I0	Q96MX6
P60510	Q9UQ80	Q9P1U1	Q00341	P26599	Q96P47
Q9UPU7	Q9Y3S2	Q9NZQ3	Q86XZ4	Q9NZN8	Q96QP1
P19838	P37108	Q5H9R7	Q92917	Q9BZZ5	Q99584
P40855	Q9Y217	O75607	Q9Y285	Q6P1J9	Q99961
O75420	Q9Y6Y8	Q86UU1	B5ME97	O15027	Q9BYI3
Q15369	Q63HR2	B0QY89	P84098	Q9UKD2	Q9H0K1
Q14674	P38606	Q86UU0	Q9UPY3	A6NMQ1	Q9H8N7
P53677	Q9Y5A9	O60518	Q96KG9	O43143	Q9H992

P13807	O95721	Q12905	Q01970	Q8TDX7	Q9HD40
Q9H446	Q8NFH4	P07900	P05387	H3BLZ8	Q9NP61
D6REX3	O94972	Q9Y5T5	P23258	Q96DH6	Q9NRH1
Q9UHD2	B4DZD6	Q14258	P27695	B4DWW4	Q9NTX7
P51617	P57772	P62316	Q68CZ6	P23284	Q9NWQ9
Q5SQN1	Q96AC1	P62266	Q9P258	G3V1L9	Q9NYY3
Q6UUV9	Q6PGN9	Q96ED9	Q7Z6B7	Q9Y314	Q9P1Y5
Q9NYF3	P48729	Q9UL18	Q9NR09	Q06265	Q9UH36
Q6AWC2	Q9H0J9	O95163	Q9BZI7	Q8NBJ5	Q9UK97
Q8N961	Q15843	Q9H0K6	Q5T6N4	Q9Y5B9	Q9ULL1
Q86VV8	O15143	O14641	Q14155	Q16513	Q9Y2J4
P47974	Q6P158	P62979	Q13393	Q14498	Q9Y333
Q8WUF5	Q9C0C7	P61964	O14920	O14777	R4GN35
Q9Y2T2	C9J6P4	P62277	Q9Y5K6	G8JLB6	J3KQN4
Q86YR5	Q9UHW5	Q9UQN3	O43172	P10155	Q5SRE5
Q9H3F6	F5H0R1	O95714	P15880	O43809	P62888
Q9H6R7	Q8IV48	B8ZZ87	P62753	E7ES43	Q9UHR6
Q9P2E3	Q96A49	P41240	Q00839	Q9Y613	Q5T1Z8
B9EGP5	Q99081	Q9Y265	Q9UH62	Q92541	Q8TCG1
Q9Y446	B4DNJ6	Q96CS2	Q5F1R6	O95456	Q93052
P46976	P62136	A8MXB7	Q01105	Q7KZF4	Q9UBB6
Q9UKE5	Q15025	Q9Y6B7	O14802	P30260	Q9BRP1
Q58A45	P62854	O14579	Q641Q2	Q9Y520	C9JDG0
Q2M2I8	J3KPL5	Q63ZY3	Q6P0Q8	Q9UKY7	P49207
Q68DC2	O60566	P46781	Q8WWM7	C9K060	Q15654
P52298	Q5JTD0	B0YIW6	P46782	Q01085	Q2TAM5
Q9BYJ9	Q96EY5	Q5SRQ6	Q92574	Q6ZSJ8	O00571
P36507	P62714	Q14008	P53999	O95155	P57678
Q70E73	P11908	P56377	P61313	P28290	Q96D46
Q96FC9	O76094	Q8TF46	Q96EV2	P35658	Q8N1F7
Q9H977	Q9H425	P52306	F8VQ10	Q00403	Q9NRX1
Q6UUV7	O43572	Q96AZ6	O60784	Q9HB71	Q15942
P11532	Q5VV41	P61221	P02794	Q9NVP2	P30876

Q9P0J7	Q7L2H7	O60231	P49005	Q9Y294	P18077
O95486	Q8N8S7	P42766	P08579	P35637	Q9UHI6
Q7Z309	Q99816	P36873	Q00577	Q01130	Q9BVS4
Q8TEU7	Q5T4S7	Q5VZK9	O15160	Q86TU7	Q13501
Q684P5	Q9Y4G8	Q9Y3A5	Q9Y6I3	O75165	Q13895
P49770	O00743	P52594	Q99543	Q16626	Q13310
F8WEA9	M0R2B7	Q92499	Q8N1G2	P63279	P63151
O43524	Q9H7P9	Q8NFH3	O00139	O15347	Q9UPW5
Q92539	P50548	Q9Y4B6	Q86US8	B4DUT8	Q6Y7W6
J3KNE0	P55010	O60256	Q99707	Q13442	O60573
Q9ULX3	P62244	F5H527	P07951	O94888	Q9UJX4
Q96FK6	Q92540	P08865	P54577	Q9BVP2	Q14247
Q9BVC4	Q66LE6	Q9Y450	O43633	Q53GS9	Q6NW29
Q9UPN7	O94804	E7EX17	Q96J02	Q99729	P39023
Q9NR50	P61158	O15357	P46063	Q16851	P11940
O95905	Q8NEB9	Q9H2U1	O00562	Q9H4M9	Q12769
Q15437	Q9NTZ6	Q2NKX8	O60506	P27348	Q2NL82
Q01968	Q8IWV8	O15371	Q6PKG0	Q7Z3B4	P62913
Q6BDS2	Q15047	Q14152	O75822	P61978	Q02543
Q53ET0	O75688	O75821	Q8N4C8	O60749	P61254
P04049	Q6ULP2	Q00536	I3L504	Q9UQE7	Q9BRP8
A4D1P6	Q8N6T3	Q14320	P40818	Q13325	P36578
P50552	Q8TEQ6	Q5VY93	Q8TCY9	Q8IWC1	Q9NVN8
Q9NT62	O75818	P14735	P35573	P78537	P62899
Q86SQ0	Q9BRX2	P85037	F5H1X8	Q9NW13	Q13451
Q9H000	P68104	O60524	Q9BPZ7	P61764	Q9H3S7
Q14232	P20042	P60891	Q8IYB7	F8WCP6	O15294
Q9NWT6	F5H874	P61353	P10644	P78527	P41743
Q2M3G4	Q9UM82	P35250	Q96P16	P07910	Q96GA3
Q9BVQ7	Q3V6T2	Q9H9G7	Q96PZ0	K7ELC2	Q9BWH6
Q15468	Q12899	Q07960	O00541	Q86V48	Q12800
Q13829	P60842	Q3MHD2	Q5GLZ8	Q5RKV6	Q9UI30
Q7L3B6	Q9UK59	H7C107	Q9NUP1	P31942	Q5T6F2

