The CAN Protein
A Mediator of Nucleocytoplasmic Transport with Oncogenic Properties

Het CAN eiwit
Een Schakel in Nucleocytoplasmatisch Transport met Oncogene Eigenschappen

PROEFSCHRIFT

Ter verkrijging van de graad van doctor
Aan de Erasmus Universiteit Rotterdam
Op gezag van Rector Magnificus
Prof. Dr P.W.C. Akkermans, M.A.
En volgens het besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
Woensdag 8 januari 1997 om 15.45 uur

Door

Maarten Willem Jan Fornerod
Geboren te Delft

Promotiecommissie

Promotor:	**Prof. Dr D. Bootsma**
Overige leden:	**Dr G.C. Grosveld (tevens co-promotor)**
	Prof. Dr A.J. van der Eb
	Prof. Dr F.G. Grosveld
	Prof. Dr B. Löwenberg

The studies described in this thesis were performed in (i) the Department of Genetics, St. Jude Children's Research Hospital, Memphis, U.S.A., supported by Cancer CORE Grant CA-21765, and by the Associated Lebanese Syrian American Charities (ALSAC) of St. Jude Children's Research Hospital, and (ii) the Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands, supported by the Dutch Cancer Society, which also supported the printing of this thesis.

Dit proefschrift is bewerkt binnen (i) de afdeling Genetica van St. Jude Children's Research Hospital, Memphis, U.S.A., gesteund door Cancer CORE Grant CA-21765, en door de Associated Lebanese Syrian American Charities (ALSAC) van St. Jude Children's Research Hospital, en (ii) de vakgroep Celbiologie en Genetica van de Faculteit der Geneeskunde en Gezondheidswetenschappen van de Erasmus Universiteit Rotterdam, gesteund door de Nederlandse Kankerbestrijding - Koningin Wilhelmina Fonds, die ook een deel van de drukkosten van dit proefschrift voor haar rekening nam.

CIP GEGEVENS KONINKLIJKE BIBLIOTHEEK DEN HAAG

Fornerod, Maarten
Het CAN Eiwit: Een Schakel in Nucleocytoplasmatisch Transport met Oncogene Eigenschappen / Maarten Fornerod. - Delft: Maarten Fornerod, 1996.
ISBN: 90-9010093-8.

Cover: Sequential sections (1 μm) through a 4 day old mouse blastocyst showing murine CRM1 protein. Sections run left to right, top to bottom. For details, see chapter 6.

CONTENTS

1.	Outline of this Thesis	5
2.	Nucleocytoplasmic Transport and Growth Control	7
3.	The translocation (6;9) associated with a specific subtype of Acute Myeloid Leukemia, results in the fusion of two genes, *dek* and *can*, and the expression of a chimeric, leukemia-specific *dek-can* mRNA. Marieke von Lindern, Maarten Fornerod, Sjozèf van Baal, Martine Jaeglé, Ton de Wit, Arjan Buijs and Gerard Grosveld (1992). *Molecular and Cellular Biology* **12**, 1687-1697.	35
4.	Relocation of the carboxyterminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. Maarten Fornerod, Judith Boer, Sjozèf van Baal, Martine Jaeglé, Marieke von Lindern, K. Gopal Murti, Donna Davis, Jacqueline Bonten, Arjan Buijs and Gerard Grosveld (1995). *Oncogene* **10**, 1739-1748.	47
5.	Interaction of cellular proteins with the leukemia specific fusion protein DEK-CAN and SET-CAN and their normal counterpart, the nucleoporin CAN. Maarten Fornerod, Judith Boer, Sjozèf van Baal, Hans Morreau and Gerard Grosveld (1996). *Oncogene* **13**, 1801-1808.	59
6.	The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. Maarten Fornerod, Jan van Deursen, Sjozèf van Baal, Donna Davis, K.Gopal Murti, Jack Fransen, and Gerard Grosveld. *EMBO Journal*, in press.	77
7.	Chromosomal localization of genes encoding CAN/Nup214-interacting proteins — *hCRM1* localizes to 2p16, whereas *Nup88* localizes to 17p13 and is physically linked to *SF2p32*. Maarten Fornerod, Sjozef van Baal, Virginia Valentine, David N. Shapiro, and Gerard Grosveld. Submitted for publication.	101
8.	Cre-mediated site specific translocation between nonhomologous mouse chromosomes. Jan van Deursen, Maarten Fornerod, Bas van Rees and Gerard Grosveld (1995). *Proceedings of the National Academy of Sciences USA* **92**, 7376-7380.	109
9.	Summary and Discussion	115

CONTENTS (Continued)

Samenvatting	125
Curriculum Vitae	129
Publications	130
Epilogue	131
Appendix 1 Nucleotide sequence for Nup88 major mRNA	134
Appendix 2 Nucleotide sequence for Nup88 alternative 3' UTR	136
Appendix 3 Nucleotide sequence for hCRM1 mRNA	137

***

Chapter 1: Outline of this Thesis

The starting point of the work described in this thesis was a novel gene, *CAN*, that had been cloned by virtue of its involvement in translocation (6;9) (von Lindern, 1990), a recurrent chromosomal aberration defining a specific subtype of acute myeloid leukemia. This gene's predicted amino acid sequence showed no homology to known proteins (chapter 3), and therefore its cellular function was unknown. The protein encoded by CAN's fusion partner in the (6;9) translocation, DEK, also had no significant homology to proteins in the database. Consequently, the role of the chimeric DEK-CAN fusion protein in t(6;9)-associated leukemogenesis was a complete mystery. In addition, *in vitro* studies to show the oncogenic function of DEK-CAN consistently proved unsuccessful (M. von Lindern, J. Boer, G. Grosveld, unpublished results). To gain understanding of the possible function of DEK-CAN, as well as the normal cellular functions of DEK and CAN, we analyzed their primary amino acid structures (chapter 3 and 4), studied their subcellular localization (chapter 3), and looked for interacting proteins (chapter 5 and 6). In these studies, we focused on the CAN protein, since this protein had also been found in another leukemia-related fusion protein, SET-CAN (von Lindern, 1992), which suggests that it could play a more general role in leukemogenesis. In view of the difficulty reproducing the oncogenic effect of the *DEK-CAN* gene *in vitro*, and toxic effects caused by overexpression of *DEK-CAN*, we also set out to create an *in vivo* system to mimic translocation (6;9) in transgenic mice using Cre-mediated recombination (chapter 8).

It appeared that DEK and SET were nuclear proteins, whereas CAN was located at the nuclear pore complex (NPC) (chapter 4). The latter observation was consistent with the presence of a long and degenerate repeat region in the C-terminal part of CAN that resembled a domain in a group of NPC proteins named nucleoporins (chapter 4). Independently, similar discoveries concerning CAN were made in the lab of Gunter Blobel (Kraemer, 1994). The DEK-CAN and SET-CAN proteins appeared to be nuclear, suggesting that relocation of the C-terminal part of CAN from the NPC to the nucleoplasm could be instrumental for DEK-CAN and SET-CAN function. Next we reasoned that identification (chapter 5) and characterization (chapter 6) of CAN-interacting proteins could shed light on the mechanism by which a nucleoporin could be involved in leukemogenesis. This question became especially compelling when two additional nuclear pore components were found to be targeted by oncogenic chromosome translocations: Nup98 (Borrow, 1996; Nakamura, 1996) and Tpr (Byrd, 1994) (chapter 2). In addition, since concurrent studies showed that CAN has an essential role in nucleocytoplasmic transport (van Deursen, 1996), study of CAN-interacting proteins could give further insights into biochemical aspects of this transport function (chapter 6). Indeed, the finding of two novel proteins that specifically interact with CAN, one of which also interacts with DEK-CAN and SET-CAN has added new and useful information in the fast moving field

of nucleocytoplasmic transport (at least according to one anonymous referee), and could open new directions in the search for the molecular actions of nucleoporin-derived fusion proteins associated with leukemogenesis (chapter 7 and 9).

References

Borrow, J., Shearman, A.M., Stanton, V.P., Jr., Becher, R., Collins, T., Williams, A.J., Dube, I., Katz, F., Kwong, Y.L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J. & Housman, D.E. (1996). *Nat. Genet.*, **12**, 159-167.

Byrd, D.A., Sweet, D.J., Panté, N., Konstantinov, K.N., Guan, T., Saphire, A.C., Mitchell, P.J., Cooper, C.S., Aebi, U. & Gerace, L. (1994). *J. Cell Biol.*, **127**, 1515-1526.

Kraemer, D., Wozniak, R.W., Blobel, G. and Radu, A. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1519-1523.

Nakamura, T., Largaespada, D.A., Lee, M.P., Johnson, L.A., Ohyashiki, K., Toyama, K., Chen, S.J., Willman, C.L., Chen, I.M., Feinberg, A.P., Jenkins, N.A., Copeland, N.G. & Shaughnessy, J.D., Jr. (1996). *Nat. Genet.*, **12**, 154-158.

van Deursen, J., Boer, J., Kasper, L. and Grosveld, G. (1996). *EMBO J.*, in press.

von Lindern, M., Poustka, A., Lerach, H. & Grosveld, G. (1990). *Mol. Cell Biol.*, **10**, 4016-4026.

von Lindern, M., van Baal, S., Wiegant, J., Raap, A., Hagemeijer, A. and Grosveld, G. (1992). *Mol. Cell. Biol.*, **12**, 3346-3355.

Chapter 2

NUCLEOCYTOPLASMIC TRANSPORT AND GROWTH CONTROL

I.	**Introduction**	**8**
II.	**Structure and function of the nuclear envelope and nuclear pore complex**	**9**
	The nuclear envelope	**9**
	The nuclear pore complex	**9**
	Mechanisms of nucleocytoplasmic transport	**10**
	Protein import	**11**
	RNA export	**14**
III.	**Regulated nucleocytoplasmic transport**	**16**
	Selective regulation	**16**
	NF-κB/Rel	**16**
	Steroid hormone receptors	**16**
	STATs	**17**
	Cell cycle-dependent nuclear import	**18**
	RNA export	**19**
	Regulation of the transport machinery	**20**
IV.	**Nucleocytoplasmic transport factors and NPC components implicated in neoplastic transformation**	**20**
	Transport factors	**20**
	Oho31/Pendulin	**20**
	NPC components	**21**
	Tpr	**21**
	Nup98	**23**
	Can/Nup214	**24**
V.	**Perspectives**	**24**
	References	**25**

I. Introduction

Multicellular organisms depend on their ability to control cell proliferation and differentiation. This control is established by a multitude of signals that all somehow feed into the nucleus of the individual cell and determine its genetic program. The execution of these programs forces the cell to divide, differentiate, or die.

Errors in the cell's genetic program can be lethal or without effect. Alternatively, they may cause the cell to escape the tight control over its division, differentiation or death, leading to uncontrolled proliferation and eventually cancer.

Signals arriving at the cell surface must cross two distinct physical barriers to reach the nucleus: the plasma membrane and the nuclear envelope. Some agents, such as steroid hormones, cross the plasma membrane and deliver their signal directly to intracellular receptors. Other agents, such as peptide hormones, growth factors, neurotransmitters, and extracellular matrix proteins cannot enter the cell, and bind receptors at the plasma membrane. These receptors, in turn, transmit the signal into the cytosol. Much progress has been made in understanding how this initial signal crosses the cytoplasm via different types of second messengers, protein kinases. Similarly, an impressive body of work describes the effects within the nucleus, where the signal can alter the genetic program of the cell by influencing gene expression. It is not surprising, therefore, that the majority of proto-oncogenes are signal transducers or transcriptional regulators. How the signal crosses the nuclear envelope is less well understood. And possibly because of this lack of information, not many nuclear envelope components or nucleocytoplasmic transport factors have been implicated in neoplastic transformation. In this chapter I discuss several aspects of macromolecular transport across the nuclear envelope and cellular activity, and possible links between the two. First, I will briefly review what is known about the structure of the nuclear envelope and nuclear pore complex (NPC), and in some more detail describe the general mechanisms of nucleocytoplasmic transport known to date. Second, I will address the question if nucleocytoplasmic transport is subject to regulation, and how this regulation is achieved. Third, I will discuss the evidence for the hypothesis that regulation of nucleocytoplasmic transport is a mechanism for growth control, and compare it to the alternative hypothesis that the rate of transport is merely a function of the cell's metabolic state. Finally, I will describe instances where mutation of nucleocytoplasmic transport factors or NPC components are implicated in deregulation of cellular growth, and discuss whether this deregulation is related to changes in nucleocytoplasmic transport, or in unrelated cellular processes such as transciptional control or signal transduction.

II. Structure and function of the nuclear envelope and nuclear pore complex

The nuclear envelope

The nuclear envelope is a multi-layer, semi-permeable structure that surrounds the nucleoplasm. It consists of two lipid layers, the outer and inner nuclear membrane, and a filamentous network that is anchored to the inner nuclear membrane, named the nuclear lamina. The outer nuclear membrane is continuous with the endoplasmic reticulum (ER) membrane, and the lumen of the double nuclear membrane structure (the perinuclear space) feeds into the lumen of the ER. Consequently, there is extensive exchange between nuclear membrane and ER components (Pathak *et al.*, 1986; Powell and Burke, 1990). The outer membrane is a major site of membrane and secretory protein synthesis. The nuclear envelope interacts with various cytoskeletal elements , including actin filaments, microtubles, and intermediate filaments, via integral nuclear membrane proteins(reviewed in Gerace and Burke, 1988). On the inside of the nuclear envelope, proteins of the nuclear lamina contact so-called scaffold attachment regions of the DNA. These regions lie within highly transcribed portions of the genome. Thus, the major sites of transcription and translation are often only separated by the thickness of the nuclear envelope, approximately 100 nm.

The nuclear pore complex

The most notable structures of the nuclear envelope are the nuclear pore complexes (NPCs), which mediate bidirectional nucleocytoplasmic transport (Figure 1). Depending on the physiological state of a cell, the number of pores may vary from 500 to 12,000 per nucleus (Maul, 1977). NPCs have been studied extensively by electron and atomic force microscopy (Hinshaw *et al.*, 1992; Oberleithner *et al.*, 1994), and these studies have revealed that the NPC consists of a central particle, sandwiched between a cytoplasmic and nuclear octagonal ring (Figure 1). The central particle contains a central plug and eight radiating spokes, whereas the outer ring projects various filaments into the cytoplasm. The inner octagonal ring projects interconnected filaments into the nucleoplasm that form a so-called nuclear basket (Figure 1). Although many of the approximately 100 different proteins that make up the ~125 megadalton NPC have been cloned and partially characterized, the mechanism by which this structure operates in nucleocytoplasmic transport is not yet known. NPC proteins that have been identified in yeast or metazoan cells, generally named nucleoporins, include proteins that are anchored to the nuclear membrane and proteins that are present at the periphery of the NPC (recently reviewed by Panté and Aebi, 1994). Many nucleoporins contain multiple copies of a short amino acid motif that always includes the dipeptide phenylalanine-glycine (FG). These repeats have been divided into two types: GLFG and FXFG, based on additional flanking amino acids. Functional differences between these two types have yet to be shown. Most studies addressing the role of individual NPC components have been

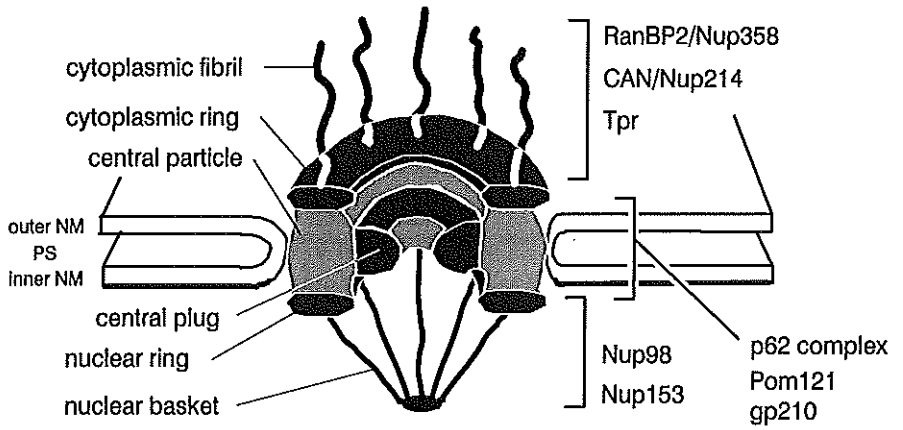


Figure 1 Schematic diagram summarizing the major structural components of the NPC, together with the localization of eight vertebrate nucleoporins. The major structural components of the NPC include the central particle, the central plug or central channel complex, the cytoplasmic ring with protruding fibrils, and the nuclear ring and nuclear basket. Positions where nucleoporins have been localized are indicated on the right. RanBP2/Nup358, CAN/Nup214 and Tpr have been immunolocalized at the cytoplasmic face of the NPC, possibly residing on the cytoplasmic fibrils, Nup153 and Nup98 have been immunolocalized at the nuclear face, while p62 epitopes were located at or near both cytoplasmic and nuclear faces of the central plug. gp210 and Pom121 are transmembrane proteins, and therefore thought to reside more towards the lateral periphery of the complex. NM, nuclear membrane; PS, perinuclear space.

conducted in yeast via gene inactivation approaches. Such approaches have shown that depletion of specific nucleoporins causes defects in nuclear protein import, mRNA export, or both (reviewed in Izaurralde and Mattaj, 1995; Panté and Aebi, 1996; Simos and Hurt, 1995), as well as defects in pre-tRNA splicing (Sharma *et al.*, 1996), and NPC morphology (Wente and Blobel, 1993; Wente and Blobel, 1994). In vertebrates, several nucleoporins have been identified, and sublocalized on the NPC (see Figure 1)(Panté *et al.*, 1994). Very recently, some important biochemical links have been established between factors involved in protein import or RNP export, and NPC components. However, the implications of these findings are still very much in debate.

Mechanisms of nucleocytoplasmic transport

The nuclear pore acts as a molecular sieve, allowing diffusion of molecules smaller than 30-60 kilodaltons (kDa)(Lang *et al.*, 1986; Paine *et al.*, 1975). The degree of freedom of this diffusion

remains to be determined, because many studies have shown a significant concentration difference between the nucleus and the cytoplasm of various ions and other small molecules (reviewed in Csermely *et al.*, 1995). Active nucleocytoplasmic transport of proteins and RNA complexes is a guided and energy consuming process that involves three major steps: recognition of the molecules to be transported, transfer through the nuclear pore, and release into the cytoplasm or nucleoplasm.

Protein import Proteins that are too big to diffuse freely through the NPC, require a nuclear localization signal (NLS) to be transported into the nucleus (Garcia-Bustos *et al.*, 1991; Hanover, 1992; Roberts, 1989). Smaller proteins, which have free, but slow, access to the nucleus via diffusion, can also use this signal to facilitate their transport (Breeuwer and Goldfarb, 1990; Pruschy *et al.*, 1994). An NLS usually consists of either one short stretch of basic amino acids, or two small basic clusters, separated by about 10 amino acids (Dingwall and Laskey, 1991). NLS-dependent nuclear uptake can experimentally be divided into two steps: i) NLS-dependent, but energy-independent, binding of the substrate to the NPC, and ii) energy-dependent translocation into the nucleoplasm (Moore and Blobel, 1992; Newmeyer and Forbes, 1988; Richardson *et al.*, 1988). Several factors have been identified that mediate protein import using an *in vitro* system originally described by Adam *et al.* (1990), based on digitonin-permeabilized mammalian cells, exogenous cytosol, ATP, and a (fluorescent) import substrate.

The first step of protein import is mediated by a heterodimeric, cytosolic protein that binds to an NLS via its smaller subunit, named importinα or karyopherinα (Görlich *et al.*, 1994; Moroianu *et al.*, 1995a). This protein docks to the cytoplasmic face of the NPC via its larger subunit, named importinβ, p97, or karyopherinβ (Adam and Adam, 1994; Chi *et al.*, 1995; Görlich *et al.*, 1995; Radu *et al.*, 1995a). Because importinβ binds nucleoporin-specific FG-repeat regions *in vitro* (Iovine *et al.*, 1995; Kraemer *et al.*, 1995; Moroianu *et al.*, 1995b; Radu *et al.*, 1995b), these regions of nucleoporins located at the cytoplasmic face of the NPC are thought to be the docking sites for the import complex. Importinα localizes to both the cytoplasm and the nucleus (Görlich *et al.*, 1996; Imamoto *et al.*, 1995; Weis *et al.*, 1996), whereas importinβ is present in the cytoplasm and at both sites of the NPC but not in the nucleus (Chi *et al.*, 1995; Koepp *et al.*, 1996; Moroianu *et al.*, 1995b). Therefore, importinα and β cannot dimerize in the nucleus, and the heterodimer probably dissociates at the nuclear side of the NPC, releasing its substrate into the nucleoplasm (Figure 2).

The energy-dependent translocation of the import substrate into the nucleoplasm requires the small Ras-like nuclear GTPase Ran and a protein called pp15, p10, or Ntf2p (Görlich *et al.*, 1995; Moore and Blobel, 1994; Nehrbass and Blobel, 1996). Experimentally, translocation of the import substrate to the nucleus can be blocked by a non-hydrolyzable form of GTP (Melchior *et al.*, 1995b), leading to accumulation of Ran at the cytoplasmic face of the NPC. Interestingly, the FG-repeat containing nucleoporin RanBP2/Nup214 is a major RanGTP (not RanGDP) binding component of the nuclear envelope (Melchior *et al.*, 1995a; Wu *et al.*,

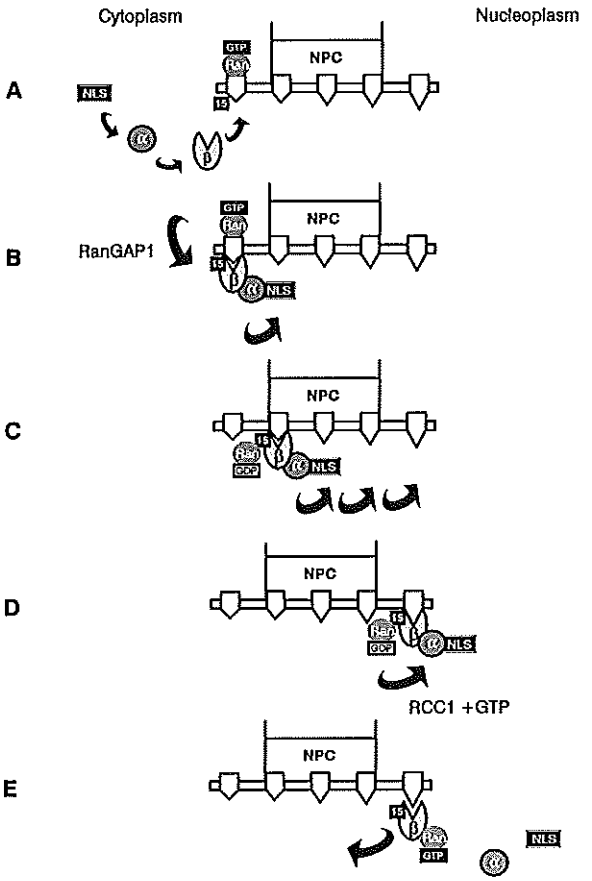


Figure 2 Model of protein import into the nucleus. The NPCs are depicted in side view, and only the upper half is shown. **A**, In the cytoplasm, NLS-containing proteins bind to the α-subunit of importin. The β-subunit mediates binding to the NPC, possibly via nucleoporin repeat regions (protruding triangles). **B**, Initiation of translocation probably involves RanGTP hydrolysis stimulated by RanGAP1. RanGTP binds to nucleoporin RanBP2/Nup358. **C**. Translocation is probably a multistep process, and requires energy, Ran, and pp15. Increasing affinity of nucleoporin repeats towards the nucleoplasm for importinβ (represented by sharpening triangles) may give direction to the process. **D**, Binding of RanGTP to importinα dissociates the import complex. Exchange of GDP to GTP is probably mediated by the Ran exchange factor RCC1. **E**, The NLS-containing protein enters the nucleoplasm, importinβ and RanGTP, possibly in complex, are returned to the cytoplasm. The return of importinα may be export-mediated (see Figure 3). α, importinα; β, importinβ; 15, pp15; NLS, nuclear localization signal containing protein; NPC, nuclear pore complex.