Q6P3W7	B3KXW5	Q9BW27	Q15477	Q8WXX5	Q8N6R0
P49841	O15144	Q13616	Q5SW79	P55072	Q9H8Y5
Q04864	A3KN83	P62750	Q9Y6A4	P84103	P50914
Q96IF1	O14730	Q10567	Q8WX93	H7BZT4	Q9BZE4
Q96EN8	Q8WVM8	Q9Y678	O15511	B4DVB8	Q6VY07
J3KPC8	Q9ULC4	Q9BPX5	Q9H6U6	P26583	Q9H1A4
Q9UPQ9	E9PB12	Q9HCE1	Q12906	Q9H7D0	Q9BV44
O00750	Q6P2E9	P23396	Q9P031	P09429	P62917
B3KS98	P62140	Q96P11	Q53EL6	D6W592	P40429
P78406	Q562E7	Q66K74	Q13618	Q99575	O00488
P32121	Q9UPQ3	Q9NVM6	Q9NZB2	Q9Y2W2	P62906
P05198	O60504	Q9NZT2	O75175	J3KQ32	P61247
Q9UKV8	F5H2M7	Q96S55	Q9UIV1	P38432	Q9BRJ6
Q8TEH3	B4DGT8	Q75MJ1	Q08J23	O75663	J3QQ67
Q8NB90	B8ZZN6	Q147X3	P27635	P11021	P18124
Q8TB45	O95628	B7WPE2	O43670	Q13283	P46776
Q969Q6	Q8IZH2	Q8IXQ3	Q14181	P22694	Q96CT7
Q0JRZ9	P62495	Q7Z2Z2	O95707	Q92878	P32969
Q9NUU7	Q15020	Q9NYL2	Q92979	Q96PK6	P52292
O00442	P63244	Q8N1G4	Q6RFH5	P33991	Q02878
G3V3G9	B3KSH1	Q14677	B4E0Y9	P49643	O00505
B4DR47	P53990	P05455	P62258	Q04323	F8W727
Q5PRF9	Q9GZZ9	P78344	P13010	O75717	P46778
Q9NRY5	P55884	Q8WU90	Q13162	P22626	Q8N6M0
O60841	B4DQM4	P35658	P42694	Q16630	P56537
Q13287	Q9H6T3	P38935	Q0VDF9	Q9UKK3	O00629
Q9H1Y0	P30154	Q14C86	Q9NQT8	Q16629	P18621
Q68CZ2	P13639	P39748	P08238	Q7Z417	Q9UJX2
P61011	G5E9Q2	M0QZW1	Q9NW64	P43246	Q9BQ67
Q8IZW8	Q7Z478	Q8IWZ3	P30419	P31327	Q9Y2L1
P28482	Q9BXS5	Q2KHT3	Q9C0C9	Q9NUQ3	Q9H7E9
P52948	Q9UET6	Q8N3C0	Q9UMY1	Q9NYF8	P61201
Q8WWN8	Q9NX04	Q96JN8	P33992	Q16543	Q7L5N1

Q3YEC7	Q09161	G3V4K3	Q14444	Q9NPD8	P14635
Q9C0D3	Q9NPI6	P62701	O00425	H3BSH7	P43487
Q9H6S3	Q15036	Q04637	O75179	Q14738	P61513
Q92747	H0YLI7	P08708	Q14683	Q9BQ70	Q99873
Q6ZRV2	Q9P2D3	P62310	E7EVA0	Q9UN86	
Q9NWS0	Q14244	Q9P2D0	P20810	P39687	
O15111	P35606	P40938	O15061	O43432	
Q9H814	P38919	Q96H20	Q08211	E9PB61	

Table 12-4 List of Uniprot IDs of proteins from MS analysis

12.2.4 321 Proteins with a Predicted NES

Q6UUV9	O95684	Q5VTI5	Q13371	Q9UNS2	Q16513
P47974	Q8IYB5	Q9ULJ7	P62851	Q8TD16	Q9Y613
Q53ET0	Q8NDV7	O15234	Q7Z460	Q8IUD2	P28290
P04049	Q9UPU7	Q13492	Q5T5C7	Q14527	Q16626
Q6P3W7	P40855	Q9P107	J3KP97	O60333	Q9UQE7
Q86XL3	Q14674	P60866	Q9BY44	Q9NSI2	P61764
B4DZD6	Q9H446	Q13615	O00443	G3V0I6	K7ELC2
Q9Y450	Q5SQN1	P24928	Q99613	O75937	Q86V48
Q00536	Q8N961	Q8IW35	J3KNR0	Q86XZ4	J3KQ32
Q8N1G4	Q8WUF5	Q9BUB5	Q5VT06	P84098	P11021
Q8IWZ3	Q9P2E3	P53992	P25098	Q9P258	Q92878
H0Y9Z5	B9EGP5	Q15652	Q8NEC7	Q5T6N4	E9PB61
Q9NR09	Q9UKE5	P52630	P41214	Q5F1R6	E9PAV3
O43633	Q58A45	Q14671	Q13617	Q641Q2	J3KN67
P26358	Q68DC2	B4DKT0	Q5H9R7	Q6P0Q8	Q15691
Q8NDH2	Q9BYJ9	Q9BSJ2	Q86UU1	P46782	P67936
P35869	P36507	O94855	Q86UU0	Q92574	E9PC69
Q86TB9	Q9P0J7	Q9Y5A9	P07900	Q96EV2	E9PF99
Q96D46	O95486	O95721	O95163	O60784	F5GZ78
Q8IXJ6	Q8TEU7	P57772	Q9UQN3	Q86US8	F5H604
Q14653	Q684P5	Q96AC1	Q96CS2	P07951	H0Y4E8
P25963	O43524	O15143	P52306	P54577	H0YEF7