1995), and this protein has been located to the cytoplasmic fibrils of the NPC (Yokoyama *et al.*, 1995). These data suggest that initial docking occurs at or near RanBP2/Nup358 (Figure 2).

RanGTP binds to importinβ and, *in vitro*, dissociates the importinαβ dimer while RanGDP has no effect (Rexach and Blobel, 1995). Since RanGTP concentrations are low in the

cytoplasm and high in the nucleoplasm, it is likely that this dissociation event occurs when importinαβ is part of the import complex at the nuclear side of the NPC (see Figure 2). Ran, like other GTPases, requires a cofactor for GTP hydrolysis (GTPase activating protein, or GAP), and an exchange factor to replace GDP with GTP. The only known guanine exchange factor for Ran is RCC1, which is located exclusively in the nucleus (Bischoff and Ponstingl, 1991) and the only known Ran GTPase activating protein (Ran-GAP1) is located in the cytoplasm (Bischoff *et al.*, 1995). The compartmentilization of these antagonists may be responsible for keeping the RanGTP concentration high in the nucleus and low in the cytoplasm (Figure 2).

The role of pp15 is not clear. It has been shown to bind several nucleoporin repeats, importinβ, and RanGDP, but not RanGDP (Nehrbass and Blobel, 1996). Furthermore, the yeast homologue is located at the nuclear envelope (Koepp *et al.*, 1996; Nehrbass and Blobel, 1996). Therefore a role in Ran shuttling can be imagined (Figure 2).

Although models are tentatively in place for docking of the import complex at the cytoplasmic side of the NPC and of its disassembly at the nucleoplasmic side, what remains controversial are the events that occur in between (Görlich and Mattaj, 1996; Nehrbass and Blobel, 1996; Rexach and Blobel, 1995; Schlenstedt, 1996). The import substrate has to move through approximately 100 nm of channel, and, therefore, the translocation is assumed to require multiple energy-dependent steps. Importinβ has been shown to bind to the FG repeat domains of several nucleoporins *in vitro* (see above), and FG-repeat containing nucleoporins have been localized to different positions along the NPC (Figure 1). Therefore, nucleoporin repeat domains are the prime candidates to act as internal guideposts. Increasing affinity of these repeats for importinβ towards the nucleus could give direction to the translocation process. Since the translocation requires GTP hydrolysis by Ran, this protein is thought to have a central role in this process. On the basis of the *in vitro* dissociation of import complex components by Ran-GTP, it has been postulated that each Ran-GTP cycle causes complete dissociation of the complex, followed by its reassembly on a higher affinity FG-repeat closer to the nucleus (Nehrbass and Blobel, 1996; Rexach and Blobel, 1995). Alternatively, the import complex could be translocated as a single unit that contains multiple FG-repeat binding sites (Görlich and Mattaj, 1996). Ran-provided energy would be used to stepwise detach one binding site from the NPC, which would subsequently bind to a higher affinity repeat closer to the nucleus, leaving the importin-import substrate complex intact. Lastly, it has been proposed that import complexes slide along the FG-repeat-coated inner surface of the NPC, driven by Ran-GTP powered proteins (Schlenstedt, 1996). It is clear that more research is needed before we will know which of these models, if any, is correct.

Although in yeast only one importinα protein is present (Srp1, Yano *et al.*, 1992), in humans at least three importinα proteins have been identified: hSrp1 (Cortes *et al.*, 1994; O'Neil and Palese, 1995) (also named importinα_1, karyopherinα_1, or NPI-1), Rch1 (Cuomo *et al.*, 1994; Weis *et al.*, 1995) (also named importinα_2, hSRPα or karyopherinα_2), and 'hImpαH', an

importinα homologue recently found in the database (Görlich *et al.*, 1996). In addition, a homologue has been described in *Drosophila*: Oho31 , also named Pendulin (Küssel and Frasch, 1995; Török *et al.*, 1995). Several of these proteins function in *in vitro* nuclear protein import assays with importinβ (Görlich *et al.*, 1996a) and their existence may reflect different substrate specificity.

Export of RNA The study of RNA export is still in its infancy. It is not yet known whether different RNA species are transported according to different mechanisms, although, in parallel to their transcription, the general principles could be the same. Our current understanding of RNA export was recently reviewed (Fabre and Hurt, 1994; Gerace, 1995; Izaurralde and Mattaj, 1995), so I will limit myself to just two examples of RNA export from which some mechanistic insights have been obtained: export of human immunodeficiency virus (HIV) RNA and that of U snRNA.

A variety of proteins are encoded by subgenomic mRNAs produced by alternative splicing of the full-length HIV transcript. To ensure propagation of the virus, full-length transcript has to be exported into the cytoplasm in order to be packaged into the retroviral coat. This requires export of intron-containing pre-mRNAs that would not normally leave the nucleus. This export is mediated by the viral Rev protein, which binds to a specific sequence in the intron-containing HIV transcripts, named the Rev response element (RRE). The Rev protein contains a nuclear export signal (NES) that interacts with an FG-repeat-containing cellular protein named hRip or Rab (Bogerd *et al.*, 1995; Fritz *et al.*, 1995). Since hRip localization is still unresolved, it is not yet clear whether Rev docks to hRip at the NPC or hRip itself is the docking component. Because the Rev NES has also been shown to bind to FG-repeat containing segments of bona fide nucleoporins (Stutz *et al.*, 1995), it is conceivable that Rev-containing ribonucleoproteins (RNP) move through the NPC via a series of docking and undocking reactions, similar to those proposed for the importinβ-containing import complex. Microinjection of Rev NES peptide-conjugated BSA into *Xenopus* oocyte nuclei suppresses not only the Rev-mediated export of RRE-containing RNA, but also that of cellular 5S rRNA and U snRNA (Fischer *et al.*, 1995). This suggests that export of these RNAs is mediated by a transport molecule with a Rev-like NES.

Most U snRNAs are transcribed by RNA polymeraseII and acquire a 7-methyl guanine (m7G) cap in the nucleoplasm. This structure is bound by a protein complex, aptly named the cap binding complex (CBC), which consists of two subunits, CPB80 and CBP20 (Izaurralde *et al.*, 1995; Izaurralde *et al.*, 1994). Intranuclear microinjection of antibodies against CBC inhibits export of U snRNA, indicating that CBC is essential function for this export process. Surprisingly, protein import factor importinα binds very strongly to CBC, which suggests that this protein plays an important role in the export of CBC-containing complexes (Görlich *et al.*, 1996b). *In vitro* studies have shown that addition of importinβ to the importinα-CBC-RNA

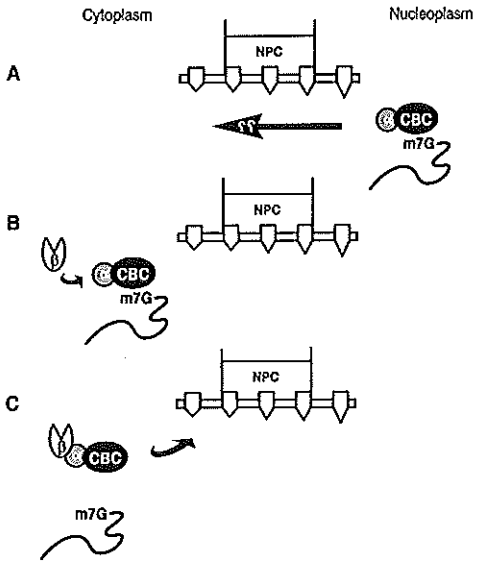


Figure 3 Elements of RNA export from the nucleus. A, Importinα is bound to cap binding complex (CBC) and methylguanine (m7G) capped RNAs in the nucleus. The mechanism of translocation is unknown. B, Importinβ dissociates the importinα-CBC complex from the RNA. C, CBC recycles to the nucleus similar to NLS-containing proteins (see Figure 2). α, importinα; β, importinβ; NPC, nuclear pore complex.

complex results in the release of the U snRNA. Because importinβ binds RanGTP with high affinity (and not RanGDP), and this binding coincides with importinαβ dissociation (Rexach and Blobel, 1995), it suggests that importinβ can bind either RanGTP or importinα, but not both at the same time. Consistent with this, dissociation of the importinα-CBC-RNA complex by importinβ is prevented by RanGTP but not by RanGDP. Thus, in the nucleus the higher concentrations of RanGTP protect the RNA export complex against dissociation by importinβ. In the cytoplasm, where RanGTP concentrations are low, and RanGDP high, importinβ can bind to importinα, resulting in RNA displacement from the complex (Figure 2) (Görlich *et al.*, 1996b). It is not yet known how the importinα-CBC-RNA complex makes contact with the NPC; however, it is tempting to postulate the existence of an importinβ-like protein that mediates this contact via nucleoporin FG-repeats.

In conclusion At least three classes of proteins are mechanistically implicated in nucleocytoplasmic transport: i) factors that contact the transport substrate in the nucleus or cytoplasm and lead it to the NPC, ii) NPC components (nucleoporins) that might guide the substrate through the pore, and iii) proteins involved in the energy supply required for the transport process.

III Regulated nucleocytoplasmic transport

Based on what we know about the transport machinery that mediates macromolecular exchanges between the cytoplasm and the nucleus, two different ways of regulating nucleocytoplasmic

transport are possible: the first is dependent on properties of the trafficking macromolecule, and therefore highly specific, the second is dependent on modifications of the transport machinery, and would therefore have a more general effect on nucleocytoplasmic traffic. I will first discuss the first mode with a number of well defined examples, and continue with the second possible mechanism, for which less direct evidence is available

Selective regulation of nucleocytoplasmic transport

In recent years, specific regulation of protein import has been shown to play a key role in several signal transduction pathways involved in growth regulation. In addition, there are indications that specific regulation of protein and RNA export exist. I will review a few of the best studied examples.

NF-κB/Rel The NF-κB protein is a transcription factor that binds to a specific element in the enhancers of many genes, including those encoding cytokines, H-Ras, and c-Myc (reviewed in Miyamoto and Verma, 1995). The transcription factor is composed of two subunits, p50 and p65 that are encoded by separate genes but share significant sequence homology. The p50 NF-κB subunit can form an active heterodimer with other p65-like partners. One of these is c-Rel, the cellular counterpart of a viral oncoprotein found in certain avian lymphomas (Bose, 1992). This association of NF-κB/Rel proteins in oncogenesis has been substantiated by findings linking increased NF-κB activity to fibroblast transformation (Kitajima *et al.*, 1992), and an NF-κB p50 homologue targeted by a chromosomal translocation in B-cell lymphomas (Neri *et al.*, 1991).

The activity of NF-κB/Rel proteins is regulated, in part, by their subcellular localization. The transcription factors are present in the cytoplasm as an inactive complex that includes an inhibitor named IκB. External or internal signals can lead to the phosphorylation and degradation of IκB and the subsequent nuclear translocation of NF-κB/Rel. The association of IκB with NF-κB p65 or c-Rel requires their intact nuclear localization signals (Beg *et al.*, 1992; Kerr *et al.*, 1991), suggesting that NF-κB/Rel nuclear import is regulated through the masking of its NLS by IκB. In some cells however, after complete degradation of IκB, less than 10-20% of the total NF-κB/Rel proteins translocates into the nucleus immediately (Verma *et al.*, 1995). It has therefore been suggested that additional unexplored regulatory steps exist in the pathway to nuclear localization of Rel/NF-κB.

Steroid hormone receptors Steroid hormones are signaling molecules that travel the blood-stream, enter cells by diffusion through the plasma membrane, and bind to specific intracellular receptors. These receptors are transcription factors that are activated by ligand binding, and subsequently transactivate target genes (reviewed in Tsai and O'Malley, 1994). Members of the receptor superfamily can be divided into two subgroups according to their functional properties. One group includes the thyroid receptor, retinoic acid receptor, 9-cis-retinoic acid receptor, and

vitamin D receptor, which are nuclear proteins that can bind DNA in the absence of ligand. Upon hormone binding, these receptors become transcriptional activators or inhibitors. The second group of steroid receptors includes the glucocorticoid receptor, androgen receptor, progesterone receptor, and estrogen receptor. These receptors form an inactive cytoplasmic complex with the heat-shock proteins hsp90, hsp70, and hsp56 in the absence of hormone. Upon hormone binding, the heat-shock proteins dissociate and the receptors enter the nucleus where they transactivate or repress target genes. These ligand-bound steroid receptors continuously shuttle between the nucleus and the cytoplasm (reviewed in Defranco *et al.*, 1995), which suggests that nucleocytoplasmic transport of this group of receptors is an important regulated step in the steroid hormone signal transduction pathway. Hormone-dependent nuclear import of the glucocorticoid receptor has been demonstrated *in vivo* (e.g. Picard and Yamamoto, 1987; Wikstrom *et al.*, 1987) and *in vitro* (Yang and DeFranco, 1994). It appears to be a fast and energy-dependent process. Therefore, the simplest model to explain the regulated import is a ligand-induced conformational change of the hormone receptor, followed by dissociation from heat-shock proteins. This dissociation would then unmask the hormone receptor's NLS, resulting in fast and energy-dependent nuclear import. The mechanism by which steroid hormone receptors move out of the nucleus is much less clear. Nuclear microinjection of the progesterone receptor NLS, in the context of *E.coli* β-galactosidase demonstrated that nuclear export was NLS-dependent; the normal β-galactosidase protein was unable to exit the nucleus (Guiochon-Mantel *et al.*, 1994). Moreover, the simian virus 40 large T antigen NLS functioned in the same manner, suggesting that NLS-dependent export could be a more general feature of nuclear proteins. Since not all nuclear proteins shuttle between the cytoplasm and the nucleus, an additional level of regulation is likely to exist. This is seemingly in disagreement with the results of Schmidt-Zachmann *et al.*, who showed efficient export of nuclear-injected pyruvate kinase, a cytoplasmic protein, suggesting that nuclear export of protein is a passive process (Schmidt-Zachmann *et al.*, 1993). However, this apparent discrepancy could merely reflect the existence of multiple, differently regulated export pathways.

STATs The Signal Transducers and Activators of Transcription (STATs) were originally identified as components of the interferon signaling pathway. It was observed that after stimulation of the interferon receptor by IFNα, a number of proteins associate on IFNα-responsive promoters. This protein complex, named ISGF3, is composed of p91 (STAT1α), p84 (STAT1β), p113 (STAT2), and p48 (reviewed in Schindler and Darnell, 1995; Taniguchi, 1995). STAT1α and β are encoded by the same gene, whereas STAT2 is about 40% identical to STAT1. Several other closely related family members have been found, named STAT3 through 6. STATs are cytoplasmic proteins in non-stimulated cells, but upon stimulation with a specific growth factor translocate to the nucleus (Dale *et al.*, 1989; Kessler *et al.*, 1990; Levy *et al.*, 1989; Waxman *et al.*, 1995). This nuclear translocation requires formation of STAT hetero- or homo-dimers, which is triggered by phosphorylation by a specific kinase of the JAK (Janus

kinase) family (e.g. Taniguchi, 1995). Several JAKs have been discovered, each involved in a different subset of cytokine signaling. These JAKs are thought to be directly activated by interaction with the ligand-occupied receptor. Why STAT dimers enter the nucleus, whereas STAT monomers remain cytoplasmic, is unknown.

Cell cycle-dependent nuclear import It has been estimated that up to 250 genes are expressed in a cell cycle-dependent manner in *Schizosaccharomyces pombe* (Price *et al.*, 1991). It is, therefore, reasonable to assume that in metazoa this number will even be higher. Correct timing of expression of these genes is essential for normal cell function, and deregulation of cell cycle proteins operating in the Retinoblastoma pathway is a major cause of oncogenic transformation (reviewed in Hall and Peters, 1996). One mechanism of cell cycle controlled gene expression is the regulated nuclear import of transcription factors. The clearest example of this phenomenon is the yeast transcription factor SWI5. SWI5 influences the mating status of yeast through transcriptional control over certain target genes (Moll *et al.*, 1991). It exhibits cell cycle-dependent nuclear exclusion, entering the nucleus specifically in G_1 (Moll *et al.*, 1991; Nasmyth *et al.*, 1990). This nuclear exclusion is mediated via phosphorylation by the *Schizosaccharomyces pombe* cyclin-dependent kinase (CDK) CDC28 of an adjacent CDK site. β-Galactosidase fusion proteins, carrying the SWI5 NLS/CDK sequences, showed a similar cell cycle-dependent nuclear localization, that was both NLS and CDK site-dependent, whereby mutation of the CDK site resulted in constitutive nuclear localization (Jans *et al.*, 1995). This suggests that phosphorylation of the CDK site next to the NLS blocks nuclear import through a charge or conformational effect. In mammalian cells, SWI5 fusion proteins were similarly transported to the nucleus in an NLS and CDK site-dependent fashion, indicating that this regulatory mechanism of protein import is conserved in evolution.

Certain CDK/cyclin complexes are also subject to regulated nuclear entry (reviewed by Gallant *et al.*, 1995). The vertebrate CDK1 $p34^{CDC2}$ interacts with at least four distinct cyclins, A, B1, B2, and B3. CDK/cyclin A and CDK/cyclin B3 complexes accumulate in the nucleus as soon as they form (e.g. Zindy *et al.*, 1992). In contrast, complexes containing cyclins B1 or B2 are cytoplasmic throughout interphase, but then translocate into the nucleus at the onset of mitosis, just prior to nuclear envelope breakdown (Gallant and Nigg, 1992; Pines and Hunter, 1991). The molecular mechanism of this abrupt nuclear translocation at the G_2/M transition is unknown; neither $p34^{CDC2}$ nor cyclins A or B contain an obvious NLS. However, it is conceivable that the nuclear translocation of CDK/cyclin complexes requires a chaperone that mediates their import into the nucleus. In the cyclin B1 protein, a cytoplasmic retention signal has been discovered, which could be a target for such regulated transport (Pines and Hunter, 1994). Why the nuclear translocation of CDK/cyclin complexes is regulated remains uncertain; deletion of the cytoplasmic retention signal of cyclin B1 (Pines and Hunter, 1994) or addition of an SV40 large T NLS to this protein (Gallant *et al.*, 1995) results in its constitutive nuclear localization, without obvious effects on cell cycle progression.

RNA export There are relatively few examples of selective RNA export known to date. The export of unspliced HIV mRNAs is probably the best studied example (discussed in part II). Infection of cells with adenovirus or influenza virus inhibits the export of some (Beltz and Flint, 1979), but not all (Moore *et al.*, 1987) cellular mRNAs, while promoting export of viral messengers. In the case of influenza, the virally encoded M1 protein promotes export of viral ribonucleoproteins from the nucleus (Martin and Helenius, 1991). The viral NS_1 protein binds to poly(A) and inhibits both splicing and export of polyadenylated mRNA (Fortes *et al.*, 1994; Qiu and Krug, 1994). As viral proteins often use cellular regulatory pathways to promote the virus's own interests, it is conceivable that this selective mRNA export is also a feature of normal cells.

Export of cyclin D1 mRNA is specifically enhanced by artificially high expression of the eukaryotic initiation factor 4E (eIF-4E) (Rousseau *et al.*, 1996). eIF-4E binds to the m7G cap of RNA polymerase II transcribed RNAs and is part of the cap-binding complex that plays a major role in the control of translation initiation (reviewed in Mader and Sonenberg, 1995). Because a significant fraction of eIF-4E is present in the nucleus (Lejbkowicz *et al.*, 1992), it has been argued that this protein is also important for the nuclear export of RNAs, in parallel to nuclear cap-binding protein complex CBC (Izaurralde *et al.*, 1995; Izaurralde *et al.*, 1992). Interestingly, overexpression of cyclin D1 in rodent cells causes transformation (Okamoto *et al.*, 1994) and the gene is amplified and overexpressed in many human carcinomas (see Hall and Peters, 1996). It would be interesting to test if exogenous overexpression of eIF-4E could increase the export of cyclin D1 sufficiently to cause enhanced cell growth.

In conclusion Specific regulation of nuclear protein import seems generally to be modulated by the accessibility of the NLS. This can be established by masking of the NLS, as illustrated by the NF-κB and steroid receptor families, or post translational modification of the NLS flanking region, as seen in SWI5. However, in a number of cases the mechanism of regulation is not yet known. There are only a few cases of selectively regulated RNA export known to date, but the common theme seems to be sequence specific binding of an export or retention promoting factor.

Regulation of the transport machinery

The first studies correlating the nucleocytoplasmic transport rate and the growth state of the cell, compared NPC numbers and nuclear accumulation of fluorescent dextran in transformed versus nontransformed fibroblasts. It appeared that both the number of NPCs and the rate of nucleocytoplasmic transport were increased in transformed cells (Jiang and Schindler, 1988). The same effect was seen when non-transformed cells were exposed to peptide growth factors (Jiang and Schindler, 1988). Furthermore, the nuclear transport capacity of proliferating BALB/c

3T3 cells, as measured by the nuclear import rate of nucleoplasmin-coated gold particles of various sizes, was found to decrease upon growth arrest (Feldherr and Akin, 1991). Fusion between proliferating and quiescent cells showed that differences in transport capacity between the two nuclei were maintained, and that the number of goldparticles located at the cytoplasmic faces of the pores was significantly greater in proliferating nuclei, suggesting that qualitative differences between the nuclear pore complexes were important in regulating nuclear transport (Feldherr and Akin, 1993).

These results can be interpreted in two ways: (i) the regulation of the transport machinery at the level of the NPC plays a role in maintaining the quiescent or proliferating state of the cell, or (ii) the number of the NPCs and their functional characteristics are regulated by factors also governing the growth state of the cell. These possibilities are not mutually exclusive; the actions of growth regulating soluble factors may lead for example to contraction or expansion of the NPC, that could serve to quickly limit or increase the export of large ribonucleic particles such as ribosomal subunits and mRNPs.

IV. Involvement of nucleocytoplasmic transport factors and NPC components in neoplastic transformation

Transport factors

Oho31/Pendulin In *Drosophila* larvae, two types of hematopoietic cells function in cellular immunity: the plasmatocytes, which resemble vertebrate macrophages; and the crystal cells, which seem to play a role in blood coagulation in response to wounding. Both cell types originate in the larval hematopoietic organs, from proplasmatocytes and procrystal cells respectively. Mutations in more than 25 genes can cause overgrowth of hematopoietic organs during larval development (reviewed Gateff, 1994). Only a few of these genes have actually been cloned, one example being the $Hop^{Tum\text{-}l}$ allele of *Hopscotch*, which causes plasmatocyte overgrowth followed by terminal differentiation (Binari and Perrimon, 1994). *Hopscotch* encodes a non-receptor tyrosine kinase that shows homology to the vertebrate Janus kinase (JAK) family. In addition, the Hopscotch/JAK kinase interacts genetically with a *Drosophila* protein named Marelle, that shows homology with vertebrate STATs (Hou, 1996, see also part II for a description of the JAK/STAT pathway). This indicates that Hopscotch is part of an invertebrate JAK/STAT system. In vertebrates no transforming mutations in the JAK/STAT pathway have been found to date, however, considering its role in many growth signalling pathways, it seems to be an ideal target for malignant transformation (Ihle, 1995). Another example is *oho31* (*overgrown hematopoietic organs-31*), a recessive gene mutation which was identified in a genetic screen designed to identify *Drosophila* genes that control cell proliferation

and differentiation (Török *et al.*, 1993). Cloning of the gene, and its subsequent mutation by P element insertion, showed that its inactivation caused not only hematopoietic overgrowth, but also the overgrowth of several imaginal organs (Küssel and Frasch, 1995; Török *et al.*, 1995). The Oho31 protein, also named Pendulin, has extensive sequence homology to members of the importinα family: yeast Srp1, *Xenopus* importinα, and mammalian hSrp1 and Rch1 (Küssel and Frasch, 1995; Török *et al.*, 1995). This homology suggests that Oho31 is also involved in nucleocytoplasmic transport. Indeed, the protein has been reported to function in protein import *in vitro* (Görlich *et al.*, 1996a). Interestingly, Oho31 localizes to the cytoplasm but transiently accumulates in the nucleus at the beginning of mitosis. In comparison, yeast importinα Srp1, is constitutive nuclear (Kussel and Frasch, 1995; Yano *et al.*, 1994), whereas *Xenopus* importinα, which is present throughout the cell (Görlich *et al.*, 1996a). Therefore it was postulated that the different importinα proteins in higher eukaryotes may have non-overlapping functions, perhaps serving distinct subsets of NLS- (or NES) containing substrates. It may be that the Oho31 phenotype can be explained as a transport deficiency in such a subset that includes factors for control of growth or differentiation.

Nucleoporins

Three vertebrate NPC proteins have so far been implicated in human cancer or leukemia. Tpr is a NPC component with extensive alpha-helical protein interaction motifs, whereas Nup98 and CAN/Nup214 are FG-repeat-containing nucleoporins.

Tpr The *tpr* gene was originally identified as a part of the *trp-met* oncogene, which results from a translocation between human chromosomes 1 and 7 in gastric tumors (Cooper *et al.*, 1984; Park *et al.*, 1986). The gene encodes a 265 kDa protein that is located at the cytoplasmic face of the NPC (Byrd *et al.*, 1994). Tpr lacks the FG-repeat region that is characteristic for many nucleoporins, but it contains an exceptionally long coiled coil domain (~1600 amino acids), and several leucine zippers, motifs involved in protein-protein interaction. Because of its predicted rod-like shape, it has been suggested that Tpr forms the flexible backbone for the cytoplasmic fibrils of the NPC (see Figure 1B). Its fusion partner in t(1;7), the Met protein, is the hepatocyte growth factor receptor, a transmembrane protein with a cytoplasmic tyrosine kinase domain. The *trp-met* fusion gene is frequently detected in human gastric tumors (Soman *et al.*, 1991), and *c-met* is consistently overexpressed in thyroid papillary carcinoma and tumors of the gastrointestinal tract (Di Renzo *et al.*, 1991; Liu *et al.*, 1992). Thus, amplification of the tyrosine kinase activity of the Met protein appears to be responsible for the transforming property. Tpr has also been found in fusion proteins that contain the receptor tyrosine kinase Trk (Greco *et al.*, 1992), or the cytoplasmic serine/threonine kinase Raf (Ishikawa *et al.*, 1987). In all cases, kinase catalytic domains are fused ~140-230 N-terminal residues of Tpr, leading to the suggestion that the N-terminal region of Tpr mediates dimerization of the kinase domains,

thereby causing kinase activation (Rodrigues and Park, 1993).

This notion seems to be supported by events surrounding an analogous oncogenic fusion protein, NPM-ALK. NPM-ALK is formed as a consequence of the t(2;5) in non-Hodgkin lymphoma and anaplastic large cell lymphoma (Elmberger *et al.*, 1995; Morris *et al.*, 1994). ALK is a receptor tyrosine kinase of unknown function, whereas NPM is a nucleolar protein, also known as B23, that forms oligomers *in vivo* (Chan and Chan, 1995). The NPM-ALK protein has been shown to form such oligomers, apparently by way of its NPM moiety (D. Bischof and S.W. Morris, unpublished results). This finding suggests that NPM oncogenically activates the ALK tyrosine kinase by promoting constitutive multimerization and phosphorylation. In agreement with this idea, a synthetic Tpr-ALK fusion protein also has transforming activity (D. Bischof and S.W. Morris, unpublished results).

Byrd *et al.* have suggested the alternative possibility that the N-terminus of Tpr targets the kinase activity to the NPC, thereby changing the phosphorylation state of NPC components and/or proteins transported through the NPC, such as transcription factors, and in this way promotes the transformed state (Byrd *et al.*, 1994). In this respect it is noteworthy that NPM/B23 is a shuttling molecular chaperone (Borer *et al.*, 1989) that is, therefore, likely to come into frequent contact with the NPC. Currently it is not known whether Tpr-containing fusion proteins are targeted to the NPC, whether the shuttling behavior of NPM/B23 is retained in NPM-ALK, and, if so, whether these attributes are essential for oncogenic activity.

Nup98 The genes for nucleoporin Nup98 and HoxA9 form a fusion gene that encodes a Nup98-HoxA9 fusion protein as a result of the translocation (7;11)(p15;p15)(Borrow *et al.*, 1996; Nakamura *et al.*, 1996a). This translocation is a rare but recurrent karyotypic abnormality mostly found in acute myeloid leukemias that have limited differentiation (Sato *et al.*, 1987). It is also found in Philadelphia chromosome positive or negative chronic myeloid leukemia (Tomiyasu *et al.*, 1982).

Nup98 is a vertebrate nucleoporin with an FXFG-type repeat that is located at the nuclear face of the NPC (Radu *et al.*, 1995b). Its nucleoporin repeat region is present in the Nup98-HoxA9 fusion protein. HoxA9 belongs to a protein family clearly implicated in oncogenesis: the homeobox protein family.

Homeobox proteins are transcription factors that are normally involved in early developmental processes (for review see e.g. Wright, 1991). They are known for their oncogenic potential (reviewed in Stuart *et al.*, 1995), for example, expression of certain Hox genes under the control of the CMV promoter/enhancer leads to transformation of NIH3T3 mouse fibroblasts (Maulbecker and Gruss, 1993). Some Hox genes have increased expression in colorectal cancer (De Vita *et al.*, 1993). More importantly, homeobox genes are frequent targets of cancer-associated chromosomal rearrangements. Translocation (1;19) creates a fusion between Pbx1, a homeobox gene on chromosome 1, and E2A, a gene that encodes the enhancer-binding transcription factors E12 and E47 (Kamps *et al.*, 1990; Nourse *et al.*, 1990). This

translocation is present in 20%-25% of all childhood pre-B cell acute lymphocytic leukemias, and E2A-PBX1 functions experimentally as a potent oncogene (Dedera *et al.*, 1993; Kamps and Baltimore, 1993). Its leukemogenic effect may in part be explained by the expression of Pbx sequences in hematopoietic cells, where it is normally absent. However, also E2A sequences are likely to contribute, since Pbx alone has no transforming activity. E2A contributes two transcriptional activation domains to the fusion protein, whereas Pbx1 contributes its DNA binding homeodomain. These two elements were thought to define E2A-PBX's transforming capability (Kamps *et al.*, 1990; Nourse *et al.*, 1990); however, it was later found that an intact homeodomain is dispensable for E2A-PBX1's oncogenic activity (Monica *et al.*, 1994). It appeared that sequences outside Hox homeodomains interact with heterologous Hox proteins and that the resulting heterodimers form the active transcription factor (Chang *et al.*, 1995). This is reflected in the recent finding that HoxA9 cooperates with the Pbx1-related homeobox gene Meis1 in murine myeloid leukemias, in which both genes were activated by proviral insertion (Nakamura *et al.*, 1996b).

Another example of the involvement of a Hox gene in cancer-related chromosome rearrangements is the Hox11 gene on chromosome 10q24. This gene is upregulated by juxtaposition to TCR enhancer sequences via translocation (10;14) in lymphoid leukemias (Dubé *et al.*, 1991; Hatano *et al.*, 1991; Kennedy *et al.*, 1991).

In view of the role of Hox proteins in transcriptional regulation, it seems likely that the Nup98-HoxA9 functions as a transcription factor, however, also other hypotheses have been put forward (see Chapter 9).

CAN/Nup214 Translocation (6;9)(p23;q34) is the hallmark of a specific subtype of acute myeloid leukemia, characterized by poor prognosis and a young age of onset. It is FAB classified mostly as M2/M4 and rarely as M1 or RAEB (Adriaansen *et al.*, 1991; Soekarman *et al.*, 1992; von Lindern, 1992). This translocation fuses the *DEK* gene, located on chromosome 6p23, to the *CAN* gene, located on chromosome 9q34 (Von Lindern *et al.*, 1992a; Von Lindern *et al.*, 1990). The *DEK-CAN* fusion gene gives rise to a leukemia-specific transcript, which encodes a chimeric DEK-CAN protein (Soekarman *et al.*, 1992; Von Lindern *et al.*, 1992a). The *CAN* gene was also rearranged in an isolated case of acute undifferentiated leukemia, where it was fused to *SET*. *SET* is also located on chromosome 9q34 (von Lindern *et al.*, 1992b).

The molecular analysis of the CAN, DEK-CAN and SET-CAN are described in the following chapters. The possible mechanisms by which the DEK-CAN and SET-CAN fusion proteins are involved in leukemogenesis, and the possible analogy to Nup98-HoxA9 are discussed in Chapter 9.

V. Perspectives

Regulated nucleocytoplasmic transport appears to be an important mechanism in the control of growth and/or differentiation. Although the number of proteins and especially RNAs that are currently known to exhibit regulated nuclear import or export is relatively low, several important systems involved in cellular growth and/or differentiation, such as cell cycle control and signal transduction, are included. In several cases the mechanism of regulation is unknown at present, but the best studied examples point to regulation of cis-acting transport signals as the main mode of control. Whether this is the only way is unknown; it is certainly not the only conceivable mechanism. Study of the regulated transport factors and the continuing dissection of the nucleocytoplasmic transport machinery may reveal new ways of regulating the nucleocytoplasmic exchange.

In this light the finding of mutated transport factors and NPC components in leukemic cells is certainly of interest. Whether these mutations directly influence transport processes, however, is doubtful. It is conceivable that domains of nucleoporins implicated in oncogenesis serve *ad hoc* functions in oncogenic fusion proteins that act in unrelated pathways leading to neoplasmic transformation, and in the case of Tpr, this is the most likely scenario. The Nup98-HoxA9 fusion protein probably serves a direct function in transcriptional control, although the presence of the Nup98 FG-repeat suggests an additional transport-linked attribute (see also Chapter 9). The importinα-like Oho31 protein is so far the only example of the direct involvement of a transport molecule in tumorigenesis. Since this protein appears to be one of a limited number of importinα-like transport factors, the question arises how a mutation in the general transport machinery can cause specific deregulation in cell growth and/or differentiation. However, rather aspecific changes in the cellular biochemistry can lead to transformation. A good example for this is the inhibition of tyrosinephosphatases by vanadate, that results in a 40-fold increase of cellular phosphotyrosine content, and hyperphosphorylation of numerous proteins (Klarlund, 1985; Klarlund *et al.*, 1988). It also results in increased cell division and anchorage independent growth (Klarlund, 1985). These effects are not dependent on the general phosphorylation state of the cell, rather, they are the result of increased phosphorylation of specific molecules in signal transduction pathways (e.g. Sun and Tonks, 1994). Thus, mutations that give rise to aspecific changes in nucleocytoplasmic transport may have similar specific effects via a modified transport of specific factors. Alternatively, mutations in the general transport machinery may function to increase the transport capacity of the cell to a level required for continuous growth, as suggested by Feldherr and Akin (1991).

Experiments to asses the role of nucleoporins and nucleocytoplasmic transport factors in growth control could include (i) identification of proteins that interact with nucleoporin-derived fusion proteins, and testing their significance for the oncogenic process, (ii) screening for rearrangements in genes encoding FG-repeat containing nucleoporins components in leukemias and other neoplasias, (iii) testing the transforming potential of transport/translation

factors including eIF-4E and CBC, and (iv) testing the oncogenic effect of mutations in human importinα homologues.

Acknowledgements

I thank Drs. Daniela Bischof, Steve Morris, and Iain Mattaj for communicating unpublished results.

References

Adam, E.J.H. and Adam, A.A. (1994) Identification of cytosolic factors required for nuclear location sequence-mediated binding to the nuclear envelope. *J Cell Biol*, **125**, 547-555.

Adriaansen, H.J., Soeting, P.W.C., Wolvers-Tettero, I.L.M. and van Dongen, J.J.M. (1991) Immunoglobulin and T cell receptor gene rearrangements in acute non-lymphocytic leukemia. *Leukemia*, **5**, 744-751.

Beg, A.A., Ruben, S.M., Scheinman, R.I., Haskill, S., Rosen, C.A. and Baldwin, A.S., Jr. (1992) I kappa B interacts with the nuclear localization sequences of the subunits of NF-kappa B: a mechanism for cytoplasmic retention. *Genes Dev*, **6**, 1899-913.

Beltz, G.A. and Flint, S.J. (1979) Inhibition of HeLa cell protein synthesis during adenovirus infection. Restriction of cellular messenger RNA sequences to the nucleus. *J Mol Biol*, **131**, 353-73.

Binari, R. and Perrimon, N. (1994) Stripe-specific regulation of pair-rule genes by hopscotch, a putative Jak family tyrosine kinase in Drosophila. *Genes Dev*, **8**, 300-12.

Bischoff, F.R., Krebber, H., Kempf, T., Hermes, I. and Ponstingl, H. (1995) Human RanGTPase-activating protein RanGAP1 is a homologue of yeast Rna1p involved in mRNA processing and transport. *Proc Natl Acad Sci U S A*, **92**, 1749-1753.

Bischoff, F.R. and Ponstingl, H. (1991) Catalysis of guanine nucleotide exchange on Ran by the mitotic regulator RCC1. *Nature*, **354**, 80-2.

Bogerd, H.P., Fridell, R.A., Madore, S. and Cullen, B.R. (1995) Identification of a novel cellular cofactor for the Rev/Rex class of retroviral regulatory proteins. *Cell*, **82**, 485-494.

Borer, R.A., Lehner, C.F., Eppenberger, H.M. and Nigg, E.A. (1989) Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell*, **56**, 379-90.

Borrow, J., Shearman, A.M., Stanton, V.P., Jr., Becher, R., Collins, T., Williams, A.J., Dube, I., Katz, F., Kwong, Y.L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J. and Housman, D.E. (1996) The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nature Genet.*, **12**, 159-167.

Bose, H.R., Jr. (1992) The Rel family: models for transcriptional regulation and oncogenic transformation. *Biochim Biophys Acta*, **1114**, 1-17.

Breeuwer, M. and Goldfarb, D.S. (1990) Facilitated nuclear transport of histone H1 and other small nucleophilic proteins. *Cell*, **60**, 999-1008.

Byrd, D.A., Sweet, D.J., Pante, N., Konstantinov, K.N., Guan, T., Saphire, A.C., Mitchell, P.J., Cooper, C.S., Aebi, U. and Gerace, L. (1994) Tpr, a large coiled coil protein whose amino terminus is involved in activation of oncogenic kinases, is localized to the cytoplasmic surface of the nuclear pore complex. *J Cell Biol*, **127**, 1515-1526.

Chan, P.K. and Chan, F.Y. (1995) Nucleophosmin/B23 (NPM) oligomer is a major and stable entity in HeLa cells. *Biochim Biophys Acta*, **1262**, 37-42.

Chang, C.P., Shen, W.F., Rozenfeld, S., Lawrence, H.J., Largman, C. and Cleary, M.L. (1995) Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev*, **9**, 663-674.

Chi, N.C., Adam, E.J.H. and Adam, S.A. (1995) Sequence and characterization of cytoplasmic nuclear protein import factor p97. *J Cell Biol*, **130**, 265-274.

Cooper, C.S., Park, M., Blair, D.G., Tainsky, M.A., Huebner, K., Croce, C.M. and Vande Woude, G.F. (1984) Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature*, **311**, 29-33.

Cortes, P., Ye, Z.S. and Baltimore, D. (1994) RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1. *Proc Natl Acad Sci U S A*, **91**, 7633-7637.

Csermely, P., Schnaider, T. and Szanto, I. (1995) Signalling and transport through the nuclear membrane. *Biochim Biophys Acta*, **1241**, 425-451.

Cuomo, C.A., Kirch, S.A., Gyuris, J., Brent, R. and Oettinger, M.A. (1994) Rch1, a protein that specifically interacts with the RAG-1 recombination-activating protein. *Proc Natl Acad Sci U S A*, **91**, 6156-60.

Dale, T.C., Imam, A.M., Kerr, I.M. and Stark, G.R. (1989) Rapid activation by interferon alpha of a latent DNA-binding protein present in the cytoplasm of untreated cells. *Proc Natl Acad Sci U S A*, **86**, 1203-7.

De Vita, G., Barba, P., Odartchenko, N., Givel, J.C., Freschi, G., Bucciarelli, G., Magli, M.C., Boncinelli, E. and Cillo, C. (1993) Expression of homeobox-containing genes in primary and metastatic colorectal cancer. *Eur J Cancer*, **29A**, 887-93.

Dedera, D.A., Waller, E.K., LeBrun, D.P., Sen-Majumdar, A., Stevens, M.E., Barsh, G.S. and Cleary, M.L. (1993) Chimeric homeobox gene E2A-PBX1 induces proliferation, apoptosis, and malignant lymphomas in transgenic mice. *Cell*, **74**, 833-43.

Defranco, D.B., Madan, A.P., Tang, Y., Chandran, U.R., Xiao, N. and Yang, J. (1995) Nucleocytoplasmic shuttling of steroid receptors. *Vitam. Horm.* **51**, 315-338.

Di Renzo, M.F., Narsimhan, R.P., Olivero, M., Bretti, S., Giordano, S., Medico, E., Gaglia, P., Zara, P. and Comoglio, P.M. (1991) Expression of the Met/HGF receptor in normal and neoplastic human tissues. *Oncogene*, **6**, 1997-2003.

Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences--a consensus? [see comments]. *Trends Biochem .ci.* **16**, 478-81.

Dubé, I.D., Kamel-Reid, S., Yuan, C.C., Lu, M., Wu, X., Corpus, G., Raimondi, S.C., Crist, W.M., Carroll, A.J., Minowada, J. and et al. (1991) A novel human homeobox gene lies at the chromosome 10 breakpoint in lymphoid neoplasias with chromosomal translocation t(10;14). *Blood*, **78**, 2996-3003.

Elmberger, P.G., Lozano, M.D., Weisenburger, D.D., Sanger, W. and Chan, W.C. (1995) Transcripts of the npm-alk fusion gene in anaplastic large cell lymphoma, Hodgkin's disease, and reactive lymphoid lesions. *Blood*, **86**, 3517-3521.

Fabre, E. and Hurt, E.C. (1994) Nuclear transport. *Curr.Opin.Cell Biol.* **6**, 335-342.

Feldherr, C. and Akin, D. (1993) Regulation of nuclear transport in proliferating and quiescent cells. *Exp. Cell Res.* **205**, 179-186.

Feldherr, C.M. and Akin, D. (1991) Signal-mediated nuclear transport in proliferating and growth-arrested BALB/c 3T3 cells. *J.Cell Biol.* **115**, 933-9.

Fischer, U., Huber, J., Boelens, W.C., Mattaj, I.W. and 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., Boer, J., van Baal, S., Jaegle, M., Von Lindern, M., Murti, K.G., Davis, D., Bonten, J., Buijs, A. and Grosveld, G. (1995) Relocation of the carboxyterminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. *Oncogene*, **10**, 1739-1748.

Fortes, P., Beloso, A. and Ortin, J. (1994) Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. *Embo J*, **13**, 704-12.

Fritz, C.C., Zapp, M.L. and Green, M.R. (1995) A human nucleoporin-like protein that specifically interacts with HIV Rev. *Nature*, **376**, 530-533.

Gallant, P., Fry, A.M. and Nigg, E.A. (1995) Protein kinases in the control of mitosis: focus on nucleocytoplasmic trafficking. *J. Cell Sci. Suppl.*, **19**, 21-28.

Gallant, P. and Nigg, E.A. (1992) Cyclin B2 undergoes cell cycle-dependent nuclear translocation and, when expressed as a non-destructible mutant, causes mitotic arrest in HeLa cells. *J Cell Biol*, **117**, 213-24.

Garcia-Bustos, J., Heitman, J. and Hall, M.N. (1991) Nuclear protein localization. *Biochim Biophys Acta*, **1071**, 83-101.

Gateff, E. (1994) Tumor suppressor and overgrowth suppressor genes of Drosophila melanogaster: developmental aspects. *Int J Dev Biol*, **38**, 565-590.

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

Gerace, L. and Burke, B. (1988) Functional organization of the nuclear envelope. *Annu Rev Cell Biol*, **4**, 335-74.

Görlich, D., Henklein, P., Laskey, R.A. and Hartmann, E. (1996a) A 41 amino acid motif in importin-alpha confers binding to importin- beta and hence transit into the nucleus. *EMBO J.*, **15**, 1810-1817.

Görlich, D., Draft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R.A., Mattaj, I.W., and Izaurralde, E. (1996b) Importin provides a link between nuclear protein import and U snRNA export. *Cell*, in press.

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

Görlich, D., Prehn, S., Laskey, R.A. and Hartmann, E. (1994) Isolation of a protein that is essential for the first step of nuclear protein import. *Cell*, **79**, 767-778.

Görlich, D., Vogel, F., Mills, A.D., Hartmann, E. and Laskey, R.A. (1995) Distinct functions for the two importin subunits in nuclear protein import. *Nature*, **377**, 246-248.

Greco, A., Pierotti, M.A., Bongarzone, I., Pagliardini, S., Lanzi, C. and Della Porta, G. (1992) TRK-T1 is a novel oncogene formed by the fusion of TPR and TRK genes in human papillary thyroid carcinomas. *Oncogene*, **7**, 237-42.

Guiochon-Mantel, A., Delabre, K., Lescop, P. and Milgrom, E. (1994) Nuclear localization signals also mediate the outward movement of proteins from the nucleus. *Proc. Natl Acad. Sci. U S A*, **91**, 7179-7183.

Hall, M. and Peters, G. (1996) Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer. *Adv. Cancer Res.*, **68**, 23-57.

Hanover, J.A. (1992) The nuclear pore: at the crossroads. *FASEB J.*, **6**, 2288-95.

Hatano, M., Roberts, C.W., Minden, M., Crist, W.M. and Korsmeyer, S.J. (1991) Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia. *Science*, **253**, 79-82.

Hinshaw, J.E., Carragher, B.O. and Milligan, R.A. (1992) Architecture and design of the nuclear pore complex. *Cell*, **69**, 1133-41.

Hou, X.S. (1996) Marelle acts downstream of the Drosophila HOP/JAK kinase and encodes a protein similar to the mammalian STATs. *Cell*, **84**, 411-419.

Ihle, J.N. (1995) The Janus protein tyrosine kinase family and its role in cytokine signalling. *Adv. Immunol.*, **60**, 1-35.

Imamoto, N., Shimamoto, T., Takao, T., Tachibana, T., Kose, S., Matsubae, M., Sekimoto, T., Shimonishi, Y. and Yoneda, Y. (1995) In vivo evidence for involvement of a 58 kDa component of nuclear pore- targeting complex in nuclear protein import. *Embo J*, **14**, 3617-3626.

Iovine, M.K., Watkins, J.L. and Wente, S.R. (1995) The GLFG repetitive region of the nucleoporin Nup116p interacts with Kap95p, an essential yeast nuclear import factor. *J. Cell Biol.*, **131**, 1699-713.

Ishikawa, F., Takaku, F., Nagao, M. and Sugimura, T. (1987) Rat c-raf oncogene activation by a rearrangement that produces a fused protein. *Mol Cell Biol*, **7**, 1226-32.

Ishimi, I. and Kikuchi, A. (1991) Identification and molecular cloning of a yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly in vitro. *J Mol Biol*, **266**, 7025-7029.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C. and Mattaj, I.W. (1995) A cap-binding protein complex mediating U snRNA export. *Nature*, **376**, 709-712.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. and Mattaj, I.W. (1994) A nuclear cap

binding protein complex involved in pre-mRNA splicing. *Cell*, **78**, 657-668.

Izaurralde, E. and Mattaj, I.W. (1995) RNA export. *Cell*, **81**, 153-159.

Izaurralde, E., Stepinski, J., Darzynkiewicz, E. and Mattaj, I.W. (1992) A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs. *J Cell Biol*, **118**, 1287-95.

Jans, D.A., Moll, T., Nasmyth, K. and Jans, P. (1995) Cyclin-dependent kinase site-regulated signal-dependent nuclear localization of the SW15 yeast transcription factor in mammalian cells. *J Biol Chem*, **270**, 17064-17067.

Jiang, L.W. and Schindler, M. (1988) Nuclear transport in 3T3 fibroblasts: effects of growth factors, transformation, and cell shape. *J Cell Biol*, **106**, 13-9.

Kamps, M.P. and Baltimore, D. (1993) E2A-Pbx1, the t(1;19) translocation protein of human pre-B-cell acute lymphocytic leukemia, causes acute myeloid leukemia in mice. *Mol Cell Biol*, **13**, 351-7.