P30291	J3KNE0	Q6P158	O60231	P46063	H3BQK9
P25054	Q9ULX3	Q96A49	F5H527	Q8TCY9	O15063
Q9NRA8	Q96FK6	P62136	Q2NKX8	P35573	O60239
Q9NQS1	Q6BDS2	Q15025	O75821	F5H1X8	P15170
O96013	Q9NT62	J3KPL5	P14735	Q8IYB7	Q14679
P42224	Q86SQ0	Q96EY5	O60524	O75175	Q15051
P14635	Q14232	P11908	P60891	B4E0Y9	Q15366
E7ENU4	Q13829	Q5VV41	Q10567	Q9C0C9	Q5T2D3
Q6NXT1	J3KPC8	Q5T4S7	Q9BPX5	Q14683	Q6P3S1
C9JDG0	Q9UPQ9	O94804	Q9NVM6	O15061	Q8WVF1
O00571	Q8TEH3	O75688	Q7Z2Z2	Q08211	Q96A73
Q02750	Q8NB90	Q9UM82	Q9NYL2	Q8N163	Q96QP1
O00401	Q0JRZ9	Q3V6T2	Q8WU90	Q00688	Q9H0K1
O15534	G3V3G9	P60842	P38935	K4DI95	Q9H8N7
B4DZH6	B4DR47	B3KXW5	Q8N3C0	Q08AD1	Q9HD40
Q9ULI2	Q9NRY5	Q562E7	Q9P2D0	Q13136	Q9NTX7
Q00537	O60841	Q9UPQ3	P62081	P42285	Q9Y2J4
Q13546	Q68CZ2	F5H2M7	O00458	E9PRY8	R4GN35
Q15434	P52948	B4DGT8	Q8ND56	B7ZKT7	Q5T1Z8
Q00653	Q8WWN8	O95628	G3XAI2	Q13151	Q9UBB6
Q5U5Q3	Q92747	Q8IZH2	Q9UPR3	P02792	Q13501
Q6P2H3	Q9C0B9	B4DQM4	Q92620	Q8TDJ6	Q9UPW5
Q13144	Q7Z739	Q15036	Q15061	P30622	Q9BRP8
Q8WUX9	Q96PU5	P35606	Q9P265	M0QZR4	Q9NVN8
Q2TAZ0	O43933	P53618	Q9P2Y5	O75116	Q9H3S7
Q8N122	Q9BQL6	P46934	O43823	Q5W0V3	Q96GA3
Q9UNY4	P05783	B3KYA7	Q6PGP7	Q9Y5S2	Q8N6R0
O00203	Q9NW82	Q9HAU5	Q68DQ2	P15311	Q9H8Y5
Q96B36	E9PBB4	Q8NE71	Q9BZF9	Q93009	Q9BZE4
Q8IX03	B7ZKM8	Q6P1N0	O60318	P52272	Q9H1A4
Q96CN4	Q92900	P49792	Q13610	B4DWW4	J3QQ67
Q9H7E9					

Table 12-5 321 Uniprot IDs of proteins with a predicted NES

13 REFERENCES

Adachi Y, Yanagida M (1989) Higher order chromosome structure is affected by cold-sensitive mutations in a Schizosaccharomyces pombe gene crm1+ which encodes a 115-kD protein preferentially localized in the nucleus and its periphery. *J Cell Biol*, **108:** 1195–1207

Alavian CN, Politz JC, Lewandowski LB, Powers CM, Pederson T (2004) Nuclear export of signal recognition particle RNA in mammalian cells. *Biochem Biophys Res Commun*, **313:** 351–355

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. (2007) *Molecular Biology of the Cell*. Garland Science

Andrade MA, Bork P (1995) HEAT repeats in the Huntington's disease protein. *Nat Genet*, **11:** 115–116

Arnold M, Nath A, Wohlwend D, Kehlenbach RH (2006) Transportin is a major nuclear import receptor for c-Fos: a novel mode of cargo interaction. *J Biol Chem*, **281:** 5492–5499

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25:** 25–29

Au S, Pante N (2012) Nuclear transport of baculovirus: revealing the nuclear pore complex passage. *J Struct Biol*, **177:** 90–98

Bayliss R, Littlewood T, Stewart M (2000) Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. *Cell*, **102:** 99–108

Bayliss R, Ribbeck K, Akin D, Kent HM, Feldherr CM, Gorlich D, Stewart M (1999) Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. *J Mol Biol*, **293:** 579–593

Beckett D, Kovaleva E, Schatz PJ (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci*, **8:** 921–929

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. *Science*, **334:** 1524–1529

Bilokapic S, Schwartz TU (2012) 3D ultrastructure of the nuclear pore complex. *Curr Opin Cell Biol*, **24:** 86–91

Bischoff FR, Klebe C, Kretschmer J, Wittinghofer A, Ponstingl H (1994) RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proceedings of the National Academy of Sciences of the United States of America*, **91:** 2587–2591

Bischoff FR, Gorlich D (1997) RanBP1 is crucial for the release of RanGTP from importin beta-related nuclear transport factors. *FEBS Lett*, **419:** 249–254

Bischoff FR, Krebber H, Smirnova E, Dong W, Ponstingl H (1995) Co-activation of RanGTPase and inhibition of GTP dissociation by Ran-GTP binding protein RanBP1. *EMBO J*, **14:** 705–715

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

Bogerd HP, Fridell RA, Benson RE, Hua J, Cullen BR (1996) Protein sequence requirements for function of the human T-cell leukemia virus type 1 Rex nuclear export signal delineated by a novel in vivo randomization-selection assay. *Molecular and cellular biology*, **16:** 4207–4214

Bohnsack MT, Czaplinski K, Gorlich D (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10:** 185–191

Bohnsack MT, Regener K, Schwappach B, Saffrich R, Paraskeva E, Hartmann E, Gorlich D (2002) Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *EMBO J*, **21:** 6205–6215