Kamps, M.P., Murre, C., Sun, X.H. and Baltimore, D. (1990) A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell*, **60**, 547-55.

Kellogg, D.R., Kikuchi, A., Fujii-Nakata, T., Turck, C.W. and Murray, A.W. (1995) Members of the NAP/SET family of proteins interact specifically with B-type cyclins. *J Cell Biol*, **130**, 661-673.

Kennedy, M.A., Gonzalez-Sarmiento, R., Kees, U.R., Lampert, F., Dear, N., Boehm, T. and Rabbitts, T.H. (1991) HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. *Proc Natl Acad Sci U S A*, **88**, 8900-4.

Kerr, L.D., Inoue, J., Davis, N., Link, E., Baeuerle, P.A., Bose, H.R., Jr. and Verma, I.M. (1991) The rel-associated pp40 protein prevents DNA binding of Rel and NF- kappa B: relationship with I kappa B beta and regulation by phosphorylation. *Genes Dev*, **5**, 1464-76.

Kessler, D.S., Veals, S.A., Fu, X.Y. and Levy, D.E. (1990) Interferon-alpha regulates nuclear translocation and DNA-binding affinity of ISGF3, a multimeric transcriptional activator. *Genes Dev*, **4**, 1753-65.

Kitajima, I., Shinohara, T., Bilakovics, J., Brown, D.A., Xu, X. and Nerenberg, M. (1992) Ablation of transplanted HTLV-I Tax-transformed tumors in mice by antisense inhibition of NF-kappa B [published erratum appears in Science 1993 Mar 12;259(5101):1523]. *Science*, **258**, 1792-5.

Klarlund, J.K. (1985) Transformation of cells by an inhibitor of phosphatases acting on phosphotyrosine in proteins. *Cell*, **41**, 707-17.

Klarlund, J.K., Latini, S. and Forchhammer, J. (1988) Numerous proteins phosphorylated on tyrosine and enhanced tyrosine kinase activities in vanadate-treated NIH 3T3 fibroblasts. *Biochim Biophys Acta*, **971**, 112-20.

Koepp, D.M., Wong, D.H., Corbett, A.H. and Silver, P.A. (1996) Dynamic localization of the nuclear import receptor and its interactions with transport factors. *J Cell Biol*, **133**, 1163-1176.

Kraemer, D., Wozniak, R.W., Blobel, G. and Radu, A. (1994) The human can protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. *Proc Natl Acad Sci USA*, **91**, 1519-1523.

Kraemer, D.M., Strambio-de-Castillia, C., Blobel, G. and Rout, M.P. (1995) The essential yeast nucleoporin NUP159 is located on the cytoplasmic side of the nuclear pore complex and serves in karyopherin-mediated binding of transport substrate. *J Biol Chem*, **270**, 19017-19021.

Kussel, P. and Frasch, M. (1995) Yeast Srp1, a nuclear protein related to Drosophila and mouse pendulin, is required for normal migration, division, and integrity of nuclei during mitosis. *Mol Gen Genet*, **248**, 351-363.

Küssel, P. and Frasch, M. (1995) Pendulin, a Drosophila protein with cell cycle-dependent nuclear localization, is required for normal cell proliferation. *J Cell Biol*, **129**, 1491-1507.

Lang, I., Scholz, M. and Peters, R. (1986) Molecular mobility and nucleocytoplasmic flux in hepatoma cells. *J Cell Biol*, **102**, 1183-90.

Lejbkowicz, F., Goyer, C., Darveau, A., Neron, S., Lemieux, R. and Sonenberg, N. (1992) A fraction of the mRNA 5' cap-binding protein, eukaryotic initiation factor 4E, localizes to the nucleus. *Proc Natl Acad Sci U S*

A, **89**, 9612-6.

Levy, D.E., Kessler, D.S., Pine, R. and Darnell, J.E., Jr. (1989) Cytoplasmic activation of ISGF3, the positive regulator of interferon- alpha-stimulated transcription, reconstituted in vitro. *Genes Dev*, **3**, 1362-71.

Liu, C., Park, M. and Tsao, M.S. (1992) Overexpression of c-met proto-oncogene but not epidermal growth factor receptor or c-erbB-2 in primary human colorectal carcinomas. *Oncogene*, **7**, 181-5.

Mader, S. and Sonenberg, N. (1995) Cap binding complexes and cellular growth control. *Biochimie*, **77**, 40-44.

Martin, K. and Helenius, A. (1991) Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell*, **67**, 117-30.

Maul, G.G. (1977) The nuclear and the cytoplasmic pore complex: structure, dynamics, distribution, and evolution. *Int Rev Cytol Suppl*, 75-186.

Maulbecker, C.C. and Gruss, P. (1993) The oncogenic potential of deregulated homeobox genes. *Cell Growth Differ*, **4**, 431-41.

Melchior, F., Guan, T., Yokoyama, N., Nishimoto, T. and Gerace, L. (1995a) GTP hydrolysis by Ran occurs at the nuclear pore complex in an early step of protein import. *J Cell Biol*, **131**, 571-581.

Melchior, F., Sweet, D.J. and Gerace, L. (1995b) Analysis of Ran/TC4 function in nuclear protein import. *Meth. Enzymol.*, **257**, 279-291.

Miyamoto, S. and Verma, I.M. (1995) Rel/NF-kappa B/I kappa B story. *Adv Cancer Res*, **66**, 255-292.

Moll, T., Tebb, G., Surana, U., Robitsch, H. and Nasmyth, K. (1991) The role of phosphorylation and the CDC28 protein kinase in cell cycle- regulated nuclear import of the S. cerevisiae transcription factor SWI5. *Cell*, **66**, 743-58.

Monica, K., LeBrun, D.P., Dedera, D.A., Brown, R. and Cleary, M.L. (1994) Transformation properties of the E2a-Pbx1 chimeric oncoprotein: fusion with E2a is essential, but the Pbx1 homeodomain is dispensable. *Mol. Cell. Biol.*, **14**, 8304-14.

Moore, M., Schaack, J., Baim, S.B., Morimoto, R.I. and Shenk, T. (1987) Induced heat shock mRNAs escape the nucleocytoplasmic transport block in adenovirus-infected HeLa cells. *Mol Cell Biol*, **7**, 4505-12.

Moore, M.S. and Blobel, G. (1992) The two steps of nuclear import, targeting to the nuclear envelope and translocation through the nuclear pore, require different cytosolic factors. *Cell*, **69**, 939-950.

Moore, M.S. and Blobel, G. (1994) Purification of a Ran-interacting protein that is required for protein import into the nucleus. *Proc Natl Acad Sci USA*, **91**, 10212-10216.

Moroianu, J., Blobel, G. and Radu, A. (1995a) Previously identified protein of uncertain function is karyopherin à and together with karyopherin á docks import substrate at nuclear pore complexes. *Proc Natl Acad Sci USA*, **92**, 2008-2011.

Moroianu, J., Hijikata, M., Blobel, G. and Radu, A. (1995b) Mammalian karyopherin à1á and à2á heterodimers: à1 or à2 subunit binds nuclear localization signal and á subunit interacts with peptide repeat-containing nucleoporins. *Proc Natl Acad Sci USA*, **92**, 6532-6536.

Morris, S.W., Kirstein, M.N., Valentine, M.B., Dittmer, K.G., Shapiro, D.N., Saltman, D.L. and Look, A.T. (1994) Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non- Hodgkin's lymphoma. *Science*, **263**, 1281-4.

Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A. and Matsumoto, K. (1995) Replication factor encoded by a putative oncogene, *set*, associated with myeloid leukemogenesis. *Proc Natl Acad Sci USA*, **92**, 4279-4283.

Nakamura, T., Largaespada, D.A., Lee, M.P., Johnson, L.A., Ohyashiki, K., Toyama, K., Chen, S.J., Willman, C.L., Chen, I.M., Feinberg, A.P., Jenkins, N.A., Copeland, N.G. and Shaughnessy, J.D., Jr. (1996a) Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nature Genet.*, **12**, 154-158.

Nakamura, T., Largaespada, D.A., Shaughnessy, J.D., Jr., Jenkins, N.A. and Copeland, N.G. (1996b) Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nature Genet.*, **12**, 149-153.

Nasmyth, K., Adolf, G., Lydall, D. and Seddon, A. (1990) The identification of a second cell cycle control on the HO promoter in yeast: cell cycle regulation of SW15 nuclear entry. *Cell*, **62**, 631-47.

Nehrbass, U. and Blobel, G. (1996) Role of the nuclear transport factor p10 in nuclear import. *Science*, **272**, 120-122.

Neri, A., Chang, C.C., Lombardi, L., Salina, M., Corradini, P., Maiolo, A.T., Chaganti, R.S. and Dalla-Favera, R. (1991) B cell lymphoma-associated chromosomal translocation involves candidate oncogene lyt-10, homologous to NF-kappa B p50. *Cell*, **67**, 1075-87.

Newmeyer, D.D. and Forbes, D.J. (1988) Nuclear import can be separated into distinct steps in vitro: nuclear pore binding and translocation. *Cell*, **52**, 641-53.

Nourse, J., Mellentin, J.D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S.D. and Cleary, M.L. (1990) Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell*, **60**, 535-45.

Oberleithner, H., Brinckmann, E., Schwab, A. and Krohne, G. (1994) Imaging nuclear pores of aldosterone-sensitive kidney cells by atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 9784-8.

Okamoto, A., Jiang, W., Kim, S.J., Spillare, E.A., Stoner, G.D., Weinstein, I.B. and Harris, C.C. (1994) Overexpression of human cyclin D1 reduces the transforming growth factor beta (TGF-beta) type II receptor and growth inhibition by TGF- beta 1 in an immortalized human esophageal epithelial cell line. *Proc Natl Acad Sci U S A*, **91**, 11576-11580.

O'Neil, R. and Palese, P. (1995) NPI-1, the human homolog of SRP-1, interacts with influenza virus nucleoprotein. *Virology*, **206**, 116-125.

Paine, P.L., Moore, L.C. and Horowitz, S.B. (1975) Nuclear envelope permeability. *Nature*, **254**, 109-14.

Panté, N. and Aebi, U. (1994) Toward the molecular details of the nuclear pore complex. *J Struct Biol*, **113**, 179-189.

Panté, N. and Aebi, U. (1996) Toward the molecular dissection of protein import into nuclei. *Curr Opin Cell Biol*, **8**, 397-406.

Panté, N., Bastos, R., McMorrow, I., Burke, B. and Aebi, U. (1994) Interactions and three-dimensional localization of a group of nuclear pore complex proteins. *J Cell Biol*, **126**, 603-617.

Park, M., Dean, M., Cooper, C.S., Schmidt, M., SJ, O.B., Blair, D.G. and Vande Woude, G.F. (1986) Mechanism of met oncogene activation. *Cell*, **45**, 895-904.

Pathak, R.K., Luskey, K.L. and Anderson, R.G. (1986) Biogenesis of the crystalloid endoplasmic reticulum in UT-1 cells: evidence that newly formed endoplasmic reticulum emerges from the nuclear envelope. *J Cell Biol*, **102**, 2158-68.

Picard, D. and Yamamoto, K.R. (1987) Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *Embo J*, **6**, 3333-40.

Pines, J. and Hunter, T. (1991) Human cyclins A and B1 are differentially located in the cell and undergo cell cycle-dependent nuclear transport. *J Cell Biol*, **115**, 1-17.

Pines, J. and Hunter, T. (1994) The differential localization of human cyclins A and B is due to a cytoplasmic retention signal in cyclin B. *Embo J*, **13**, 3772-3781.

Powell, L. and Burke, B. (1990) Internuclear exchange of an inner nuclear membrane protein (p55) in heterokaryons: in vivo evidence for the interaction of p55 with the nuclear lamina. *J Cell Biol*, **111**, 2225-34.

Price, C., Nasmyth, K. and Schuster, T. (1991) A general approach to the isolation of cell cycle-regulated genes in the budding yeast, Saccharomyces cerevisiae. *J Mol Biol*, **218**, 543-56.

Pruschy, M., Ju, Y., Spitz, L., Carafoli, E. and Goldfarb, D.S. (1994) Facilitated nuclear transport of calmodulin in tissue culture cells. *J. Cell Biol.*, **127**, 1527-36.

Qiu, Y. and Krug, R.M. (1994) The influenza virus NS1 protein is a poly(A)-binding protein that inhibits nuclear export of mRNAs containing poly(A). *J. Virol.*, **68**, 2425-32.

Radu, A., Blobel, G. and Moore, M.S. (1995a) Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. *Proc Natl Acad Sci USA*, **92**, 1769-1773.

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

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

Richardson, W.D., Mills, A.D., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1988) Nuclear protein migration involves two steps: Rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. *Cell*, **52**, 655-664.

Roberts, B. (1989) Nuclear location signal-mediated protein transport. *Biochim Biophys Acta*, **1008**, 263-80.

Rodrigues, G.A. and Park, M. (1993) Dimerization mediated through a leucine zipper activates the oncogenic potential of the met receptor tyrosine kinase. *Mol Cell Biol*, **13**, 6711-22.

Rousseau, D., Kaspar, R., Rosenwald, I., Gehrke, L. and Sonenberg, N. (1996) Translation initiation of ornithine decarboxylase and nucleocytoplasmic transport of cyclin D1 mRNA are increased in cells overexpressing eukaryotic initiation factor 4E. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 1065-1070.

Sato, Y., Abe, S., Mise, K., Sasaki, M., Kamada, N., Kouda, K., Musashi, M., Saburi, Y., Horikoshi, A., Minami, Y. and et al. (1987) Reciprocal translocation involving the short arms of chromosomes 7 and 11, t(7p-;11p+), associated with myeloid leukemia with maturation. *Blood*, **70**, 1654-8.

Schindler, C. and Darnell, J.E., Jr. (1995) Transcriptional responses to polypeptide ligands: the JAK-STAT pathway. *Annu Rev Biochem*, **64**, 621-651.

Schlenstedt, G. (1996) Minireview: Protein import into the nucleus. *FEBS Lett*, **389**, 75-79.

Schmidt-Zachmann, M.S., Dargemont, C., Kuhn, L.C. and Nigg, E.A. (1993) Nuclear export of proteins: the role of nuclear retention. *Cell*, **74**, 493-504.

Sharma, K., Fabre, E., Tekotte, H., Hurt, E.C. and Tollervey, D. (1996) Yeast nucleoporin mutants are defective in pre-tRNA splicing. *Mol Cell Biol*, **16**, 294-301.

Simos, G. and Hurt, E.C. (1995) Nucleocytoplasmic transport: Factors and mechanisms. *FEBS Lett*, **369**, 107-112.

Soekarman, D., Von Lindern, M., Daenen, S., de Jong, B., Fonatsch, C., Heinze, B., Bartram, C., Hagemeijer, A. and Grosveld, G. (1992) The translocation (6;9)(p23;q34) shows consistent rearrangement of two genes and defines a myeloproliferative disorder with specific clinical features. *Blood*, **79**, 1-8.

Soman, N.R., Correa, P., Ruiz, B.A. and Wogan, G.N. (1991) The TPR-MET oncogenic rearrangement is present and expressed in human gastric carcinoma and precursor lesions. *Proc Natl Acad Sci U S A*, **88**, 4892-6.

Stuart, E.T., Yokota, Y. and Gruss, P. (1995) PAX and HOX in neoplasia. *Adv Genet*, **33**, 255-274.

Stutz, F., Neville, M. and Rosbash, M. (1995) Identification of a novel nuclear pore-associated protein as a functional target of the HIV-1 Rev protein in yeast. *Cell*, **82**, 495-506.

Sun, H. and Tonks, N.K. (1994) The coordinated action of protein tyrosine phosphatases and kinases in cell signaling. *Trends Biochem. Sci.*, **19**, 480-485.

Taniguchi, T. (1995) Cytokine signaling through nonreceptor protein tyrosine kinases. *Science*, **268**, 251-5.

Tomiyasu, T., Sasaki, M., Kondo, K. and Okada, M. (1982) Chromosome banding studies in 106 cases of chronic myelogenous leukemia. *Jap J Human Genet*, **27**, 243-58.

Török, I., Strand, D., Schmitt, R., Tick, G., Torok, T., Kiss, I. and Mechler, B.M. (1995) The overgrown hematopoietic organs-31 tumor suppressor gene of Drosophila encodes an Importin-like protein accumulating in the nucleus at the onset of mitosis. *J Cell Biol*, **129**, 1473-1489.

Török, T., Tick, G., Alvarado, M. and Kiss, I. (1993) P-lacW insertional mutagenesis on the second chromosome of Drosophila melanogaster: isolation of lethals with different overgrowth phenotypes. *Genetics*, **135**, 71-80.

Tsai, M.J. and O'Malley, B.W. (1994) Molecular mechanisms of action of steroid/thyroid receptor superfamily

members. *Annu Rev Biochem*, **63**, 451-486.

Verma, I.M., Stevenson, J.K., Schwarz, E.M., Van Antwerp, D. and Miyamoto, S. (1995) Rel/NF-kappa B/I kappa B family: intimate tales of association and dissociation. *Genes Dev*, **9**, 2723-2735.

von Lindern, M. (1992) *Molecular characterization of translocation (6;9) in acute nonlymphocytic leukemia.* Erasmus University, Rotterdam, pp. 172.

von Lindern, M., Fornerod, M., van Baal, S., Jaeglé, M., de Wit, T., Buijs, A. and Grosveld, G. (1992a) The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA. *Mol Cell Biol*, **12**, 1687-1697.

von Lindern, M., Poustka, A., Lerach, H. and Grosveld, G. (1990) The (6;9) chromosome translocation, associated with a specific subtype of acute nonlymphocytic leukemia, leads to aberrant transcription of a target gene on 9q34. *Mol Cell Biol*, **10**, 4016-4026.

von Lindern, M., van Baal, S., Wiegant, J., Raap, A., Hagemeijer, A. and Grosveld, G. (1992b) can, a putative oncogene associated with myeloid leukemogenesis, can be activated by fusion of its 3' half to different genes: characterization of the set gene. *Mol Cell Biol*, **12**, 3346-3355.

Waxman, D.J., Ram, P.A., Park, S.H. and Choi, H.K. (1995) Intermittent plasma growth hormone triggers tyrosine phosphorylation and nuclear translocation of a liver-expressed, Stat 5-related DNA binding protein. Proposed role as an intracellular regulator of male- specific liver gene transcription. *J Biol Chem*, **270**, 13262-13270.

Weis, K., Mattaj, I.W. and Lamond, A.I. (1995) Identification of hSRP1 alpha as a functional receptor for nuclear localization sequences. *Science*, **268**, 1049-1053.

Weis, K., Ryder, U. and Lamond, A.I. (1996) The conserved amino-terminal domain of hSRP1 alpha is essential for nuclear protein import. *EMBO J.*, **15**, 1818-1825.

Wente, S.R. and Blobel, G. (1993) A temperature-sensative NUP116 null mutant forms a nuclear envelope seal over the yeast nuclear pore complex thereby blocking nucleocytoplasmic traffic. *J Cell Biol*, **123**, 275-284.

Wente, S.R. and Blobel, G. (1994) NUP145 encodes a novel yeast glycine-leucine-phenylalanine-glycine (GLFG) nucleoporin required for nuclear envelope structure. *J Cell Biol*, **125**, 955-969.

Wikstrom, A.C., Bakke, O., Okret, S., Bronnegard, M. and Gustafsson, J.A. (1987) Intracellular localization of the glucocorticoid receptor: evidence for cytoplasmic and nuclear localization. *Endocrinology*, **120**, 1232-42.

Wright, C.V. (1991) Vertebrate homeobox genes. *Curr Opin Cell Biol*, **3**, 976-82.

Wu, J., Matunis, M.J., Kraemer, D., Blobel, G. and Coutavas, E. (1995) Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J Biol Chem*, **270**, 14209-14213.

Yang, J. and DeFranco, D.B. (1994) Differential roles of heat shock protein 70 in the in vitro nuclear import of glucocorticoid receptor and simian virus 40 large tumor antigen. *Mol Cell Biol*, **14**, 5088-5098.

Yano, R., Oakes, M., Yamaghishi, M., Dodd, J.A. and Nomura, M. (1992) Cloning and characterization of SRP1, a suppressor of temperature- sensitive RNA polymerase I mutations, in Saccharomyces cerevisiae. *Mol Cell Biol*, **12**, 5640-51.

Yano, R., Oakes, M.L., Tabb, M.M. and Nomura, M. (1994) Yeast Srp1p has homology to armadillo/plakoglobin/beta-catenin and participates in apparently multiple nuclear functions including the maintenance of the nucleolar structure. *Proc Natl Acad Sci U S A*, **91**, 6880-6884.

Yokoyama, N., Hayashi, N., Seki, T., Panté, N., Ohba, T., Nishii, K., Kuma, K., Hayashida, T., Miyata, T., Aebi, U., Fukui, M. and Nishimoto, T. (1995) A giant nucleopore protein that binds Ran/TC4. *Nature*, **376**, 184-188.

Zindy, F., Lamas, E., Chenivesse, X., Sobczak, J., Wang, J., Fesquet, D., Henglein, B. and Brechot, C. (1992) Cyclin A is required in S phase in normal epithelial cells. *Biochem Biophys Res Commun*, **182**, 1144-54.

Chapter 3

The Translocation (6;9), Associated with a Specific Subtype of Acute Myeloid Leukemia, Results in the Fusion of Two Genes, *dek* and *can*, and the Expression of a Chimeric, Leukemia-Specific *dek-can* mRNA

MARIEKE VON LINDERN, MAARTEN FORNEROD, SJOZÈF VAN BAAL, MARTINE JAEGLE, TON DE WIT, ARJAN BUIJS, AND GERARD GROSVELD*

Department of Cell Biology and Genetics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Received 11 December 1991/Accepted 24 January 1992

The translocation (6;9) is associated with a specific subtype of acute myeloid leukemia (AML). Previously, it was found that breakpoints on chromosome 9 are clustered in one of the introns of a large gene named *Cain* (*can*). cDNA probes derived from the 3′ part of *can* detect an aberrant, leukemia-specific 5.5-kb transcript in bone marrow cells from t(6;9) AML patients. cDNA cloning of this mRNA revealed that it is a fusion of sequences encoded on chromosome 6 and 3′ *can*. A novel gene on chromosome 6 which was named *dek* was isolated. In *dek* the t(6;9) breakpoints also occur in one intron. As a result the *dek-can* fusion gene, present in t(6;9) AML, encodes an invariable *dek-can* transcript. Sequence analysis of the *dek-can* cDNA showed that *dek* and *can* are merged without disruption of the original open reading frames and therefore the fusion mRNA encodes a chimeric DEK-CAN protein of 165 kDa. The predicted DEK and CAN proteins have molecular masses of 43 and 220 kDa, respectively. Sequence comparison with the EMBL data base failed to show consistent homology with any known protein sequences.

Defined karyotypic aberrations are associated with specific subtypes of leukemia. Detailed molecular characterization of these aberrations may identify genes involved in leukemogenesis and in the precise regulation of proliferation and differentiation in the hematopoietic system. Translocations are the best-studied chromosomal abnormalities. As the result of a translocation, the function or activity of oncogenes located at or near the translocation breakpoint is altered. In myeloid leukemia three translocation breakpoints have been cloned and analyzed at the molecular level.

The two best studied, t(9;22) in chronic myeloid leukemia (27, 43) and t(15;17) in acute promyelocytic leukemia (2, 8, 12), result in the formation of chimeric genes that encode fusion proteins. In chronic myeloid leukemia this is a BCR-ABL protein that has an enhanced tyrosine kinase activity (34, 49) directly responsible for its in vivo tumorigenic potential (14, 25). In acute promyelocytic leukemia a PML-RARα fusion protein that represents an altered transcription factor (16, 33) is found.

The third translocation is the t(6;9) (p23;q34), found in a specific subtype of acute myeloid leukemia (AML) (1, 39, 41). This leukemia is characterized by a poor prognosis, affects young adults, and is classified mostly as M2 or M4 and rarely as M1 (according to the French-American-British classification of AML). A region on chromosome 9 situated 360 kb downstream of the c-*abl* gene was cloned and analyzed. It was found that breakpoints were clustered in a region of 8 kb in five patients, four with t(6;9) AML and one with acute undifferentiated leukemia (AUL) (47). Through cDNA cloning this region could be identified as one of the introns of a large gene (>100 kb) encoding a 7-kb transcript. This intron was named *icb*-9; the intron containing the breakpoints on chromosome 9 and situated in the middle of a gene named *Cain* (*can*). The 3′ part of *can* is translocated to the 6p− chromosome, and only 3′ *can* probes detect an additional, leukemia-specific 5.5-kb transcript in bone marrow cells from t(6;9) AML patients. No additional transcripts were detected with 5′ *can* probes. The breakpoint region on chromosome 6p23 was isolated from a genomic λEMBL3 library constructed of bone marrow DNA from one of the t(6;9) patients. An area of 40 kb of chromosome 6 DNA was cloned in overlapping phages. Southern blot analysis showed that chromosomal breakpoints t(6;9) AML patients are clustered in a relatively small region of 12 kb.

This article reports the cloning of a cDNA representing the 5.5-kb aberrant transcript specific for t(6;9) AML; the isolation of a novel gene, *dek*, on chromosome 6p23; and the sequence analysis of both *can* and *dek* cDNAs.

MATERIALS AND METHODS

Northern (RNA) blotting. Patient material and cell lines used were described previously (47, 48). RNA of mouse tissue was isolated from BCBA mice. RNA was isolated by either the guanidinium isothiocyanate (11) or the LiCl-Ureum method (5). Total RNA was electrophoresed and blotted as described by Fourney et al. (20). Equal amounts of rRNA were loaded; before the samples were loaded on a denaturing gel, 5% of each sample was loaded on a nondenaturing agarose gel to estimate the amount of rRNA and to adapt the sample quantity if necessary. Northern blots were hybridized in 10% dextran (40). Northern blots of mouse tissues were hybridized with human probes with 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 65°C, and filters were washed with 1× SSC at 65°C for *dek* probes and with 0.3× SSC at 65°C for *can* probes. Probes were labelled by the method of Feinberg and Vogelstein (19).

cDNA cloning. One hundred micrograms of total RNA from patient DK was heat denatured and annealed to 10 μg

* Corresponding author.

of a 21-mer, 5'GAAGGACTAGGTGCACCATGT3', at 55°C. First-strand synthesis was done with avian reverse transcriptase (26). Second-strand synthesis was done according to the RNaseH method (24). The DNA was blunt ended with T4 polymerase and treated with *Eco*RI methylase (Sigma). *Eco*RI linkers were ligated onto the cDNA with T4 ligase and RNA ligase (40), and after *Eco*RI digestion, the cDNA was size selected on a Sephacryl S-1000 column. cDNA larger than 1 kb was ligated into the *Eco*RI site of λgt10 (31). Phage DNA was packaged by using packaging extracts (GIGA gold; Stratagene). PFU (19×10^6) were generated, of which only 10% contained inserts, estimated by analysis of randomly picked phages. The other 90% most likely contained linker sequences. The human testis cDNA library in λgt11 was purchased from Clonetech (Palo Alto, Calif.). The CMLO λEMBL3 library was described by Hermans et al. (28).

Sequence determination and analysis. Restriction fragments of cDNA clones were subcloned in M13. Overlapping cDNA sequences on both strands were determined by dideoxy sequencing (42). Initially, M13 primers were used; when no suitable restriction sites were present a primer was generated on the basis of already available cDNA sequence. To establish intron-exon borders, genomic fragments containing the exon of interest were subcloned into M13 and a primer near the putative intron-exon border was generated to prime the sequence reaction. Sequences were analyzed with the computer program Microgenie, and the EMBL data base was used to search for homologous sequences at both the nucleotide and amino acid levels.

Cloning of the 3' end of *can*. Thirty micrograms of total RNA of bone marrow cells from AUL patient SE was heat denatured, and first-strand cDNA was synthesized with avian reverse transcriptase by using 100 pmol of the 35-mer 5'GTCGCGAATTCGTCGACGCGTTTTTTTTTTTTTT3' as a primer (21, 28). Excess primer was removed by isopropanol precipitation. One hundredth of the cDNA reaction was amplified by using *Taq* polymerase (Perkin-Elmer Cetus) and the primers 5'GTCGCGAATTCGTCGACGCG3' and 5'GCCTTTGGATCCCTGGGACCAACCGC3'. The latter primer is located 180 bp upstream of the poly(A) signal in *can* cDNA. The amplified fragment of 230 bp was sequenced by using a protocol for direct sequencing of fragments produced by an asymmetric polymerase chain reaction (32).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this article have been submitted to the EMBL, Genbank, and DDBJ nucleotide sequence data bases under accession numbers X64228 (*can*) and X64229 (*dek*).

RESULTS

Analysis of the *can* gene and transcript. As reported previously, a nearly full-length *can* cDNA was isolated in the overlapping cDNA clones hXT23, hXT37, hXT54, and hXT65 (47). Originally, cDNA clone hX8 was thought to represent the 5' part of the *can* mRNA. However, a more detailed mapping analysis showed that the 5' part of hX8 does not belong to the *can* gene and is in fact not even located on chromosome 9 (data not shown). Therefore, hX8 must be considered a cloning artifact. As several (11) independent cDNA clones appeared to have 5' ends mapping close to the 5' end of hXT23, we assumed that the 5' end of the latter clone maps in the vicinity of the *can* mRNA cap site.

The genomic map of *can*, reported previously, extended over 70 kb but did not include the 3' part of the *can* gene. Therefore, cDNA clone hXT65 was used to screen a genomic λEMBL3 library, and many hybridizing phages were isolated. Clones Al1F10.6, Al1F10.2, Al1F10.8, and Al1F10.12 were selected since they covered the largest stretch of DNA, and they were analyzed in more detail. As indicated in Fig. 1A, a gap is still present between Al1F10.8 and Al1F10.12. The total amount of *can* sequences cloned in phages is 130 kb. Since the gene is located on a *Bss*HII fragment of 170 kb (47) and no *Bss*HII site is present in Al1F10.12, it was deduced from Fig. 1A that the gap between Al1F10.8 and Al1F10.12 can range between 1 and 40 kb. Restriction enzyme fragments that contain exons were determined by hybridization of Southern blots containing *Eco*RI, *Bam*HI, and *Hin*dIII digests of the phages with *can* cDNA clone hXT65 (Fig. 1C).

The overlapping *can* cDNA clones were sequenced and appeared to contain a large open reading frame (ORF) of 6,270 nucleotides (nt) encoding a putative protein of 220 kDa (Fig. 2). This ORF starts in clone hXT23 and ends in clone hXT65. A 700-bp *Hin*dIII-*Pst*I fragment of phage Al1F3, in which the *Bss*HII site is located (Al1F3E4HP), was also sequenced. Figure 3A shows that the sequence of Al1F3E4HP is colinear with hXT23 up to its 5' end. Other cDNA clones have 5' ends mapping near the 5' end of hXT23. Whether this region contains *can* promoter sequences has to be tested. At the 5' end, the *can* cDNA contains ATG start codons at positions 95 to 97, 107 to 109, and 115 to 117. The sequence around the codon at position 95 is concordant with the consensus sequence postulated by Kozak (35), which suggests that this methionine is probably the start of the CAN protein. The first stop codon in this frame is at position 6365. The sequence of cDNA clone hXT65 ends immediately 3' of what appeared to be a variant polyadenylation signal: ATTAAA (nt 6562 to 6567). As no poly(A) tail was present in this clone, the 3' end of the *can* transcript was amplified by using the protocol for rapid amplification of cDNA ends (21) from a position 180 bp 5' of the poly(A) signal to the poly(A) tail. The sequence of this amplified fragment showed that the poly(A) tail starts 16 nt downstream of the ATTAAA signal. The 3' end of hXT65 hybridized to genomic λEMBL3 phage Al1F10.12. Sequence analysis showed that the 3' exon of *can* is present in this phage. Its sequence is colinear with the cDNA sequence down to the poly(A) tail (Fig. 3B).

Since previous mapping data localized the t(6;9) breakpoints in the middle of cDNA clone hXT37 (Fig. 1C), the breakpoints must occur within the *can* ORF. To exactly localize the position of *icb*-9 within the ORF, genomic clones were used to sequence the intron-exon borders delineating this intron. This showed that the translocation breakpoints occur between codons 812 and 813 (nt 2530 to 2531) in the ORF of the *can* mRNA (Fig. 2 and 3C). Because of the translocation, 4,053 nt of the *can* cDNA are encoded on the 6p− chromosome. As a consequence, cDNA probes located within these 4,053 nt recognize a specific 5.5-kb transcript in bone marrow cells from t(6;9) AML patients (47).

Cloning the *dek-can* hybrid cDNA. To resolve the identity of the t(6;9) AML-specific 5.5-kb mRNA, a primed cDNA library was constructed by using total RNA of bone marrow from t(6;9) patient DK and a 21-nt primer mapping 800 bp downstream of the translocation breakpoint in the *can* cDNA (Fig. 1C). Part of the library (2×10^6 PFU) was screened with a 360-bp *Bam*HI-*Rsa*I (hXT37BR) fragment, indicated in Fig. 1C. Two clones (DK1 and DK2 [1.3 and 1.5 kb, respectively]) were isolated and characterized. They

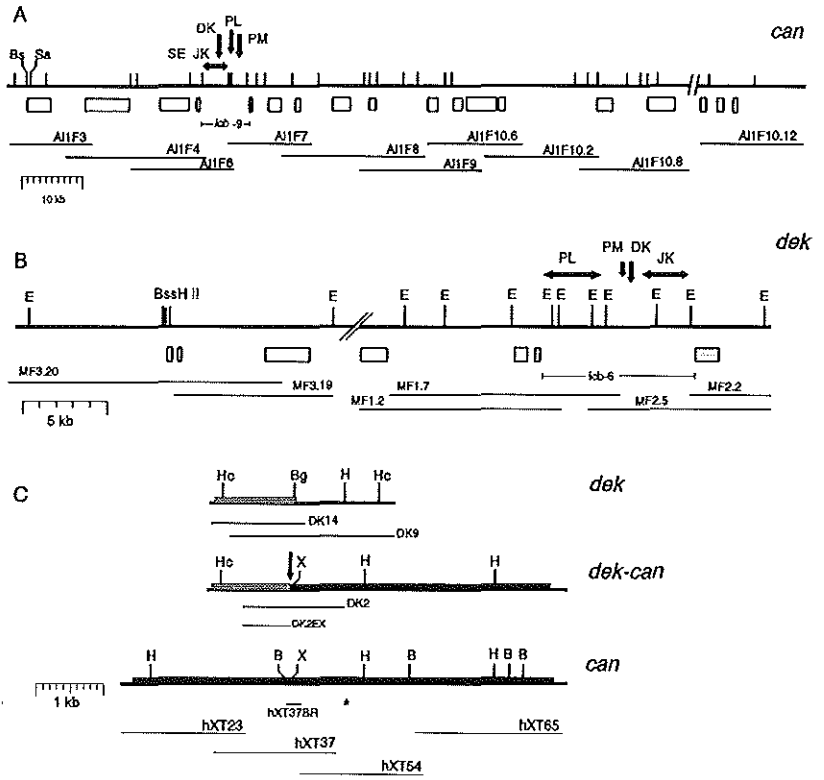
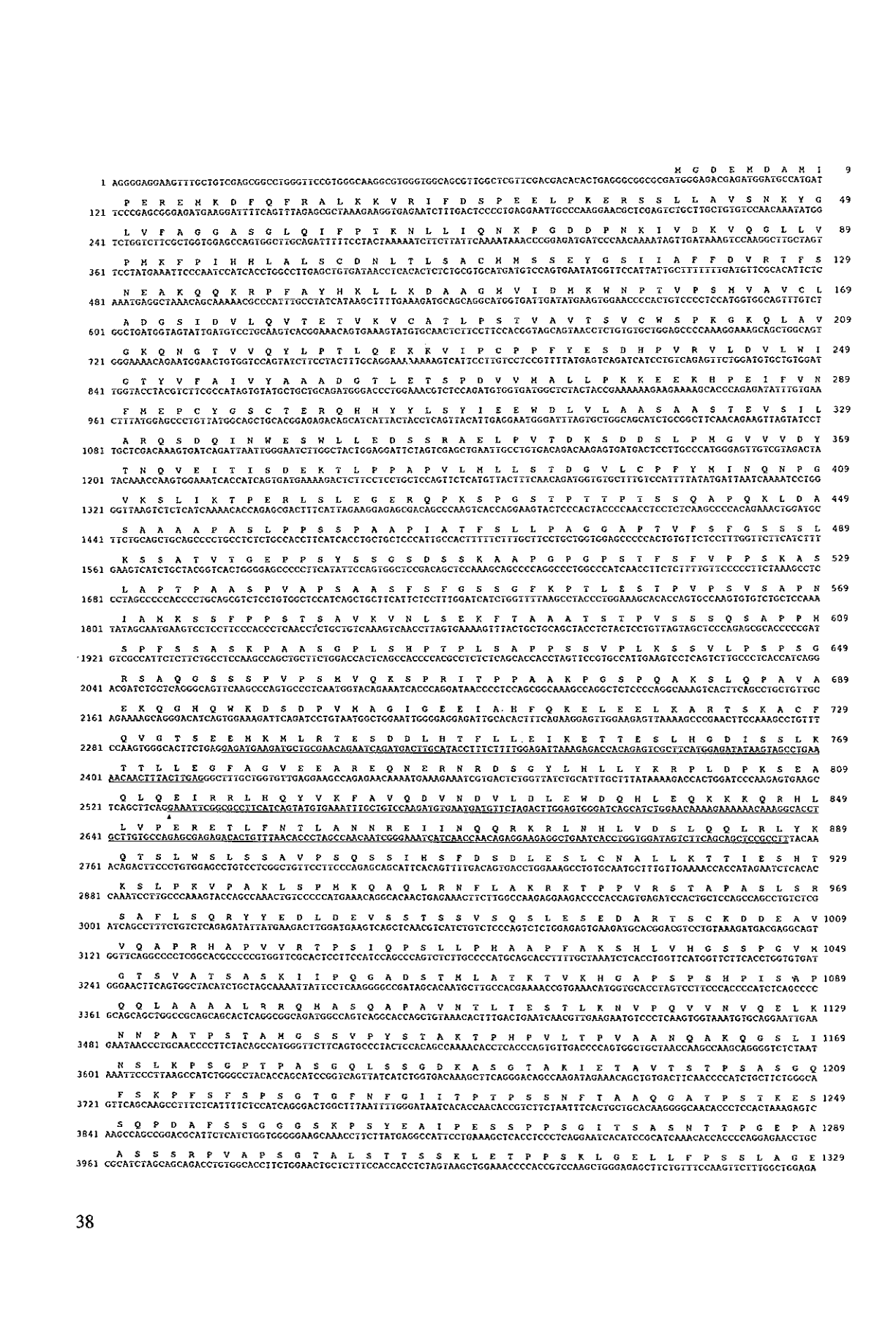


FIG. 1. Restriction maps of the *can* and *dek* genes and *dek*, *dek-can*, and *can* cDNAs. (A) Genomic map of the *can* gene. Vertical lines represent *Eco*RI sites. Open boxes represent restriction enzyme fragments hybridizing to *can* cDNA probes. The positions of the breakpoints of chromosomes from t(6;9) AML patients DK, PM, JK, and PL and AUL patient SE are indicated by arrows. They are all located in *icb*-9. Below the map, isolated genomic phages are depicted. AllF3, AllF4, AllF6, AllF7, AllF8, and AllF9 were reported previously. AllF10.6, AllF10.2, AllF10.8, and AllF10.12 were isolated by using cDNA clone hXT65 as probe. The gap between AllF10.8 and AllF10.12 is at maximum 40 kb. The scale is indicated in kilobases. (B) Genomic map of *dek*. Open boxes indicate restriction fragments hybridizing to cDNA probes; these fragments were delimited by various restriction enzyme sites not shown in this map. Stippled boxes are mapped exons. The positions of the breakpoints of chromosomes from t(6;9) AML patients DK, PM, PL, and JK are indicated by arrows; they are all located in *icb*-6. MF3.20, MF3.19, MF1.2, MF1.7, MF2.5, and MF2.2 are λEMBL3 phages from which the map has been deduced. The gap between MF3.19 and MF1.2 is estimated to be only a few kilobases. (C) Restriction maps of the cDNAs of *dek*, *dek-can*, and *can* are depicted. A scale for the cDNA maps is indicated. Arrows indicate the position of the breakpoints. The ORF of *dek* is indicated by a cross-hatched bar, and the ORF of *can* is indicated by a solid bar on top of the lines that indicate the cDNAs. The chimeric cDNA DK2 has been isolated from a primed cDNA library that was made with a primer, indicated by an asterisk. This library was screened with probe hXT37BR. DK9 and DK14 are *dek* cDNAs isolated with probe DK2EX from a λgt11 cDNA library derived from human testis RNA. Abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; X, *Xba*I.

appeared to be colinear with *can* cDNA from the primer at the 3′ end of the cDNA clones, exactly up to the 5′ end of the exon flanking *icb*-9 at its 3′ side. Upstream of this point, both clones are identical but deviate completely from the *can* sequence. To determine the chromosomal origin of these sequences, a 5′ DK2 fragment (probe DK2EX, a 700-bp *Eco*RI-*Xba*I fragment [Fig. 1C]) was hybridized to a Southern blot containing DNA of a hybrid cell panel with the segregated translocation chromosomes involved in the t(6;9) (48). The probe hybridized to DNA of cell lines containing chromosome 6 and 6p− (results not shown). The same probe was hybridized to a Northern blot containing RNA of HeLa cells, hematopoietic cell lines (Daudi, HL60, KG1, and K562), and bone marrow cells from t(6;9) AML patient DK and AUL patient SE. This revealed the presence of a 2.7-kb transcript in all lanes and an additional 5.5-kb transcript in the t(6;9) AML patient bone marrow sample (Fig. 4). This 5.5-kb transcript is identical in size to the aberrant transcript detected with 3′ *can* probes in a sample from this patient (47). These results proved that sequences encoded by a gene on chromosome 6 are present in the t(6;9) AML-specific 5.5-kb transcript, which is thus identified as a chimeric mRNA.

It is noteworthy that in a sample from AUL patient SE no aberrant transcript was detected by the chromosome 6 probe, while hybridization with 3′ *can* probes clearly de-



M G D E M D A M I 9
1 AGGGGAGGAAGTTTGCTGTCGAGCGGCCTGGGTTCCGTGGGCAAGGCGTGGGTGGCAGCGTTGGCTCGTTCGACGACACACTGAGGGCGGCGCGGATGGGAGACGAGATGGATGCCATGAT

P E R E M K D F Q F R A L K K V R I F D S P E E L P K E R S S L L A V S N K Y G 49
121 TCCCGAGCGGGAGATGAAGGATTTCAGTTAGAGCGCTAAAGAAGGTGAGAATCTTTGACTCCCCTGAGGAATTGCCCAAGGAACGCTCGAGTCTGCTTGCTGTGTCCAACAAATATGG

L V F A G G A S G L Q I F P T K N L L I Q N K P G D D P N K I V D K V Q G L L V 89
241 TCTGGTCTTCGGCTGGTGGAGCCAGTGGCTTGCAGATTTTTCCTACTAAAAATCTTCTTATTCAAAATAAACCCGGAGATGATCCCAACAAAATAGTTGATAAAGTCCAAGGCTTGCTAGT

P M K F P I H H L A L S C D N L T L S A C H M S S E Y G S I I A F F D V R T F S 129
361 TCCTATGAAATTCCCAATCCATCACCTGGCCTTGAGCTGTGATAACCTCACACTCTCTGCGTGCATGATGTCCAGTGAATATGGTTCCATTATTGCTTTTTTTGATGTTCGCACATTCTC

N E A K Q Q K R P F A Y H K L L K D A A G M V I D M K W N P T V P S M V A V C L 169
481 AAATGAGGCTAAACAGCAAAAACGCCCATTTGCCTATCATAAGCTTTTGAAAGATGCAGCAGGCATGGTGATTGATATGAAGTGGAACCCCACTGTCCCCTCCATGGTGGCAGTTTGTCT

A D G S I D V L Q V T E T V K V C A T L P S T V A V T S V C W S P K G K Q L A V 209
601 GGCTGATGGTAGTATTGATGTCCTGCAAGTCACGGAAACAGTGAAAGTATGTGCAACTCTTCCTTCCACGGTAGCAGTAACCTCTGTGTGCTGGAGCCCCAAAGGAAAGCAGCTGGCAGT

G K Q N G T V V Q Y L P T L Q E K K V I P C P P F Y E S D H P V R V L D V L W I 249
721 GGGAAAACAGAATGGAACTGTGGTCCAGTATCTTCCTACTTTGCCAGGAAAAAAAAGTCATTCCTTGTCCTCCGTTTTATGAGTCAGATCATCCTGTCAGAGTTCTGGATGTGCTGTGGAT

G T Y V F A I V Y A A A D G T L E T S P D V V M A L L P K K E E K H P E I F V N 289
841 TGGTACCTACGTCATTCGCCATAGTGTATGCTGCAGATGGGACCCTGGAAACGTCCAGATGTGGTGATGGCTCTACTACCGAAAAAAGAAGAAAAGCACCCAGAGATATTTGTGAA

F M E P C Y G S C T E R Q H H Y Y L S Y I E E W D L V L A A S A A S T E V S I L 329
961 CTTTATGGAGCCCTGTTATGGCAGCTGCACGGAGAGACAGCATCATTACTACCTCAGTTACATTGAGGAATGGGATTTAGTGCTGGCAGCATCTGCGGCTTCAACAGAAGTTAGTATCCT

A R Q S D Q I N W E S W L L E D S S R A E L P V T D K S D D S L P M G V V V D Y 369
1081 TGCTCGACAAAGTGATCAGATTAATTGGGAATCTTGGCTACTGGAGGATTCTAGTCGAGCTGAATTGCCTGTGACAGACAAGAGTGATGACTCCTTGCCCATGGGAGTTGTCGTAGACTA

T N Q V E I T I S D E K T L P P A P V L M L L S T D G V L C P F Y M I N Q N P G 409
1201 TACAAACCAAGTGGAAATCACCATCAGTGATGAAAAGACTCTTCCTCCTGCTCCAGTTCTCATGTTACTTTCAACAGATGGTGTGCTTTGTCCATTTTATATGATTAATCAAAATCCTGG

V K S L I K T P E R L S L E G E R Q P K S P G S T P T T P T S S Q A P Q K L D A 449
1321 GGTTAAGTCTCTCATCAAAACACCAGAGCGACTTTCATTAGAAGGAGAGCGACAGCCCAAGTCACCAGGAAGTACTCCCACTACCCCAACCTCCTCTCAAGCCCCACAGAAACTGGATGC

S A A A A A P A S L P P S S P A A P I A T F S L L P A G G A P T V F S F G S S S L 489
1441 TTCTGCAGCTGCAGCCCCTGCCTCTCTCCCACCTTCATCACCTGCTGCTCCCATTGCCACTTTTTCTTTGGTTCCTGCTGGTGGAGCCCCCACTGTGTTCTCCTTTGGTTCTTCATCTTT

K S S A T V T G E P P S Y S S G S D S S K A A P G P G P S T F S F V P P S K A S 529
1561 GAAGTCATCTGCTACGGTCACTGGGGAGCCCCCTTCATATTCCAGTGGCTCCGACAGCTCCAAAGCAGCCCCAGGCCCTGGCCCATCAACCTTCTCTTTTGTTCCCCCTTCTAAAGCCTC

L A P T P A A S P V A P S A A S F S F G S S G F K P T L E S T P V P S V S A P N 569
1681 CCTAGCCCCCACCCCTGCAGCGTCTCCTGTGGCTCCATCAGCTGCTTCATTCTCCTTGGATCATCTGGTTTTAAGCCTACCCTGGAAAGCACACCAGTGCCAAGTGTGTCTGCTCCAAA

I A M K S S F P P S T S A V K V N L S E K F T A A A T S T P V S S S Q S A P P H 609
1801 TATAGCAATGAAGTCCTCCTTCCCACCCTCAACCTCTGCTGTCAAAGTCAACCTTAGTGAAAAGTTTACTGCTGCAGCTACCTCTACTCCTGTTAGTAGCTCCCAGAGCGCACCCCCGAT