Bonner WM (1975) Protein migration into nuclei. II. Frog oocyte nuclei accumulate a class of microinjected oocyte nuclear proteins and exclude a class of microinjected oocyte cytoplasmic proteins. *J Cell Biol*, **64:** 431–437

Brownawell AM, Macara IG (2002) Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. *J Cell Biol*, **156:** 53–64

Brunet A, Kanai F, Stehn J, Xu J, Sarbassova D, Frangioni JV, Dalal SN, DeCaprio JA, Greenberg ME, Yaffe MB (2002) 14-3-3 transits to the nucleus and participates in dynamic nucleocytoplasmic transport. *J Cell Biol*, **156:** 817–828

Bullock TL, Clarkson WD, Kent HM, Stewart M (1996) The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (NTF2). *J Mol Biol*, **260:** 422–431

Chang C-C, Lin C-J (2001) LIBSVM: A library for support vector machines.

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

Chuderland D, Konson A, Seger R (2008) Identification and characterization of a general nuclear translocation signal in signaling proteins. *Mol Cell*, **31:** 850–861

Cingolani G, Petosa C, Weis K, Muller CW (1999) Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature*, **399:** 221–229

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25:** 1422–1423

Colwell LJ, Brenner MP, Ribbeck K (2010) Charge as a selection criterion for translocation through the nuclear pore complex. *PLoS Comput Biol*, **6:** e1000747

Craig E, Zhang ZK, Davies KP, Kalpana GV (2002) A masked NES in INI1/hSNF5 mediates hCRM1-dependent nuclear export: implications for tumorigenesis. *EMBO J*, **21:** 31–42

Cullen BR (1992) Mechanism of action of regulatory proteins encoded by complex retroviruses. *Microbiol Rev*, **56:** 375–394

Dean KA, von Ahsen O, Gorlich D, Fried HM (2001) Signal recognition particle protein 19 is imported into the nucleus by importin 8 (RanBP8) and transportin. *J Cell Sci*, **114:** 3479–3485

Delphis A, Beman A, Batavorum L (1719) Leeuwenhoek, A. van: Opera Omnia, seu Arcana Naturae ope exactissimorum Microscopiorum detecta, experimentis variis comprobata. *Epistolis ad varios illustres viros J*,

Denning DP, Patel SS, Uversky V, Fink AL, Rexach M (2003) Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc Natl Acad Sci U S A*, **100:** 2450–2455

Dian C, Bernaudat F, Langer K, Oliva MF, Fornerod M, Schoehn G, Muller CW, 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

Dinkel H, Michael S, Weatheritt RJ, Davey NE, Van Roey K, Altenberg B, Toedt G, Uyar B, Seiler M, Budd A, Jodicke L, Dammert MA, Schroeter C, Hammer M, Schmidt T, Jehl P, McGuigan C, Dymecka M, Chica C, Luck K, Via A, Chatr-Aryamontri A, Haslam N, Grebnev G, Edwards RJ, Steinmetz MO, Meiselbach

H, Diella F, Gibson TJ (2012) ELM--the database of eukaryotic linear motifs. *Nucleic Acids Res*, **40:** D242–51

Dong X, Biswas A, Suel KE, Jackson LK, Martinez R, Gu H, Chook YM (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature*, **458:** 1136–1141

Drivas GT, Shih A, Coutavas E, Rush MG, D'Eustachio P (1990) Characterization of four novel ras-like genes expressed in a human teratocarcinoma cell line. *Mol Cell Biol*, **10:** 1793–1798

Engel K, Kotlyarov A, Gaestel M (1998) Leptomycin B-sensitive nuclear export of MAPKAP kinase 2 is regulated by phosphorylation. *EMBO J*, **17:** 3363–3371

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

Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3:** e3647

Englmeier L, Olivo JC, Mattaj IW (1999) Receptor-mediated substrate translocation through the nuclear pore complex without nucleotide triphosphate hydrolysis. *Curr Biol*, **9:** 30–41

Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23:** 257–258

Fantozzi DA, Harootunian AT, Wen W, Taylor SS, Feramisco JR, Tsien RY, Meinkoth JL (1994) Thermostable inhibitor of cAMP-dependent protein kinase enhances the rate of export of the kinase catalytic subunit from the nucleus. *J Biol Chem*, **269:** 2676–2686

Feng W, Benko AL, Lee JH, Stanford DR, Hopper AK (1999) Antagonistic effects of NES and NLS motifs determine S. cerevisiae Rna1p subcellular distribution. *Journal of cell science*, **112:** 339–347

Fischer U, Huber J, Boelens WC, Mattaj IW, Luhrmann 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, Mattaj IW (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, **90:** 1051–1060

Franke WW, Scheer U (1974) Pathways of nucleocytoplasmic translocation of ribonucleoproteins. *Symp Soc Exp Biol*, 249–282

Frey S, Gorlich D (2007) A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell*, **130:** 512–523

Frey S, Gorlich D (2009) FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. *EMBO J*, **28:** 2554–2567

Frey S, Richter RP, Gorlich D (2006) FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science*, **314:** 815–817

Fridell RA, Fischer U, Luhrmann R, Meyer BE, Meinkoth JL, Malim MH, Cullen BR (1996) Amphibian transcription factor IIIA proteins contain a sequence element functionally equivalent to the nuclear export signal of human immunodeficiency virus type 1 Rev. *Proc Natl Acad Sci U S A*, **93:** 2936–2940

Fu SC, Imai K, Horton P (2011) Prediction of leucine-rich nuclear export signal containing proteins with NESsential. *Nucleic acids research*, **39:** e111–e111

Fukuda M, Asano S, Nakamura T, Adachi M, Yoshida M, Yanagida M, Nishida E (1997) CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature*, **390:** 308–311

Fukuda M, Gotoh I, Gotoh Y, Nishida E (1996) Cytoplasmic localization of mitogen-activated protein kinase kinase directed by its NH2-terminal, leucine-rich short amino acid sequence, which acts as a nuclear export signal. *J Biol Chem*, **271:** 20024–20028

Fuxreiter M, Tompa P, Simon I (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23:** 950–956

Gareau JR, Reverter D, Lima CD (2012) Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. *J Biol Chem*, **287:** 4740–4751

Gerace L (1995) Nuclear export signals and the fast track to the cytoplasm. *Cell*, **82:** 341–344

Gontan C, Guttler T, Engelen E, Demmers J, Fornerod M, Grosveld FG, Tibboel D, Gorlich D, Poot RA, Rottier RJ (2009) Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors. *J Cell Biol*, **185:** 27–34

Gorlich D, Dabrowski M, Bischoff FR, Kutay U, Bork P, Hartmann E, Prehn S, Izaurralde E (1997) A novel class of RanGTP binding proteins. *J Cell Biol*, **138:** 65–80

Gorlich D, Kostka S, Kraft R, Dingwall C, Laskey RA, Hartmann E, Prehn S (1995) Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. *Curr Biol*, **5:** 383–392

Gorlich D, Kutay U (1999) Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol*, **15:** 607–660

Gorlich D, Mattaj IW (1996) Nucleocytoplasmic transport. *Science*, **271:** 1513–1518

Gorlich D, Pante N, Kutay U, Aebi U, Bischoff FR (1996) Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J*, **15:** 5584–5594

Gorlich D, Prehn S, Laskey RA, Hartmann E (1994) Isolation of a protein that is essential for the first step of nuclear protein import. *Cell*, **79:** 767–778

Gorlich D, Seewald MJ, Ribbeck K (2003) Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. *EMBO J*, **22:** 1088–1100

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

Güttler T, Madl T, Neumann P, Deichsel D, Corsini L, Monecke T, Ficner R, Sattler M, 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

Harootunian AT, Adams SR, Wen W, Meinkoth JL, Taylor SS, Tsien RY (1993) Movement of the free catalytic subunit of cAMP-dependent protein kinase into and out of the nucleus can be explained by diffusion. *Mol Biol Cell*, **4:** 993–1002

Heerklotz D, Döring P, Bonzelius F, Winkelhaus S, Nover L (2001) The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2. *Molecular and cellular biology*, **21:** 1759–1768

Henderson BR (2000) Nuclear-cytoplasmic shuttling of APC regulates beta-catenin subcellular localization and turnover. *Nat Cell Biol*, **2:** 653–660

Hillig RC, Renault L, Vetter IR, Drell T, Wittinghofer A, Becker J (1999) The crystal structure of rna1p: a new fold for a GTPase-activating protein. *Molecular cell*, **3:** 781–791

Hopper AK, Traglia HM, Dunst RW (1990) The yeast RNA1 gene product necessary for RNA processing is located in the cytosol and apparently excluded from the nucleus. *J Cell Biol*, **111:** 309–321

Huang TT, Kudo N, Yoshida M, Miyamoto S (2000) A nuclear export signal in the N-terminal regulatory domain of IkappaBalpha controls cytoplasmic localization of inactive NF-kappaB/IkappaBalpha complexes. *Proc Natl Acad Sci U S A*, **97:** 1014–1019

Huber J, Cronshagen U, Kadokura M, Marshallsay C, Wada T, Sekine M, Luhrmann R (1998a) Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J*, **17:** 4114–4126

Huber J, Cronshagen U, Kadokura M, Marshallsay C, Wada T, Sekine M, Luhrmann R (1998b) Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J*, **17:** 4114–4126

Hutchison CAr, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M (1978) Mutagenesis at a specific position in a DNA sequence. *J Biol Chem*, **253:** 6551–6560

Iovine MK, Watkins JL, Wente SR (1995) The GLFG repetitive region of the nucleoporin Nup116p interacts with Kap95p, an essential yeast nuclear import factor. *J Cell Biol*, **131:** 1699–1713

Izaurralde E, Mattaj IW (1995) RNA export. *Cell*, **81:** 153–159

Jakel S, Albig W, Kutay U, Bischoff FR, Schwamborn K, Doenecke D, Gorlich D (1999) The importin beta/importin 7 heterodimer is a functional nuclear import receptor for histone H1. *EMBO J*, **18:** 2411–2423

Jakel S, Gorlich D (1998) Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J*, **17:** 4491–4502

Jakel S, Mingot JM, Schwarzmaier P, Hartmann E, Gorlich D (2002) Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *EMBO J*, **21:** 377–386

Johnson C, Van Antwerp D, Hope TJ (1999) An N-terminal nuclear export signal is required for the nucleocytoplasmic shuttling of IkappaBalpha. *EMBO J*, **18:** 6682–6693

Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC (1990) Atomic structure of the actin:DNase I complex. *Nature*, **347:** 37–44

Kaffman A, O'Shea EK (1999) Regulation of nuclear localization: a key to a door. *Annu Rev Cell Dev Biol*, **15:** 291–339

Kahle J, Baake M, Doenecke D, Albig W (2005) Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13. *Mol Cell Biol*, **25:** 5339–5354

Kahle J, Piaia E, Neimanis S, Meisterernst M, Doenecke D (2009) Regulation of nuclear import and export of negative cofactor 2. *J Biol Chem*, **284:** 9382–9393

Kalland KH, Szilvay AM, Langhoff E, Haukenes G (1994) Subcellular distribution of human immunodeficiency virus type 1 Rev and colocalization of Rev with RNA splicing factors in a speckled pattern in the nucleoplasm. *J Virol*, **68:** 1475–1485

Kanehisa M (2013) Molecular network analysis of diseases and drugs in KEGG. *Methods Mol Biol*, **939:** 263–275

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, **28:** 27–30

Kataoka N, Bachorik JL, Dreyfuss G (1999) Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol*, **145:** 1145–1152

Kobayashi T, Kamitani W, Zhang G, Watanabe M, Tomonaga K, Ikuta K (2001) Borna disease virus nucleoprotein requires both nuclear localization and export activities for viral nucleocytoplasmic shuttling. *J Virol*, **75:** 3404–3412

Kose S, Imamoto N, Tachibana T, Shimamoto T, Yoneda Y (1997) Ran-unassisted nuclear migration of a 97-kD component of nuclear pore-targeting complex. *J Cell Biol*, **139:** 841–849