S P F S S A S K P A A S G P L S H P T P L S A P P S S V P L K S S V L P S P S G 649
1921 GTCGCCATTCTCTTCTGCCTCCAAGCCAGCTGCTTCTGGACCACTCAGCCACCCCACGCCTCTCTCAGCACCACCTAGTTCCGTGCCATTGAAGTCCTCAGTCTTGCCCTCACCATCAGG

R S A Q G S S S P V P S M V Q K S P R I T P P A A K P G S P Q A K S L Q P A V A 689
2041 ACGATCTGCTCAGGGCAGTTCAAGCCCAGTGCCCTCAATGGTACAGAAATCACCCAGGATAACCCCTCCAGCGGCAAAGCCAGGCTCTCCCCAGGCAAAGTCACTTCAGCCTGCTGTTGC

E K Q G N Q W K D S D P V M A G I G E E I A H F Q K E L E E L K A R T S K A C F 729
2161 AGAAAAGCAGGGACATCAGTGGAAAGATTCAGATCCTGTAATGGCTGGAATTGGGGAGGAGATTGCACACTTCAGAAGGAGTTGGAAGAGTTAAAAGCCCGAACTTCCAAAGCCTGTTT

Q V G T S E E M K M L R T E S D D L H T F L L E I K E T T E S L H G D I S S L K 769
2281 CCAAGTGGGCACTTCTGAGGAGATGAAGATGCTGCGAACAGAATCAGATGACTTGCATACCTTTCTTTTGGAGATTAAAGAGACCACAGAGTCGCTTCATGGAGATATAAGTAGCCTGAA

T T L L E G F A G V E E A R E Q N E R N R D S G Y L H L L Y K R P L D P K S E A 809
2401 AACAACTTTACTTGAGGGCTTGCTGGTGTTGAGGAAGCCAGAGAACAAAATGAAAGAAATCGTGACTCTGGTTATCTGCATTTGCTTTATAAAAAGACCACTGGATCCCAAGAGTGAAGC

Q L Q E I R R L H Q Y V K F A V Q D V N D V L D L E W D Q H L E Q K K K Q R H L 849
2521 TCAGCTTCAGGAAATTCGGCGCCTTCATCAGTATGTGAAATTTGCTGTCCAAGATGTGAATGATGTTCTAGACTTGGAGTGGGATCAGCATCTGGAACAAAAGAAAAAACAAAGGCACCT

L V P E R E T L F N T L A N N R E I I N Q Q R K R L N H L V D S L Q Q L R L Y K 889
2641 GCTTGTGCCAGAGCGAGAGACACTGTTTAACACCCTAGCCAACAATCGGGAAATCATCAACCAACAGAGGAAGAGCGCTGAATCACCTGGTGGATAGTCTTCAGCAGCTCCGCCTTTACAA

Q T S L W S L S S A V P S Q S S I H S F D S D L E S L C N A L L K T T I E S H T 929
2761 ACAGACTTCCCTGTGGAGCCTGTCCTCGGCTGTTCCTTCCCAGAGCAGCATTCACAGTTTTGACAGTGACCTGGAAAGCCTGTGCCAATGCTTTGTTGAAAACCACCATAGAATCTCACAC

K S L P K V P A K L S P M K Q A Q L R N F L A K R K T P P V R S T A P A S L S R 969
2881 CAAATCCTTGCCCAAAGTACCAGCCAAACTGTCCCCCATGAAACAGGCACAACTGAGAAACTTCTTGGCCAAGAGGAAGACCCCACCAGTGAGATCCACTGCTCCAGCCAGCCTGTCTCG

S A F L S Q R Y Y E D L D E V S S T S S V S Q S L E S E D A R T S C K D D E A V 1009
3001 ATCAGCCTTCTGTCTCAGAGATATTATGAAGACTTGGATGAAGTCAGCTCAACGTCATCTGTCTCCCAGTCTCTGGAGAGTGAAGATGCACGGACGTCCTGTAAAGATGACGAGGCAGT

V Q A P R H A P V V R T P S I Q P S L L P H A A P F A K S H L V H G S S P G V K 1049
3121 GGTTCAGGCCCCTCGGCACGCCCCCGTGGTTCGCACTCCTTCCATCCAGCCCAGTCTCTTGCCCCATGCAGCACCTTTTGCTAAATCTCACCTGGTTCATGGTTCTTCACCTGGTGTGAT

G T S V A T S A S K I I P Q G A D S T M L A T K T V K H G A P S P S H P I S A P 1089
3241 GGGAACTTCAGTGGCTACATCTGCTACGAAATTATTCCTCAAGGGGCCGATAGCACAATGCTTGCCACGAAAACCGTGAAACATGGTGCACCTAGTCCTTCCCACCCCATCTCAGCCCC

Q Q L A A A A A L R R Q M A S Q A P A V N T L T E S T L K N V P Q V V N V Q E L K 1129
3361 GCAGCAGCTGGCCGCAGCAGCACTCAGGCGGCAGATGGCCAGTCAGGCACCAGCTGTAAACACTTTGACTGAATCAACGTTGAAGAATGTCCCTCAAGTGGTAAATGTGCAGGAATTGAA

N N P A T P S T A M G S S V P Y S T A K T P H P V L T P V A A N Q A K Q G S L I 1169
3481 GAATAACCCTGCAACCCCTTCTACAGCCATGGGTCTTCAGTGCCCTACTCCACAGCCAAAACACCTCACCCAGTGTTGACCCCAGTGGCTGTAACCAAGCCAAGCAGGGGTCTCTAAT

N S L K P S G P T P A S G Q L S S G D K A S G T A K I E T A V T S T P S A S G Q 1209
3601 AAATTCCCTTAAGCCATCTGGGCCTACACCAGCATCCGGTCAGTTATCATCTGGTGACAAAGCTTCAGGGACAGCCAAGATAGAAACAGCTGTGACTTCAACCCCATCTGCTTCTGGGCA

F S K P F S F S P S G T G F N F G I I T P T P S S N F T A A Q Q A T P S T K E S 1249
3721 GTTCAGCAAGCCTTTCTCATTTTCTCCATCAGGGACTGGCTTTAATTTTGGGATAATCACACCAACACCGTCTTCATTTCACTGCTGCACAAGGGGCAACACCCTCCACTAAAGAGTC

S Q P D A F S S G G G S K P S Y E A I P E S S P P S G I T S A S N T T P G E P A 1289
3841 AAGCCAGCCGGACGCATTCTCATCTGGTGGGGGAAGCAAACCTTCTTATGAGGCCATTCCTGAAAGCTCACCTCCCTCAGGAATCACATCCGCATCAAACACCACCCCAGGAGAACCTGC

A S S S R P V A P S G T A L S T T S S K L E T P P S K L G E L L F P S S L A G E 1329
3961 CGCATCTAGCAGCAGACCTGTGGCACCTTCTGGAACTGCTCTTTCCACCACCTCTAGTAAGCTGGAAACCCCACCGTCCAAGCTGGGAGAGCTTCTGTTTCCCAAGTTCTTTGGCTGGAGA

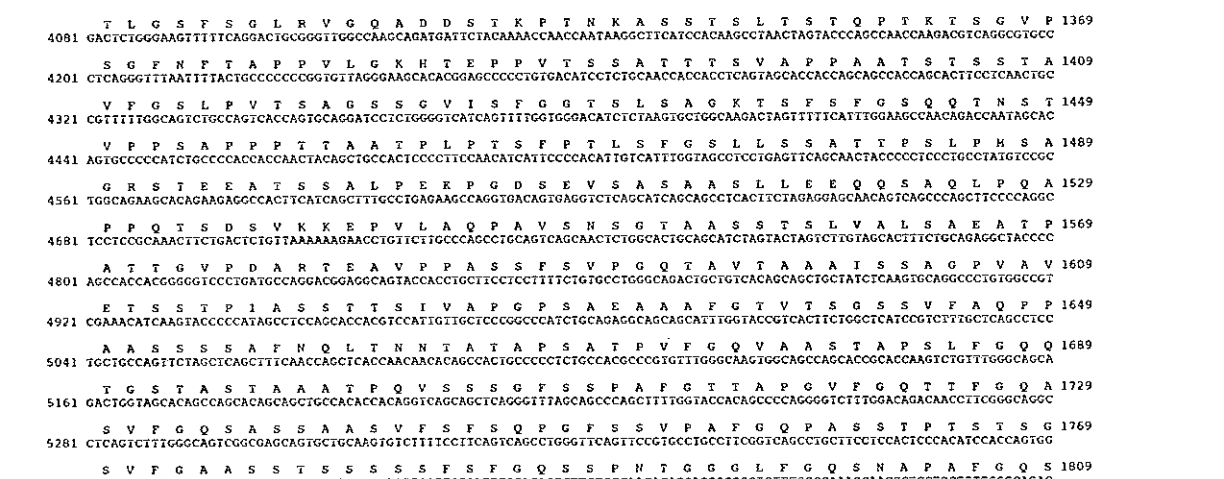
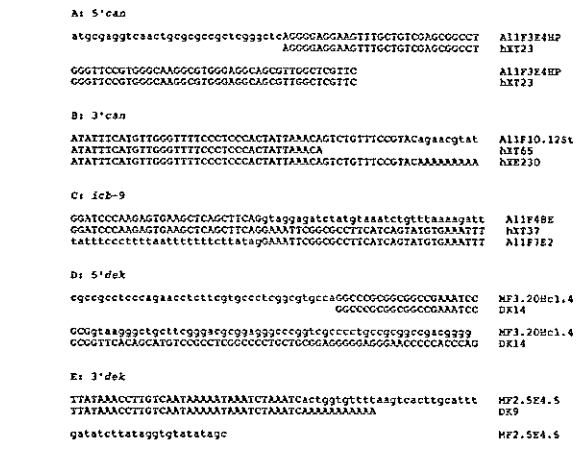


FIG. 2. cDNA sequence and putative amino acid sequence of the CAN protein. The position of the t(6;9) breakpoint is indicated by a solid triangle (nt 2530 to 2531). The putative leucine zipper (Fig. 6A, box A) and amphipathic helices (Fig. 6A, box B) are underlined.

tected an aberrant mRNA of 5.5 kb in the Northern blot of a sample from this patient (47). This result is in agreement with the observation that bone marrow cells from AUL patient SE contained a breakpoint in *icb*-9 but failed to show a breakpoint in band p23 on chromosome 6, which contains breakpoints of chromosomes from t(6;9) AML patients.

Cloning *dek* cDNA. To isolate a cDNA of the normal 2.7-kb transcript encoded on chromosome 6, a λgt11 cDNA library, derived from human testis RNA, was screened with probe DK2EX. In total, 24 clones were isolated and analyzed by restriction enzyme mapping. Two overlapping cDNA clones (DK9 and DK14) that contained 2.7 kb of contiguous sequences, probably representing the full-length transcript, were identified. The full-length *dek* cDNA clone of 2,699 nt was sequenced (Fig. 5). It contains an ORF of 1,125 nt encoding a putative protein of 375 amino acids and with an estimated molecular mass of 43 kDa, which is followed by a large 3′ untranslated region of 1,541 nt. The ATG codon is located at positions 34 to 36, which matches the Kozak consensus sequence (35). The predicted amino acid sequence of the ORF is shown in Fig. 5. The 3′ end of the cDNA sequence contains two AATAAA poly(A) addition sites next to each other, at positions 2682 and 2688, followed by a poly(A) stretch at position 2702.

As reported previously (47), the genomic area of chromosome 6p23, containing the breakpoints of chromosomes from four t(6;9) AML patients, was isolated in four overlapping λEMBL3 clones. The t(6;9) breakpoints appeared to map in a stretch of 12 kb in the middle of this region. The cDNA clones DK9 and DK14 were hybridized to Southern blots containing DNA of these genomic phages, which was digested with several restriction enzymes. Genomic fragments hybridizing to cDNA probes are present at either side of, but not within, the region that contains the chromosome 6 translocation breakpoints (Fig. 1B). Initially, the 12-kb breakpoint region was mapped by Southern blot analysis of genomic DNA and was delineated by two *Eco*RV sites. More precise mapping of the cloned chromosomal DNA reduced the size of the breakpoint region on chromosome 6p23 to an intron of 9 kb. Like what was done for the *can* gene, the intron-exon borders of the *dek* exons that flank



MF3.20Hc1.4
DK14
MF3.20Hc1.4
DK14
MF2.5E4.5
DK9
MF2.5E4.5
MF1.7E2.7
DK9
MF2.5E4.5

FIG. 3. Comparison of genomic and cDNA sequences of *dek* and *can*. (A) Genomic 700-bp *Hin*dIII-*Pst*I fragment (AllF3E4HP) isolated from phage AllF3 (Fig. 1A) contains the most 5′ sequences of cDNA clone hXT23. Presumably, these sequences belong to the first exon of *can*. (B) Genomic 900-bp *Stu*I fragment (AllF10.12St) isolated from phage AllF10.12 (Fig. 1A) contains the most 3′ exon of *can*. cDNA clone hXT65 ends immediately 3′ of the poly(A) signal; therefore, a 230-bp cDNA fragment (hXE230) that contains the poly(A) signal and poly(A) tail was generated by using the protocol for rapid amplification of cDNA ends. The poly(A) signal is underlined. (C) The intron-exon borders flanking *icb*-9 were determined to prove that the t(6;9) breakpoints are located in a single intron. A 1.2-kb *Bam*HI-*Eco*RI fragment (AllF4BE) from phage AllF4 (Fig. 1A) contains part of the exon flanking *icb*-9 upstream. In fact, the *Bam*HI site is located in the exon. A 2-kb *Eco*RI fragment (AllF7E2) from phage AllF7 (Fig. 1A) contains the exon flanking *icb*-9 downstream. The sequence of cDNA hXT37 shows both exons joined together. (D) A 1.4-kb *Hin*cII fragment (MF3.20Hc1.4) from the genomic *dek* phage MF3.20 (Fig. 1B) contains the most 5′ *dek* cDNA sequences of clone DK14 in a 240-bp *Bss*HII fragment. (E) *dek* has a large 3′ exon located in a 4.5-kb *Eco*RI fragment (MF2.5E4.5) of phage MF2.5 (Fig. 1B). The sequence encompassing the 3′ end of this exon is shown together with the 3′ end of cDNA clone DK9. (F) The intron-exon borders flanking *icb*-6 were determined. Downstream of *icb*-6 only one large 3′ exon is present in a 4.5-kb *Eco*RI fragment (MF2.5E4.5) of phage MF2.5 (Fig. 1B), the 5′ border of which is shown. The exon flanking *icb*-6 upstream is located in a 2.7-kb *Eco*RI fragment (MF1.7E2.7) of phage MF1.7 (Fig. 1B). The sequence of cDNA clone DK9 shows both exons joined together.

icb-6 were sequenced. This showed that the translocation breakpoints occur between codons 349 and 350, almost at the C terminus of the *dek* ORF (Fig. 5 and 3F). The intron containing the breakpoints on chromosome 6 was termed *icb*-6. Fusion of *dek* and *can* via the introns *icb*-9 and *icb*-6 results in transcription of a chimeric mRNA in which the ORFs of *dek* and *can* are merged without disruption of their original reading frames. As a result, the DEK-CAN protein has a predicted molecular mass of 165 kDa.

Analysis of the *dek* gene. Hybridization of *dek* cDNA

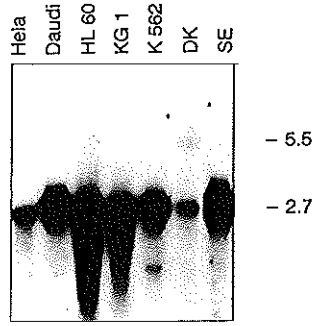


FIG. 4. Northern blot containing 20 μg of total RNA extracted from the cell lines HeLa, Daudi, HL60, KG1, and K562 and from bone marrow cells from t(6;9) AML patient DK and AUL patient SE. The sizes of the mRNAs hybridizing to the *dek* cDNA probe DK14 are indicated in kilobases.

probes to the previously isolated λEMBL3 phages occurred 3′ of *icb*-6, to only one apparently continuous stretch of genomic DNA. As the restriction maps of this genomic DNA and the 3′ *dek* cDNA are colinear, it is likely that only a single 1.6-kb 3′ exon is present downstream of *icb*-6. To further substantiate this point, the 3′ end of this exon was sequenced and was found to be colinear with *dek* cDNA down to the poly(A) tail (Fig. 3E). However, a tiny intron in this exon can be excluded only by complete sequencing of the exon.

Hybridization of 5′ *dek* cDNA probes to genomic phages indicated that the 5′ end of the *dek* gene was not contained within them. Therefore, a genomic λEMBL3 library was screened with a 270-bp *Eco*RI-*Hin*cII fragment derived from the 5′ end of cDNA clone DK14. Phages (24) were isolated and characterized. Two overlapping phages that spanned the largest stretch of DNA were phage MF3.19 and MF3.20 (Fig. 1B). MF3.20 contains five *Bss*HII sites mapping close together in a 1.4-kb *Hin*cII fragment (MF3.20 Hc1.4). MF3.20Hc1.4 is the most 5′ fragment hybridizing to the 270-bp *Eco*RI-*Hin*cII cDNA probe. Fine mapping and subsequent sequence analysis showed that the most 5′ cDNA sequences are present in a 240-bp *Bss*HII fragment which is preceded by a region rich in G/C (Fig. 3D).

*Bss*HII sites are mostly clustered in CpG islands which often appear to represent promoter sequences (7). Therefore, this region in *dek* may well encompass the promoter area. Moreover, linking of a 520-bp genomic fragment, mapping immediately 5′ of the cDNA homologous sequence, to a chloramphenicol acetyltransferase reporter gene showed that this DNA region contains a strong promoter activity (results not shown).

The 3′ end of MF3.19 did not overlap with the most 5′ end of the previously isolated phages (MF1.2); a gap is still present in the map. Long-range mapping analysis (unpublished results) indicated that the distance between the translocation breakpoint of the chromosome from patient DK in *icb*-6 and the *Bss*HII sites is approximately 30 kb. As 27 kb of DNA between the *Bss*HII sites and the breakpoint of the chromosome from patient DK is present in the genomic phages, the gap is estimated to measure only a few kilobases.

Analysis of the *can* and *dek* cDNA sequences. Comparison of the predicted amino acid sequences of both *can* and *dek* with the EMBL data base failed to reveal any substantial

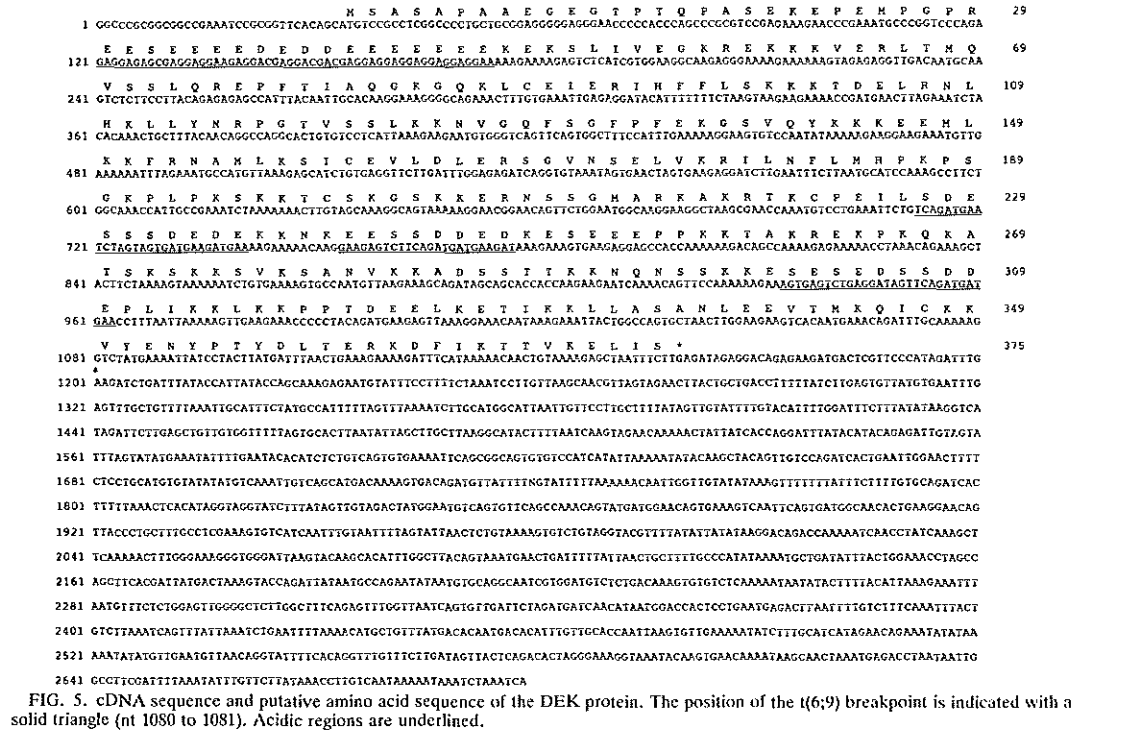


FIG. 5. cDNA sequence and putative amino acid sequence of the DEK protein. The position of the t(6;9) breakpoint is indicated with a solid triangle (nt 1080 to 1081). Acidic regions are underlined.

homology to known protein sequences. However, detailed analysis identified some structures that may have functional significance. In CAN, from amino acids (aa) 736 to 774, N terminal of *icb*-9, a leucine zipper motif is located. The leucine repeat consists of L-740, L-747, I-754, L-761, and L-768 and may represent a protein-protein dimerization domain. Projected on a helical wheel (Fig. 6A), hydrophobic residues at position +1 (relative to the leucines), acidic residues at position −1, and basic residues at position +2 are present. These residues may stabilize the formation of protein dimers through additional electrostatic interactions between the leucine repeats of proteins.

Immediately C terminal of the breakpoint in CAN from aa 811 to 887 two putative amphipathic helices are present and are separated by a loop of 25 residues, of which 13 are charged either positively (7) or negatively (6). The hydrophobic backbone of the first amphipathic helix is formed by I-814, L-817, V-821, A-824, V-828, and V-831 (Fig. 6A). The second amphipathic helix contains a heptad leucine repeat consisting of L-861, I-868, L-875, and L-882 (Fig. 6A). A region encompassing the C-terminal part of the loop and the C-terminal amphipathic helix (aa 840 to 887) shows homology to the human estrogen receptor (ER) dimerization domain: 30% of the residues is identical, 57% similar (Fig. 6B). The C-terminal 22 aa of this homology region in the ER were shown to be essential for the formation of ER homodimers (38).

Many SP or TP and SS, ST, TS, or TT dimers are present both N terminal and C terminal of the amino acid stretch containing the putative leucine zipper and amphipathic helices (Fig. 6A). This sequence motif has been proposed to have an ancillary role in DNA binding. At the C-terminus there is a recurrence of phenylalanine residues often in combination with S/T-P or S/T-S/T dimers.

In the predicted DEK protein no specific structures could be recognized apart from a continuous stretch of acidic residues at the N terminus, three acidic regions interspersed with serines, and a very high overall percentage (42%) of charged amino acids (H, R, K, E, and D).

Expression of *dek* and *can*. The expression patterns of *dek* and *can* in different mouse tissues may give a clue to the possible function of these genes. Twenty micrograms of total RNA of bone marrow, spleen, thymus, brain, liver, kidney, testes, ovary, placenta, and whole embryos 10, 13, 16, and 19 days after conception was loaded on a denaturing agarose gel. Hybridization of *dek* and *can* cDNA probes to hamster- and mouse-derived hybrid cell lines showed that both genes are conserved between species (unpublished results). Thus, the human *dek* cDNA clone DK14 and *can* cDNA clones hXT37 and hXT56 were used to screen for mouse *dek* and *can* transcripts. As shown in Fig. 7, *dek* and *can* are expressed in all tissues. *dek* is expressed at a relatively high level, while *can* seems to have a more restricted expression pattern. *can* expression was easily detected in RNA of thymus, spleen, bone marrow, kidney, brain, and testes but was hardly visible in all other tissues or in whole embryos during development.