Kosugi S, Hasebe M, Tomita M, Yanagawa H (2008) Nuclear Export Signal Consensus Sequences Defined Using a Localization-Based Yeast Selection System. *Traffic (Copenhagen, Denmark)*, **9:** 2053–2062

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

Kudo N, Matsumori N, Taoka H, Fujiwara D, Schreiner EP, Wolff B, Yoshida M, Horinouchi S (1999a) Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc Natl Acad Sci U S A*, **96:** 9112–9117

Kudo N, Taoka H, Toda T, Yoshida M, Horinouchi S (1999b) A novel nuclear export signal sensitive to oxidative stress in the fission yeast transcription factor Pap1. *J Biol Chem*, **274:** 15151–15158

Kuge S, Arita M, Murayama A, Maeta K, Izawa S, Inoue Y, Nomoto A (2001) Regulation of the yeast Yap1p nuclear export signal is mediated by redox signal-induced reversible disulfide bond formation. *Mol Cell Biol*, **21:** 6139–6150

Kurisaki A, Kurisaki K, Kowanetz M, Sugino H, Yoneda Y, Heldin CH, Moustakas A (2006) The mechanism of nuclear export of Smad3 involves exportin 4 and Ran. *Mol Cell Biol*, **26:** 1318–1332

Kutay U, Bischoff FR, Kostka S, Kraft R, Gorlich D (1997) Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor. *Cell*, **90:** 1061–1071

Kutay U, Lipowsky G, Izaurralde E, Bischoff FR, Schwarzmaier P, Hartmann E, Gorlich D (1998) Identification of a tRNA-specific nuclear export receptor. *Mol Cell*, **1:** 359–369

la Cour T, Gupta R, Rapacki K, Skriver K, Poulsen FM, Brunak S (2003) NESbase version 1.0: a database of nuclear export signals. *Nucleic Acids Res*, **31:** 393–396

la Cour T, Kiemer L, Molgaard A, Gupta R, Skriver K, Brunak S (2004) Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel*, **17:** 527–536

Labokha AA, Gradmann S, Frey S, Hulsmann BB, Urlaub H, Baldus M, Gorlich D (2013) Systematic analysis of barrier-forming FG hydrogels from Xenopus nuclear pore complexes. *EMBO J*, **32:** 204–218

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227:** 680–685

Li W, Yu SW, Kong AN (2006) Nrf2 possesses a redox-sensitive nuclear exporting signal in the Neh5 transactivation domain. *J Biol Chem*, **281:** 27251–27263

Li Y, Yamakita Y, Krug RM (1998) Regulation of a nuclear export signal by an adjacent inhibitory sequence: the effector domain of the influenza virus NS1 protein. *Proceedings of the National Academy of Sciences*, **95:** 4864–4869

Lipowsky G, Bischoff FR, Schwarzmaier P, Kraft R, Kostka S, Hartmann E, Kutay U, Gorlich D (2000) Exportin 4: a mediator of a novel nuclear export pathway in higher eukaryotes. *EMBO J*, **19:** 4362–4371

Madan AP, DeFranco DB (1993) Bidirectional transport of glucocorticoid receptors across the nuclear envelope. *Proc Natl Acad Sci U S A*, **90:** 3588–3592

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

Mandell RB, Feldherr CM (1990) Identification of two HSP70-related Xenopus oocyte proteins that are capable of recycling across the nuclear envelope. *J Cell Biol*, **111:** 1775–1783

Matsuura Y, Stewart M (2004) Structural basis for the assembly of a nuclear export complex. *Nature*, **432:** 872–877

Matunis MJ, Coutavas E, Blobel G (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol*, **135:** 1457–1470

McKinsey TA, Zhang CL, Olson EN (2001) Identification of a signal-responsive nuclear export sequence in class II histone deacetylases. *Mol Cell Biol*, **21:** 6312–6321

Melchior F, Paschal B, Evans J, Gerace L (1993a) Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. *J Cell Biol*, **123:** 1649–1659

Melchior F, Weber K, Gerke V (1993b) A functional homologue of the RNA1 gene product in Schizosaccharomyces pombe: purification, biochemical characterization, and identification of a leucine-rich repeat motif. *Mol Biol Cell*, **4:** 569–581

Meyer BE, Malim MH (1994) The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev*, **8:** 1538–1547

Milburn MV, Tong L, deVos AM, Brunger A, Yamaizumi Z, Nishimura S, Kim SH (1990) Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science*, **247:** 939–945

Mingot JM, Bohnsack MT, Jakle U, Gorlich D (2004) Exportin 7 defines a novel general nuclear export pathway. *EMBO J*, **23:** 3227–3236

Mingot JM, Kostka S, Kraft R, Hartmann E, Gorlich D (2001) Importin 13: a novel mediator of nuclear import and export. *EMBO J*, **20:** 3685–3694

Mohr D, Frey S, Fischer T, Guttler T, Gorlich D (2009) Characterisation of the passive permeability barrier of nuclear pore complexes. *EMBO J*, **28:** 2541–2553

Monecke T, Guttler T, Neumann P, Dickmanns A, Gorlich D, 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, Dickmanns A, Ficner R (2013) Structural basis for cooperativity of CRM1 export complex formation. *Proc Natl Acad Sci U S A*, **110:** 960–965

Mosammaparast N, Jackson KR, Guo Y, Brame CJ, Shabanowitz J, Hunt DF, Pemberton LF (2001) Nuclear import of histone H2A and H2B is mediated by a network of karyopherins. *J Cell Biol*, **153:** 251–262

Muhlhausser P, Muller EC, Otto A, Kutay U (2001) Multiple pathways contribute to nuclear import of core histones. *EMBO Rep*, **2:** 690–696

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, **51 Pt 1:** 263–273

Munoz-Fontela C, Collado M, Rodriguez E, Garcia MA, Alvarez-Barrientos A, Arroyo J, Nombela C, Rivas C (2005) Identification of a nuclear export signal in the KSHV latent protein LANA2 mediating its export from the nucleus. *Exp Cell Res*, **311:** 96–105

Nachury MV, Weis K (1999) The direction of transport through the nuclear pore can be inverted. *Proc Natl Acad Sci U S A*, **96:** 9622–9627

Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*, **7:** 548

Nakielny S, Dreyfuss G (1998) Import and export of the nuclear protein import receptor transportin by a mechanism independent of GTP hydrolysis. *Curr Biol*, **8:** 89–95

Nelson DE, Laman H (2011) A Competitive binding mechanism between Skp1 and exportin 1 (CRM1) controls the localization of a subset of F-box proteins. *J Biol Chem*, **286:** 19804–19815

Neuhoff V, Arold N, Taube D, Ehrhardt W (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, **9:** 255–262

Neville M, Stutz F, Lee L, Davis LI, Rosbash M (1997) The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr Biol*, **7:** 767–775

Nishi K, Yoshida M, Fujiwara D, Nishikawa M, Horinouchi S, Beppu T (1994) Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. *J Biol Chem*, **269:** 6320–6324

O'Neill RE, Talon J, Palese P (1998) The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO J*, **17:** 288–296

Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell*, **101:** 187–198

Ohtsubo M, Okazaki H, Nishimoto T (1989) The RCC1 protein, a regulator for the onset of chromosome condensation locates in the nucleus and binds to DNA. *J Cell Biol*, **109:** 1389–1397

Oliver FW (1913) Makers of British botany. *London: Cambridge University Press*,

Ong SE, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc*, **1:** 2650–2660

Ori A, Banterle N, Iskar M, Andres-Pons A, Escher C, Khanh Bui H, Sparks L, Solis-Mezarino V, Rinner O, Bork P, Lemke EA, Beck M (2013) Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. *Mol Syst Biol*, **9:** 648

Ossareh-Nazari B, Bachelerie F, Dargemont C (1997) Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science*, **278:** 141–144

Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*, **4:** 2411–2423

Paine PL, Moore LC, Horowitz SB (1975) Nuclear envelope permeability. *Nature*, **254:** 109–114

Pante N, Kann M (2002) Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol Biol Cell*, **13:** 425–434

Paraskeva E, Izaurralde E, Bischoff FR, Huber J, Kutay U, Hartmann E, Luhrmann R, Gorlich D (1999) CRM1-mediated recycling of snurportin 1 to the cytoplasm. *J Cell Biol*, **145:** 255–264

Partridge JR, Schwartz TU (2009) Crystallographic and biochemical analysis of the Ran-binding zinc finger domain. *J Mol Biol*, **391:** 375–389

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

Pinol-Roma S, Dreyfuss G (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355:** 730–732

Plafker K, Macara IG (2000a) Facilitated nucleocytoplasmic shuttling of the Ran binding protein RanBP1. *Mol Cell Biol*, **20:** 3510–3521

Plafker SM, Macara IG (2000b) Importin-11, a nuclear import receptor for the ubiquitin-conjugating enzyme, UbcM2. *EMBO J*, **19:** 5502–5513

Plafker SM, Macara IG (2002) Ribosomal protein L12 uses a distinct nuclear import pathway mediated by importin 11. *Mol Cell Biol*, **22:** 1266–1275

Pollard VW, Michael WM, Nakielny S, Siomi MC, Wang F, Dreyfuss G (1996) A novel receptor-mediated nuclear protein import pathway. *Cell*, **86:** 985–994

Radu A, Moore MS, Blobel G (1995) The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. *Cell*, **81:** 215–222

Rawlinson SM, Pryor MJ, Wright PJ, Jans DA (2009) CRM1-mediated nuclear export of dengue virus RNA polymerase NS5 modulates interleukin-8 induction and virus production. *J Biol Chem*, **284:** 15589–15597

Reichelt R, Holzenburg A, Buhle ELJ, Jarnik M, Engel A, Aebi U (1990) Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *J Cell Biol*, **110:** 883–894

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

Ribbeck K, Gorlich D (2001) Kinetic analysis of translocation through nuclear pore complexes. *EMBO J*, **20:** 1320–1330

Ribbeck K, Kutay U, Paraskeva E, Gorlich D (1999) The translocation of transportin-cargo complexes through nuclear pores is independent of both Ran and energy. *Curr Biol*, **9:** 47–50

Ribbeck K, Lipowsky G, Kent HM, Stewart M, Gorlich D (1998) NTF2 mediates nuclear import of Ran. *EMBO J*, **17:** 6587–6598

Richards SA, Lounsbury KM, Carey KL, Macara IG (1996) A nuclear export signal is essential for the cytosolic localization of the Ran binding protein, RanBP1. *J Cell Biol*, **134:** 1157–1168

Rout MP, Blobel G (1993) Isolation of the yeast nuclear pore complex. *J Cell Biol*, **123:** 771–783

Saitoh H, Pu R, Cavenagh M, Dasso M (1997) RanBP2 associates with Ubc9p and a modified form of RanGAP1. *Proc Natl Acad Sci U S A*, **94:** 3736–3741

Sambrook J, Russell DW. (2001) *Molecular Cloning: A Laboratory Manual, Third Edition (3 volume set)*. Cold Spring Harbor Laboratory Press,

Scheffzek K, Klebe C, Fritz-Wolf K, Kabsch W, Wittinghofer A (1995) Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. *Nature*, **374:** 378–381

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*, **95:** 5857–5864

Schwartz TU (2005) Modularity within the architecture of the nuclear pore complex. *Curr Opin Struct Biol*, **15:** 221–226

Schwoebel ED, Talcott B, Cushman I, Moore MS (1998) Ran-dependent signal-mediated nuclear import does not require GTP hydrolysis by Ran. *J Biol Chem*, **273:** 35170–35175

Seewald MJ, Korner C, Wittinghofer A, Vetter IR (2002) RanGAP mediates GTP hydrolysis without an arginine finger. *Nature*, **415:** 662–666

Seimiya H, Sawada H, Muramatsu Y, Shimizu M, Ohko K, Yamane K, Tsuruo T (2000) Involvement of 14-3-3 proteins in nuclear localization of telomerase. *The EMBO journal*, **19:** 2652–2661