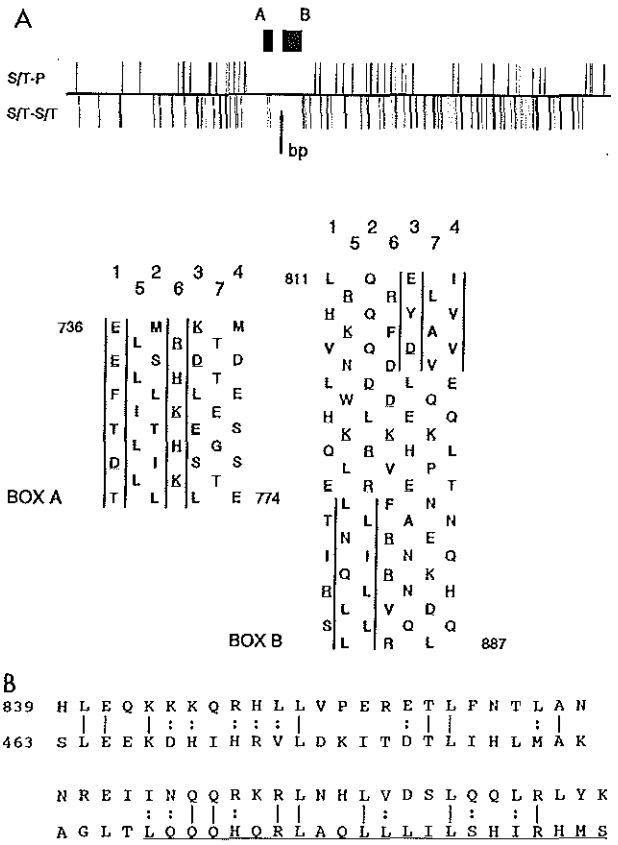


FIG. 6. (A) Domains of the putative CAN protein that may have functional significance. The top line represents the 2,192-aa putative CAN protein. The positions of S/T-P and S/T-S/T dimers are indicated with vertical lines. Box A (aa 736 to 775) represents a putative leucine zipper. Box B (aa 811 to 887) represents two amphipathic helices separated by a region of charged amino acids. An arrow indicates the position of the translocation breakpoint between boxes A and B. The amino acid sequences of boxes A and B are given underneath the CAN protein domains. The first four amino acid residues are written in a horizontal row and the next three are placed below and between them. In this way, the sequence can be read as an helical wheel, cut open at one side. Charged amino acids are underlined, and hydrophobic residues are in boldface type. Vertical lines indicate hydrophobic or charged sides of the predicted helical structure. (B) A part of the predicted CAN protein sequence present in box B (shown above) is homologous to the human ER. The homologous sequences are aligned. Identical amino acids are indicated by vertical lines, and similar amino acids are indicated by colons. The C-terminal 22 aa of the ER (underlined) are essential for ER homodimerization.

DISCUSSION

A novel fusion gene is present in leukemic cells carrying t(6;9) (p23;q34). The translocation breakpoints on chromosome 9 occur in one intron of the *can* gene, *icb*-9. Translocation breakpoints on chromosome 6 occur in one intron of the *dek* gene, *icb*-6. As a result of the translocation, a *dek-can* fusion gene encoding a chimeric *dek-can* transcript is generated. The sequence of this chimeric cDNA predicts it to encode a 165-kDa DEK-CAN protein.

Although the precise position of the breakpoints in *icb*-9 and *icb*-6 may vary, the same exons of *dek* and *can* are joined by splicing of the primary transcript of the fusion gene. The invariable *dek-can* transcript can be used as a marker of t(6;9) AML that can be sensitively monitored by the polymerase chain reaction (44). This may be a great advantage for diagnosis, monitoring of response to chemotherapy, and detection of minimal residual disease after bone marrow transplantation.

If steady-state levels of *dek-can* and *dek* transcripts in bone marrow from patient DK are compared, it appears that *dek* mRNA is much more abundant than *dek-can* mRNA. The bone marrow from patient DK contains >90% leukemic cells, of which every cell contains one chromosome 6 and one chromosome 6p−. In this cell population, the overall number of alleles of the normal *dek* gene and the fusion gene are about equal, and both are driven by the *dek* promoter. Higher steady-state levels of *dek* mRNA could be due to a longer half-life of *dek* transcripts compared with that of *dek-can* transcripts. Alternatively, enhancer sequences which are involved in transcription activation could be

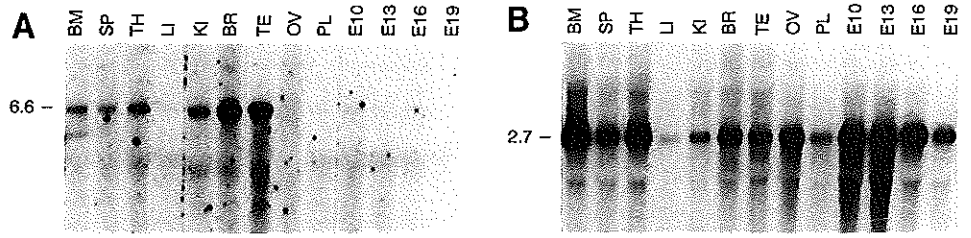


FIG. 7. Northern blot containing total RNA of various mouse tissues hybridized to *can* cDNA probe hXT37 and hXT56 (A) and to *dek* cDNA probe DK14 (B). The size of the transcript is indicated in kilobases. Abbreviations: BM, bone marrow; SP, spleen; TH, thymus; LI, liver; KI, kidney; BR, brain; TE, testes; OV, ovary; PL, placenta (13 days after conception); E10, E13, E16, and E19, embryos aged 10, 13, 16, and 19 days after conception, respectively.

present at the 3′ side of the *dek* gene. The enhancer would be removed from the fusion gene by the translocation.

The cellular function of DEK and CAN and the way DEK-CAN may interfere with normal hematopoiesis are still obscure. Neither of the two genes shows expression that is confined to the hematopoietic system. In fact, screening of a Northern blot containing RNA samples of different mouse tissues showed that *dek* is expressed ubiquitously. *can* is also expressed in all tissues, though at much lower and more variable levels. The tissues expressing *can* at a relatively high level include spleen and bone marrow. Since *can* mRNA is also found in human hematopoietic cell lines, it is unlikely that, because of the translocation, ectopic expression of *can* in hematopoietic cells would directly be involved in leukemogenic transformation. More likely, replacement of the N-terminal part of CAN by DEK sequences generates a protein that has different properties and is involved in transformation.

A breakpoint in *can* (*icb-9*) was also demonstrated in bone marrow cells from an AUL patient (SE) with an apparently normal karyotype (47). However, no breakpoint could be found in *dek*. In concordance with this observation, an aberrant transcript of 5.5 kb, detected by 3′ *can* probes in bone marrow RNA from this AUL patient, failed to hybridize to 5′ *dek* probes. cDNA cloning results strongly suggest that in this patient, *can* forms a fusion gene in which the 5′ sequences are derived from another, as yet unknown gene (46a). Therefore, it is possible that the C-terminal part of CAN contains domains involved in the leukemogenic process that may be activated by different N-terminal moieties.

Preliminary protein localization data were obtained by immunocytochemistry with antibodies directed against DEK and CAN and COS cells transiently expressing high levels of CAN, DEK, or DEK-CAN protein. CAN appears to be mainly cytoplasmic, while DEK has a strictly nuclear localization. The fusion of DEK to CAN results in a protein with a nuclear localization (19a).

In view of these data, the analysis of the *can* cDNA sequence revealed some structures that may be indicative for its function.

(i) An amphipathic helix with a heptad leucine repeat is predicted by the sequence just 5′ of *icb-9*. This leucine zipper motif has been detected in many proteins such as FOS, JUN, GCN4, and CCAAT/enhancer binding protein (10, 50), in which it mediates the formation of either homo- or heterodimers. A basic stretch of amino acids, juxtaposed to the leucine zipper, can function as a DNA binding element. In CAN no basic region is present adjoining the leucine zipper, and hence this helix most likely functions as a dimerization domain. Not only the addition of novel sequences to the 3′ part of *can* but also the removal of the original 5′ part of the gene may contribute to the putatively tumorigenic properties of the *dek-can* fusion gene. As the leucine zipper is detached from the C-terminal CAN sequences by the translocation, it is tempting to speculate that this structure may be the interaction site for a factor that could regulate CAN activity.

(ii) The protein sequence just 3′ of *icb-9* predicts two amphipathic helices separated by a stretch of 25 amino acids, containing many charged residues. Several arguments suggest that this domain may function in protein dimerization. (a) The C-terminal putative helix and part of the preceding charged amino acids show homology to the hormone binding region of the human and mouse ERs (23). It has been shown that the mouse ER contains a strong dimerization domain adjoining the hormone binding domain (18, 38). The entire sequence containing both domains is conserved within the steroid receptor family. At the N-terminal side of this domain, the homology between the ER and CAN extends beyond the homology between the ER and other steroid hormone receptors. However, CAN has no homology to the hormone binding domain immediately C-terminal of the dimerization domain. It is interesting that the homologous protein domain in another member of the steroid hormone receptor family, the retinoic acid receptor type α, was shown to dimerize with multiple cell-type-specific proteins which have not yet been characterized. Dimerization increased the affinity of the receptor for its cognate binding sequence (22). In addition, homology of CAN to the ER is noteworthy with regard to the finding that the retinoic acid receptor type α is involved in t(15;17) in acute promyelocytic leukemia (2, 8, 15). It will be interesting to analyze whether CAN can form heterodimers with the ER or other members of the steroid hormone receptor family. (b) Although no direct homology is present, the putative structure of CAN just C-terminal of *icb-9* (aa 811 to 887) architecturally resembles aa 82 to 162 of transcription factor AP-4, a basic stretch-helix-loop-helix protein (30). This part of AP-4 contains an additional dimerization domain, which, like this region in CAN, consists of two amphipathic helices separated by a stretch of 28 aa, containing many charged residues (30).

(iii) Many SP and TP dimers are present both N-terminal and C-terminal of the region containing the putative leucine zipper and amphipathic helices. A proline preceded by a serine or a threonine forms a β turn I, which is stabilized by formation of hydrogen bonds between the serine or threo-

nine and the backbone of 2 aa following the proline (45). A β turn I conformation can also be assumed by serine or threonine dimers. S/T-P dimers are clustered around DNA binding domains of many proteins that associate with DNA in a sequence-specific manner. Suzuki (45) proposes that the S/T-P-X-X (X for any amino acid) motif will bind in the minor groove of DNA in a sequence-independent manner. This may stabilize a specific interaction of the DNA binding motif in the major groove.

In the C-terminal cluster of S/T-P and S/T-S/T dimers in CAN, the aromatic residue phenylalanine is often recurring. The C-terminal part of RNA polymerase II of both the yeast *Saccharomyces cerevisiae* and mammals contains a SPTSPSY repeat (3, 13), which is essential for its function (4). Suzuki (46) argues that the β turn I-X-Y motif may be essential for DNA binding and shows that the aromatic ring of the tyrosine residue in this repeat can intercalate into the DNA. In *Drosophila* RNA polymerase II, tyrosine is replaced by another aromatic residue, phenylalanine (6). A structure of β turns combined with aromatic residues is therefore postulated to be a novel type of DNA binding domain. In the 3′ part of CAN a S/T-S/T/P-X-F sequence occurs 14 times. We will study whether this region has DNA binding capacity, either by itself or by stabilizing DNA binding domains of transcription factor complexes that contain the CAN protein.

The predicted protein sequence of DEK contains a remarkably high percentage of charged amino acids. At the N terminus (aa 30 to 47), DEK contains a continuous stretch of acidic residues. Three other acidic stretches are present, from aa 227 to 236, 241 to 248, and 301 to 310. They contain acidic residues interspersed only by serine residues. Acidic regions were found mainly in two types of nuclear proteins (17). (i) Chromatin-associated proteins such as nucleolin and high mobility group proteins contain acidic regions that can interact with the basic domains of histones (36, 37). These proteins also contain a conserved DNA binding domain, the high mobility group box, a sequence motif that is not present in DEK. (ii) A class of transcriptional activators, among which are herpes simplex virus VP16 protein and the yeast transcription factor GCN4, contain an acidic patch that can interact with the RNA polymerase II complex (9, 29).

Many basic amino acids are present in the DEK protein next to the acidic regions. The calculated pI of DEK is 8.9. Because of these basic stretches, several putative nuclear localization signals can be recognized. DEK is completely devoid of hydrophobic stretches.

We speculate that replacement of N-terminal CAN sequences by almost the entire DEK protein may activate the transforming potential of CAN. However, the mechanism of this putative activation remains to be determined. Analysis of the primary structure of DEK and CAN combined with the preliminary localization data suggests that these proteins may have a function in the cell nucleus.

Up to now, breakpoints of three different translocations in myeloid leukemia have been cloned and molecularly analyzed. Thus far the formation of fusion genes seems to be the predominant effect of translocations in myeloid leukemia.

ACKNOWLEDGMENTS

We thank D. Bootsma and D. Meijer for continuous support and discussion; J. Abels, H. Adriaansen, A. Hagemeijer, and B. Löwenberg for patient material; and M. Kuit and T. de Vries Lentsch for photographic work. We are indebted to A. Coulson (MRC, Edinburgh) for help with computer homology searches.

This work was supported by the Dutch Cancer Foundation.

REFERENCES

1. **Adriaansen, H. J., J. J. M. van Dongen, H. Hooijkaas, K. Hählen, M. B. van't Veer, B. Löwenberg, and A. Hagemeijer.** 1988. Translocation (6;9) may be associated with a specific TdT-positive immunological phenotype in ANLL. Leukemia **2**:136–140.
2. **Alcalay, M., D. Zangrilli, P. P. Pandolfi, L. Longo, A. Mencarelli, A. Giacomucci, M. Rocchi, A. Biondi, A. Rambaldi, F. Lo Coco, D. Diverio, E. Donti, F. Griniani, and P. G. Pelicci.** 1991. Translocation breakpoint of acute promyelocytic leukemia lies within the retinoic acid receptor alpha locus. Proc. Natl. Acad. Sci. USA **88**:1977–1981.
3. **Allison, L. A., M. Moyle, M. Shales, and C. J. Ingles.** 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell **42**:599–610.
4. **Allison, L. A., J. K. Wong, V. D. Fitzpatrick, M. Moyle, and C. J. Ingles.** 1988. The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. Mol. Cell. Biol. **8**:321–329.
5. **Auffray, C., and F. Rougeon.** 1980. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
6. **Biggs, J., L. L. Searles, and A. L. Greenleaf.** 1985. Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II. Cell **42**:611–621.
7. **Bird, A. P.** 1986. CpG islands and the function of DNA methylation. Nature (London) **321**:209–213.
8. **Borrow, J., A. D. Goddard, D. Sheer, and E. Solomon.** 1990. Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. Science **249**:1577–1580.
9. **Brent, R., and M. Ptashne.** 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43**:729–736.
10. **Busch, S. J., and P. Sassone-Corsi.** 1990. Dimers, leucine zippers and DNA-binding domains. Trends Genet. **6**:36–40.
11. **Chirgwin, J. M., A. E. Przybyla, J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
12. **Chomienne, C., P. Ballerini, N. Balitrand, M. E. Huang, I. Krawice, S. Castaigne, P. Fenaux, P. Tiollais, A. Dejean, L. Degos, and H. de Thé.** 1990. The retinoic acid receptor alpha gene is rearranged in retinoic acid-sensitive promyelocytic leukemias. Leukemia **4**:802–807.
13. **Corden, J. L., D. L. Cadena, J. J. Ahearn, and M. E. Dahmus.** 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc. Natl. Acad. Sci. USA **82**:7934–7938.
14. **Daley, G. Q., R. Van Etten, and D. Baltimore.** 1990. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science **247**:824–830.
15. **de Thé, H., C. Chomienne, M. Lanotte, L. Degos, and A. Dejean.** 1990. The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor alpha gene to a novel transcribed locus. Nature (London) **347**:558–561.
16. **De Thé, H., C. Lavau, A. Marchio, C. Chomienne, L. Degos, and A. Dejean.** 1991. The PML-RARα fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RARα. Cell **66**:675–684.
17. **Earnshaw, W. C.** 1987. Anionic regions in nuclear proteins. J. Cell Biol. **105**:1479–1482.
18. **Fawell, S. E., J. A. Lees, R. White, and M. G. Parker.** 1990. Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor. Cell **60**:953–962.
19. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.

19a. **Fornerod, M., M. von Lindern, and G. Grosveld.** Unpublished data.

20. **Fourney, R. M., J. Miyakoshi, R. S. Day III, and M. C. Paterson.**

1988. Northern blotting: efficient RNA staining and transfer. Focus 10:5–6.
21. **Frohman, M. A., M. K. Dash, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85:8998–9002.
22. **Glass, C. K., O. V. Devary, and M. G. Rosenfeld.** 1990. Multiple cell type-specific proteins differentially regulate target sequence recognition by the alpha retinoic acid receptor. Cell 63:729–738.
23. **Green, S., P. Walter, V. Kumar, A. Krust, J. M. Bornert, P. Argos, and P. Chambon.** 1986. Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature (London) 320:134–139.
24. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene 25:263–269.
25. **Heisterkamp, N., G. Jenster, H. J. Ten Hoeve, D. Zovich, P. K. Pattengale, and J. Groffen.** 1990. Acute leukaemia in bcr/abl transgenic mice. Nature (London) 344:251–253.
26. **Hermans, A., J. Gow, L. Selleri, M. von Lindern, A. Hagemeijer, L. M. Wiedemann, and G. Grosveld.** 1988. bcr-abl oncogene activation in Philadelphia chromosome-positive acute lymphoblastic leukemia. Leukemia 2:628–633.
27. **Hermans, A., N. Heisterkamp, M. von Lindern, S. van Baal, D. Meijer, D. van der Plas, L. M. Wiedemann, J. Groffen, D. Bootsma, and G. Grosveld.** 1987. Unique fusion of bcr and c-abl genes in Philadelphia chromosome positive acute lymphoblastic leukemia. Cell 51:33–40.
28. **Hermans, A., L. Selleri, J. Gow, L. Wiedeman, and G. Grosveld.** 1989. Molecular analysis of the Philadelphia translocation in myelogenous and acute lymphoblastic leukemia. Cancer Cells 7:21–26.
29. **Hope, I. A., and K. Struhl.** 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46:885–894.
30. **Hu, Y. F., B. Luscher, A. Admon, N. Mermod, and R. Tjian.** 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. 4:1741–1752.
31. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1984. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. Glover (ed.), DNA cloning techniques: a practical approach. IRL Press, Oxford.
32. **Kadowaki, T., H. Kadowaki, and S. I. Taylor.** 1990. A nonsense mutation causing decreased levels of insulin receptor mRNA: detection by a simplified technique for direct sequencing of genomic DNA amplified by the polymerase chain reaction. Proc. Natl. Acad. Sci. USA 87:658–662.
33. **Kakizuka, A., W. H. Miller, K. Umesono, R. P. Warrell, S. R. Frankel, V. V. V. S. Murty, E. Dmitrovsky, and R. M. Evans.** 1991. Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RARα with a novel putative transcription factor, PML. Cell 66:663–674.
34. **Konopka, J. B., S. M. Watanabe, and O. N. Witte.** 1984. An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. Cell 37:1035–1042.
35. **Kozak, M.** 1988. A profusion of controls. J. Cell Biol. 107:1–7.
36. **Krohne, G., and W. W. Franke.** 1980. A major soluble acidic protein located in nuclei of diverse vertebrate species. Exp. Cell Res. 129:167–189.
37. **Lapeyre, B., H. Bourbon, and F. Amalric.** 1987. Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. Proc. Natl. Acad. Sci. USA 84:1472–1476.
38. **Lees, J. A., S. E. Fawell, R. White, and M. G. Parker.** 1990. A 22-amino-acid peptide restores DNA-binding activity to dimerization-defective mutants of the estrogen receptor. Mol. Cell. Biol. 10:5529–5531.
39. **Pearson, M. G., J. W. Vardiman, B. M. LeBean, J. D. Rowley, S. Schwartz, S. L. Kerman, M. M. Cohen, E. W. Fleischman, and E. L. Prigogina.** 1985. Increased numbers of marrow basophils may be associated with a t(6;9) in ANLL. Am. J. Hematol. 18:393–403.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Sandberg, A. A., R. Morgan, J. A. McCallister, M. B. Kaiser, and F. Hecht.** 1983. Acute myeloblastic leukemia (AML) with t(6;9) (p23;q34): a specific subgroup of AML? Cancer Genet. Cytogenet. 10:139–142.
42. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
43. **Shtivelman, E., B. Lifshitz, R. P. Gale, and E. Canaani.** 1985. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature (London) 315:550–554.
44. **Soekarman, D., M. von Lindern, S. Daenen, B. de Jong, C. Fonatsch, B. Heinze, C. R. Bartram, A. Hagemeijer, and G. Grosveld.** Submitted for publication.
45. **Suzuki, M.** 1989. SPXX, a frequent sequence motif in gene regulatory proteins. J. Mol. Biol. 207:61–84.
46. **Suzuki, M.** 1990. The heptad repeat in the largest subunit of RNA polymerase II binds by intercalating into DNA. Nature (London) 344:562–565.
46a. **von Lindern, M.** Unpublished data.
47. **von Lindern, M., A. Poustka, H. Lerach, and G. Grosveld.** 1990. The (6;9) chromosome translocation, associated with a specific subtype of acute nonlymphocytic leukemia, leads to aberrant transcription of a target gene on 9q34. Mol. Cell. Biol. 10:4016–4026.
48. **von Lindern, M., T. van Agthoven, A. Hagemeijer, H. Adriaansen, and G. Grosveld.** 1989. The human pim-1 gene is not directly activated by the translocation (6;9) in acute nonlymphocytic leukemia. Oncogene 4:75–79.
49. **Walker, L. C., T. S. Ganesan, S. Dhut, B. Gibbons, T. A. Lister, J. Rothbard, and B. D. Young.** 1987. Novel chimaeric protein expressed in Philadelphia positive acute lymphoblastic leukaemia. Nature (London) 329:851–853.
50. **Ziff, E. B.** 1990. Transcription factors: a new family gathers at the cAMP response site. Trends Genet. 6:69–72.

Chapter 4

Relocation of the carboxyterminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements

Maarten Fornerod[1], Judith Boer[1], Sjozèf van Baal[1], Martine Jaeglé[3], Marieke von Lindern[3,4], K Gopal Murti[2], Donna Davis[2], Jacqueline Bonten[1], Arjan Buijs[1] and Gerard Grosveld[1]

Departments of [1]Genetics and [2]Virology & Molecular Biology, St Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, Tennessee 38105, USA; [3]Department of Cell Biology and Genetics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands; [4]Research Institute for Molecular Pathology (IMP), Dr Bohrgasse 7, 1030 Vienna, Austria

Fusion genes encoding the 3' part of the *can* gene are implicated in two types of leukemia. The *dek-can* fusion gene is present in t(6;9) acute myeloid leukemia and the *set-can* fusion gene is present in one case of acute undifferentiated leukemia. In order to obtain leads towards the molecular basis of these diseases, we have studied the cellular localization of the DEK-CAN and SET-CAN fusion proteins and their normal counterparts. DEK-CAN and SET-CAN were localized exclusively in the nucleus, and also DEK and SET were found to be nuclear proteins. However, CAN was mainly located at the nuclear and cytoplasmic face of the nuclear envelope. This observation is in accordance with the presence of an amino acid repeat in the C-terminal part of CAN, common to the family of nucleoporins. The C-terminal part also contains a nuclear location domain as shown by deletion analysis. This domain may be important for the presence of CAN at the nucleoplasmic side of the nuclear envelope. The relocation of the carboxyterminal part of CAN due to DEK-CAN and SET-CAN may reinforce a nuclear function of the CAN protein.

Keywords: leukemia; fusion proteins; nucleoporins; translocations; *can* proto-oncogene

Introduction

The *can* gene, located on chromosome 9q34, is involved in at least two chromosomal rearrangements associated with human leukemia. First, translocation (6;9) is associated with a specific subtype of acute myeloid leukemia (AML) (Rowley and Potter, 1976). This subtype is characterized by poor response to therapy, young age at onset and a primary FAB classification of M1, M2 or M4 (Adriaansen *et al.*, 1988). It has been shown that the translocation event fuses a gene, *dek*, on chromosome 6p23 to the *can* gene in a head-to-tail manner (Von Lindern *et al.*, 1990; 1992a). A leukemia-specific *dek-can* transcript originates from the 6p-chromosome and is invariably found in leukemic cells carrying the (6;9) translocation (Soekarman *et al.*, 1992a,b). This fusion transcript encodes a chimeric DEK-CAN protein consisting of most of the entire DEK polypeptide, fused to the two-third C-terminal part of the CAN protein. The consistent presence of the same *dek-can* transcript in t(6;9) leukemic cells suggests a causative role of the translocation in this subtype of acute myeloid leukemia.