Seiser RM, Sundberg AE, Wollam BJ, Zobel-Thropp P, Baldwin K, Spector MD, Lycan DE (2006) Ltv1 is required for efficient nuclear export of the ribosomal small subunit in Saccharomyces cerevisiae. *Genetics*, **174:** 679–691

Smillie DA, Sommerville J (2002) RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. *J Cell Sci*, **115:** 395–407

Stevens BJ, Swift H (1966) RNA transport from nucleus to cytoplasm in Chironomus salivary glands. *J Cell Biol*, **31:** 55–77

Stewart M (2003) Structural biology. Nuclear trafficking. *Science*, **302:** 1513–1514

Stommel JM, Marchenko ND, Jimenez GS, Moll UM, Hope TJ, Wahl GM (1999a) A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. *The EMBO journal*, **18:** 1660–1672

Stommel JM, Marchenko ND, Jimenez GS, Moll UM, Hope TJ, Wahl GM (1999b) A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. *EMBO J*, **18:** 1660–1672

Stuven T, Hartmann E, Gorlich D (2003) Exportin 6: a novel nuclear export receptor that is specific for profilin.actin complexes. *EMBO J*, **22:** 5928–5940

Subramanian K, Meyer T (1997) Calcium-induced restructuring of nuclear envelope and endoplasmic reticulum calcium stores. *Cell*, **89:** 963–971

Team RD (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2007.

Thakar K, Karaca S, Port SA, Urlaub H, Kehlenbach RH (2013) Identification of CRM1-dependent Nuclear Export Cargos Using Quantitative Mass Spectrometry. *Mol Cell Proteomics*, **12:** 664–678

Thomas F, Kutay U (2003) Biogenesis and nuclear export of ribosomal subunits in higher eukaryotes depend on the CRM1 export pathway. *J Cell Sci*, **116:** 2409–2419

Trotta CR, Lund E, Kahan L, Johnson AW, Dahlberg JE (2003) Coordinated nuclear export of 60S ribosomal subunits and NMD3 in vertebrates. *EMBO J*, **22:** 2841–2851

Truant R, Cullen BR (1999) The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol Cell Biol*, **19:** 1210–1217

Truant R, Kang Y, Cullen BR (1999) The human tap nuclear RNA export factor contains a novel transportin-dependent nuclear localization signal that lacks nuclear export signal function. *J Biol Chem*, **274:** 32167–32171

van Hengel J, Vanhoenacker P, Staes K, van Roy F (1999) Nuclear localization of the p120(ctn) Armadillo-like catenin is counteracted by a nuclear export signal and by E-cadherin expression. *Proc Natl Acad Sci U S A*, **96:** 7980–7985

Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A (1999a) Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell*, **97:** 635–646

Vetter IR, Nowak C, Nishimoto T, Kuhlmann J, Wittinghofer A (1999b) Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature*, **398:** 39–46

Vielhaber EL, Duricka D, Ullman KS, Virshup DM (2001) Nuclear export of mammalian PERIOD proteins. *J Biol Chem*, **276:** 45921–45927

Wada A, Fukuda M, Mishima M, Nishida E (1998) Nuclear export of actin: a novel mechanism regulating the subcellular localization of a major cytoskeletal protein. *EMBO J*, **17:** 1635–1641

Wahlberg E, Lendel C, Helgstrand M, Allard P, Dincbas-Renqvist V, Hedqvist A, Berglund H, Nygren PA, Hard T (2003) An affibody in complex with a target protein: structure and coupled folding. *Proc Natl Acad Sci U S A*, **100:** 3185–3190

Walker P, Doenecke D, Kahle J (2009) Importin 13 mediates nuclear import of histone fold-containing chromatin accessibility complex heterodimers. *J Biol Chem*, **284:** 11652–11662

Watson ML (1959) Further observations on the nuclear envelope of the animal cell. *J Biophys Biochem Cytol*, **6:** 147–156

Weinmann L, Hock J, Ivacevic T, Ohrt T, Mutze J, Schwille P, Kremmer E, Benes V, Urlaub H, Meister G (2009) Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, **136:** 496–507

Wen W, Harootunian AT, Adams SR, Feramisco J, Tsien RY, Meinkoth JL, Taylor SS (1994) Heat-stable inhibitors of cAMP-dependent protein kinase carry a nuclear export signal. *J Biol Chem*, **269:** 32214–32220

Wen W, Meinkoth JL, Tsien RY, Taylor SS (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell*, **82:** 463–473

Whittaker GR, Helenius A (1998) Nuclear import and export of viruses and virus genomes. *Virology*, **246:** 1–23

Wilm M (2009) Quantitative proteomics in biological research. *Proteomics*, **9:** 4590–4605

Xu D, Grishin NV, Chook YM (2012a) NESdb: a database of NES-containing CRM1 cargos. *Mol Biol Cell*,

Xu D, Farmer A, Collett G, Grishin NV, Chook YM (2012b) Sequence and structural analyses of nuclear export signals in the NESdb database. *Molecular biology of the cell*,

Yan C, Lee LH, Davis LI (1998) Crm1p mediates regulated nuclear export of a yeast AP-1-like transcription factor. *EMBO J*, **17:** 7416–7429

Yang J, Bardes ES, Moore JD, Brennan J, Powers MA, Kornbluth S (1998) Control of cyclin B1 localization through regulated binding of the nuclear export factor CRM1. *Genes Dev*, **12:** 2131–2143

Yang JM, Wang AH (2004) Engineering a thermostable protein with two DNA-binding domains using the hyperthermophile protein Sac7d. *J Biomol Struct Dyn*, **21:** 513–526

Yokoyama N, Hayashi N, Seki T, Pante N, Ohba T, Nishii K, Kuma K, Hayashida T, Miyata T, Aebi U, et a (1995) A giant nucleopore protein that binds Ran/TC4. *Nature*, **376:** 184–188

Zeitler B, Weis K (2004) The FG-repeat asymmetry of the nuclear pore complex is dispensable for bulk nucleocytoplasmic transport in vivo. *J Cell Biol*, **167:** 583–590

Zhang Y, Xiong Y (2001) A p53 amino-terminal nuclear export signal inhibited by DNA damage-induced phosphorylation. *Science*, **292:** 1910–1915