A second chromosome rearrangement found in one case of acute undifferentiated leukemia fuses *can* to *set*, located on chromosome 9q34 at the centromeric side of *can* and c-*abl* (Von Lindern *et al.*, 1992b). Whether the *set-can* fusion gene results from a (9;9) translocation or an inversion within the 9q34 band is unknown at present. The *can* sequences in the *set-can* fusion gene are identical to the ones found in the *dek-can* fusion gene, indicating *can* as the gene with oncogenic potential. Furthermore, in one case of refractory anemia with an excess of blast cells (RAEB), a subtype of myelodysplastic syndrome (MDS), and in a case of common acute lymphoblastic leukemia (c-ALL), a chromosomal breakpoint in the *can* gene was detected, while no breakpoint in *dek* (Soekarman *et al.*, 1992a) or *set* (unpublished results) was present. This indicates that, besides *dek* and *set*, at least one additional fusion partner of the *can* gene exists.

Chromosome translocations that fuse genes within their transcribed regions have been strongly associated with specific types of myeloid and lymphoid malignancies (Young, 1992) and solid tumors (Delattre *et al.*, 1992; Sorensen *et al.*, 1994). These translocations generate fusion proteins with deregulated properties compared to the individual fusion partners, which likely form the very basis of these malignancies. Therefore, study of these fusion proteins and their normal counterparts may provide important clues for the understanding of the mechanism of carcinogenesis, as well as normal cell function.

Not very much is currently known about the SET, CAN and DEK proteins. However, recently Adachi *et al.* (1994) reported that SET is a nuclear phosphoprotein. The DEK protein often acts as a major immunoreactive antigen in patients with auto-immune diseases such as juvenile rheumatoid arthritis or lupus erythemathosis (Sierakowski *et al.*, 1993). Furthermore, a cDNA encoding the last 65 amino acids of the DEK protein could confer partial complementation of the genetic instability of ataxia-telangiectasia fibroblasts of the complementation-group D (Meyn *et al.*, 1993). Relevance of this finding to the normal function of DEK, or its role in leukemogenesis, is presently unclear. Less is known about the CAN protein, the common fusion partner of DEK and SET. Expression of the protein may not be confined to the hematopoietic system since, in mouse tissues, *can* mRNA is found in kidney, brain, testis, bone marrow, thymus and spleen (Von Lindern *et al.*, 1992a).

Correspondence: G Grosveld
Received 18 January 1995; accepted 27 January 1995

This study is designed to establish and compare the subcellular localization of the CAN protein and its two chimeric derivations, DEK-CAN and SET-CAN. To this end, we have made cDNA constructs of the genes, expressed them in cells, and localized the respective proteins by immunofluorescence and immuno-electron microscopy using specific antisera or monoclonal antibodies that recognize epitope-tagged versions of the proteins. These studies reveal that CAN protein is associated with the nuclear envelope and, in particular, the nuclear pores, and is sometimes present in the nucleus; whereas the DEK-CAN and SET-CAN proteins localize always to the nucleoplasm. In addition to motifs in the primary sequence of the CAN protein that are typically present in nucleoporins, we identified a 227 aminoacid sequence in the carboxyterminal part of CAN sufficient for nuclear localization. We propose from our results that the nuclear localization of the CAN moiety is instrumental in the aberrant growth of *set-can* and *dek-can* positive leukemic cells. While this manuscript was in preparation, a publication appeared (Kraemer *et al.*, 1994) confirming our observation that the CAN protein is present in the nucleopore complex. However, in contrast to our data, these authors find CAN present only at the cytoplasmic side of the complex.

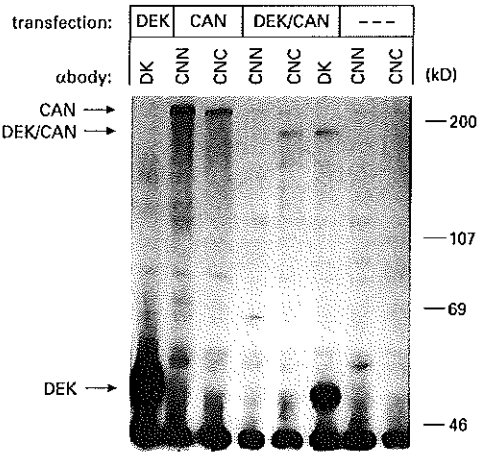


Figure 1 Immunoprecipitation of DEK, CAN and DEK-CAN protein. COS-1 cells were transfected with pCDXX-Dek, pCDXX-Dek-Can, pCDXX-Can or pCDXX-antisense-Can (- - -) and metabolically labelled with [^{3}H]leucine. Cell lysates were subjected to immunoprecipitation with antisera directed against DEK (αDK), the N-terminal part of CAN (αCNN), and the C-terminal part of CAN (αCNC). Proteins were separated on 6% SDS-polyacrylamide gels. Arrows mark the position of CAN, DEK-CAN and CAN proteins. Molecular weight standards were run in an adjacent lane; their masses (in kilodalton) are indicated

Results

Specificity of antisera

To investigate the nature of the DEK, CAN and SET proteins, as well as the DEK-CAN and SET-CAN fusion proteins, antisera were raised against an N-terminal part of CAN (αCNN), a C-terminal part of CAN (αCNC), and against an N-terminally truncated version of the DEK protein.

In order to test the specificity of the antibodies, *dek*, *can* and *dek-can* cDNAs were cloned into an SV40-based expression vector (pCDXX) that directs high levels of expression after transfection into COS-1 cells. For technical reasons, a mini-intron was positioned between the *dek* and *can* moieties of the *dek-can* cDNA construct. The three antisera were then used to immunoprecipitate ^{3}H-leucine-labelled DEK, CAN and DEK-CAN protein after transient transfection in COS-1 cells (Figure 1). As shown in lane 1, DEK antibodies precipitate a polypeptide of 52 kDa that migrates slightly faster than the rabbit IgG heavy chain [54–55 kDa (Yarmush *et al.*, 1981)]. CAN antibodies (lane 2 and 3) recognize a polypeptide of 220 kDa. The DEK-CAN chimeric protein is precipitated as a polypeptide of 180 kDa by the DEK antiserum (lane 6) as well as the CAN antiserum that recognizes the C-terminal part of the protein (lane 5), but not by antiserum raised against the N-terminal part of CAN (lane 4). The endogenous (i.e., COS-1 cell-derived) DEK and CAN polypeptides of 52 and 220 kDa are also precipitated from cells transfected with either the *dek-can* or an antisense *can* construct (lanes 6, 7 and 8), albeit in lower amounts. It is noteworthy that estimated molecular mobilities of DEK, DEK-CAN and CAN proteins after SDS–PAGE approximate the predicted sizes of 43, 165 and 220 kDa, respectively. The somewhat lower mobility of DEK and DEK-CAN is presumably due to the high content of charged amino acids present in the DEK polypeptide (43%). Thus, *dek*, *can* and *dek-can* cDNA constructs direct synthesis of the correct proteins in COS-1 cells and proteins produced are recognized specifically by the DEK and CAN antisera.

In contrast with the COS cell experiments, we have been unable to detect the CAN and DEK-CAN proteins in immunoprecipitations from metabolically-labelled primary bone marrow cells of t(6;9) AML patients. The proteins may be below detection levels because *can* and *dek-can* mRNA levels are very low in these cells (von Lindern *et al.*, 1990), and the antibodies are raised against denatured protein, which limits their effectiveness in immunoprecipitation experiments.

Phosphorylation of DEK and CAN

It has been shown that SET is a phosphoprotein (Adachi *et al.*, 1994). To determine if DEK and CAN share a similar post-translational modification, we examined whether DEK or CAN polypeptides were phosphorylated, by metabolic labelling of transiently transfected COS-1 cells with [^{32}P]orthophosphate. In a parallel experiment, COS-1 cells were metabolically labelled with [^{3}H]leucine. Immunoprecipitation of DEK and CAN from labelled cell lysates (Figure 2) was performed with anti-DEK and anti-CAN antisera, the respective pre-immune sera serving as negative controls. From both ^{3}H- and ^{32}P-labelled lysates, the DEK-specific antiserum precipitates a protein of 52 kDa not present in the pre-immune lanes, showing that DEK is a phosphoprotein. The DEK amino acid sequence contains two potential phosphorylation sites for cAMP-dependent protein kinase, 16 for casein kinase

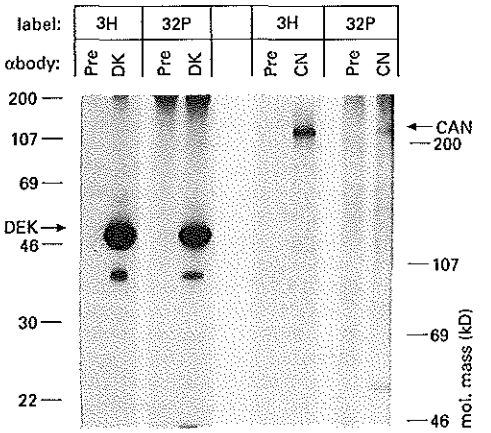


Figure 2 Identification of DEK and CAN as phosphoproteins. COS-1 cells were transfected with pCDXX-Dek or pCDXX-Can and metabolically labelled with either [^{3}H]leucine or [^{32}P]orthophosphate. Cell lysates were subjected to immunoprecipitation with antisera directed against DEK (αDK), or a mixture of antisera against the N- and C-terminal part of CAN (αCN). Proteins were separated on 11% (DEK) or 6% (CAN) SDS–PAGE. Lanes labeled with 'Pre' represent immunoprecipitations with the respective pre-immune sera. Arrows mark positions of DEK and CAN proteins. Molecular weight standards were run in an adjacent lane; their masses (in kilodalton) are indicated

II and 20 for protein kinase C. A second phosphoprotein of approximately 40 kDa is also recognized by the anti-DEK antiserum. The same protein is also detected in Western blot experiments (data not shown), indicating that it represents a truncated version of the DEK protein, possibly generated by specific degradation or internal start codon usage. The CAN-specific antiserum precipitates labelled protein of 220 kDa from both ^{3}H- and ^{32}P-labelled cell lysates not present in the pre-immune precipitation, indicating that CAN too, is a phosphoprotein. CAN contains various potential phosphorylation sites: one for cAMP-dependent protein kinase, 23 for casein kinase II, 30 for protein kinase C, and one (Y^{877}) for tyrosine kinase. Which of the sites serve as phosphorylation targets, and the functional implications of these phosphorylations will be subject of future studies.

Immunofluorescence analysis

Since there are no cell lines available from t(6;9) leukemia patients and the leukemic cells are hard to obtain due to the rare occurrence of the disease, we decided to study the subcellular localization of DEK, CAN, SET, DEK-CAN and SET-CAN in transfected and normal HeLa cells. The results obtained with HeLa cells are identical to those with COS-1 cells, NIH3T3 cells and the chronic myeloid leukemia derived K562 cells. Since SET, DEK and CAN are ubiquitously expressed proteins, localization data obtained with these cell lines may closely represent the situation in leukemia cells of t(6;9) patients. Only results obtained with HeLa cells are presented here and consist of a compendium of a large series of experiments in which the effects were studied of different levels of expression of these proteins on their localization.

After transfection, the level of protein expression in individual cells varies considerably; therefore, the only reliable way to roughly estimate the level of expression is to visually compare the relative fluorescence signal with that in untransfected cells. Although the method is not quantitative, it allows us to make the distinction between low, moderate and high protein expression in cells.

In normal HeLa cells, the anti-DEK antibodies reveal DEK to be a nuclear protein (Figure 3A) during interphase. In mitotic cells, however, the protein is dispersed throughout the cell body, but a distinct and easily visible fraction of DEK co-localizes with the condensed chromosomes. Figure 3B shows a cell in metaphase; Figure 3C, a cell in late telophase. Anti-CAN antisera αCNC (Figure 3D) and αCNN (not shown) display a weak punctuated staining of the

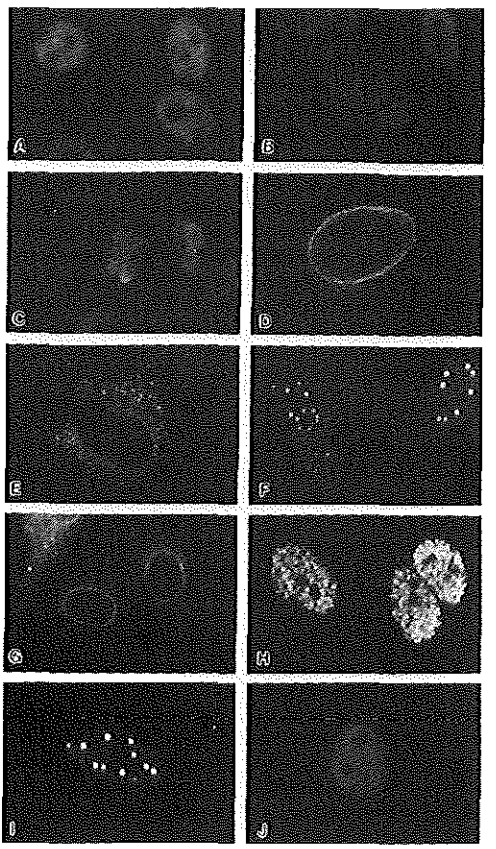


Figure 3 Subcellular distribution of DEK, CAN, DEK-CAN, SET-CAN and SET proteins. (A–D) Normal HeLa cells labelled with DEK-specific antiserum αDK (A–C) or CAN-specific antiserum αCNC (D). (E–G) HeLa cells transfected with pCDXX-HA1-Can expressing high (E–F) or low (G) amounts of HA1-tagged CAN protein, detected with monoclonal antibody 12CA5. (H–I) cells transfected with pCDXX-Dek-Can (H) or pCDXX-Set-Can (I). DEK-CAN or SET-CAN protein was detected with CAN-specific antiserum αCNC. (J) HeLa cells transfected with pSCTOP-HA1-Set and expressing HA1-tagged SET protein, detected with antibody 12CA5. The presence of first antibodies was visualized by a FITC-labelled second antibody

nuclear envelope in normal HeLa cells. To further examine the localization of the CAN protein in the nuclear envelope and to increase the sensitivity of detection, two copies of the HA1 hemagglutinin epitope were fused to the N-terminus of the CAN protein. The resulting HA1-CAN protein was transiently expressed in HeLa cells. In immunofluorescence analysis, this tagged version of the CAN protein could be detected by monoclonal antibody 12CA5, as well as by the anti-CAN antisera, αCNN and αCNC. Due to low background fluorescence, small amounts of HA1-CAN protein were more readily detected by the monoclonal antibody than by the anti-CAN antisera (not shown). In cells that express low amounts of the protein, the CAN signal is mainly confined to the nuclear membrane (Figure 3G), indicating that this is indeed the CAN primary location. However, in cells that show high expression of HA1-CAN, the protein was predominantly present in the cytoplasm (Figure 3E) or, in a minority of overexpressing cells (approximately 5–10%), mainly in the nucleus (Figure 3F). Independent of its subcellular localization, highly expressed HA1-CAN protein accumulates in aggregated masses of unknown identity. Identical results were obtained using expression constructs encoding the normal CAN protein, immuno-detected by polyclonal antisera, indicating that the intracellular distribution of highly expressed CAN is not an artefact created by addition of the HA1 sequences. In contrast to the CAN protein, the DEK-CAN (Figure 3H) and SET-CAN (Figure 3I) fusion proteins, and their HA1-tagged equivalents (not shown), have a strictly nuclear localization in HeLa cells, regardless of their level of expression. Using anti-CAN antisera, also a weak nuclear-rim staining is present in these cells (not visible in micrographs 3H and 3I). Inside the nucleus, the fusion proteins seem to be unevenly distributed. The DEK-CAN protein is present in many small granular patches, whereas the SET-CAN protein is more evenly distributed, apart from a few (2–10) rather large dots, resembling those seen in the nuclei of CAN overexpressing cells (Figure 3F). The significance of these protein masses, if any, is not understood at present. In order to establish the subcellular localization of the SET protein, DNA sequences encoding the HA1 tag were introduced at codon 12 of the *set* cDNA. Transiently-transfected HeLa cells having low-to-moderate expression of the HA1-SET construct resulted in a staining strictly confined to the nucleus, as shown in Figure 3J. At high expression levels, the protein was detected throughout the cell (data not shown).

Influence of DEK-CAN expression on the normal CAN protein

Since leukemic cells of t(6;9) patients express mRNA of the normal as well as the translocated allele of the CAN gene (Von Lindern *et al.*, 1990), we were interested if expression of DEK-CAN in HeLa cells

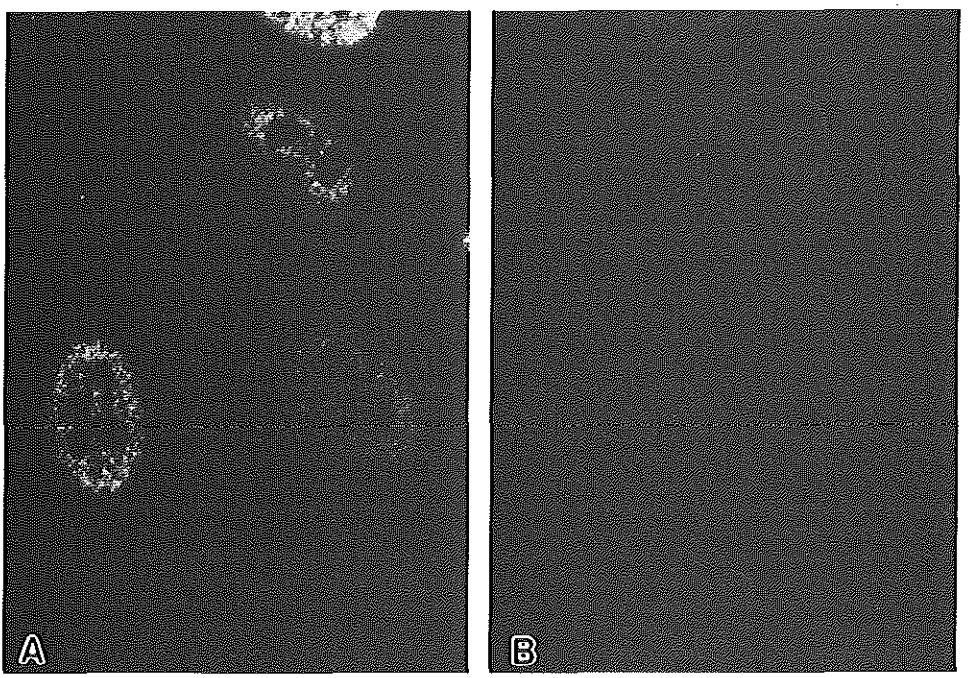


Figure 4 Simultaneous subcellular localization of CAN (red) and DEK-CAN (green). Cell line C4322 (see Material and methods) in which HA1 tagged DEK-CAN expression is tetracycline repressible, was grown in the absence (A) or presence (B) of 2.5 $\mu g\,ml^{-1}$ tetracycline for 3 days prior to immunofluorescent analysis (as described in Materials and methods). In both cells CAN is detected with antiserum αCNN, directed against an N-terminal part of the CAN protein (not present in DEK-CAN), followed by Texas Red coupled goat-anti-rabbit second antibody. HA1 tagged DEK-CAN is detected by monoclonal antibody 12CA5 followed by FITC coupled goat anti-mouse second antibody

would influence the localization of the normal CAN protein. To this end, cell line C4322 was created, carrying the HA1 tagged *dek-can* mini-gene under the control of the tetracycline repression system (Gossen and Bujard, 1992), and CAN localization was monitored with released versus repressed DEK-CAN expression. As seen in Figure 4, expression of DEK-CAN had no effect on CAN localization, suggesting that the effect of the fusion protein in the cell is dominant rather than dominant negative over the normal function of CAN. In addition, this experiment shows that essentially no DEK-CAN protein is present in the nuclear envelope.

Immuno-electron microscopy

To confirm that CAN is mainly localized in the nuclear envelope and to identify the region of the nuclear envelope with which CAN is associated at high resolution, immuno-electron microscopic studies were conducted on CE490S cells. This cell line was created by introducing the *can* cDNA, under the control of the heptamerized tet-operator (UHD-Can), into HeLa derived cell line HtTA-1 (Gossen and Bujard, 1992). In the absence of tetracycline, CAN expression in individual cells varies from moderate to low and is mainly located in the nuclear envelope, as observed by immunofluorescence (data not shown). Cells were sectioned and incubated with a mixture of pre-adsorbed (see Materials and methods) rabbit polyclonal antisera (αCNN and αCNC) recognizing the CAN protein, followed by incubation with goat anti-rabbit antibodies conjugated with 10 nm colloidal gold particles. The micrographs in Figure 5 show the pattern of distribution of the gold particles at the level of the nuclear envelope. In Figure 5A, a large stretch of the nuclear envelope is shown, with the outer and inner layers visible. In this image, the gold particles are seen predominantly at the inner surface of the nuclear envelope. Figures 5B–D illustrate the gold label at the level of the nuclear pore. The label appears to be at the outer (Figure 5B), middle (Figure 5C) or inner (Figure 5D) aspect of the nuclear pore. Occasionally, the gold-label is seen near the cytoplasmic face of the nuclear pore and presumably associated with the outer fibrillar component of the nuclear pore complex (Figure 5E). That the label is predominantly at the nuclear pores is confirmed by tangential sections of the nuclear envelope (Figure 5F). In these sections, the nuclear pores are visible as circular openings measuring approximately 60 nm, and some of these are decorated with gold particles. Quantitation of the gold particles (Table I) substantiates our observation that the nuclear envelope/pore complexes are predominant sites where CAN is located. Of the total number of gold particles within the nuclear envelope, 93 are observed at the outer and 65 at the inner region, usually in the vicinity of the nuclear pore complexes.

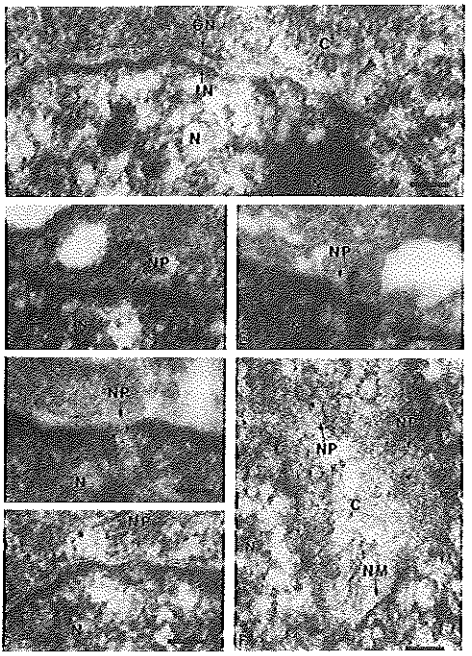


Figure 5 Immuno-electron microscopic localization of the CAN protein in CE490S cells. (A) Section through the nucleus (N), showing the outer (ON) and inner (IN) nuclear membrane and nucleus (Nu); C, cytoplasm. (B–D) Three sections through a nuclear pore (NP), showing gold particles at the cytoplasmic side (B), in the middle (C) or at the nucleoplasmic side (D) of a nuclear pore (NP). (E) Cross section showing gold particles decorating the cytoplasmic side of a nuclear pore. (F) Tangential section through the nuclear envelope. Nuclear pores (NP, arrows) appear as ring-like structures, two of which are decorated with gold particles; NM, nuclear membrane. Bar represents 100 nm

Similar results with this cell line are obtained using the αCNN antiserum only (Table 1), while incubation with pre-immune sera gives background levels (Table 1). Sections of normal HtTA-1 cells incubated with preadsorbed αCNN+αCNC gives a signal in the nuclear envelope that is slightly above background

Table 1 Distribution of gold particles in immuno-electron microscopic analysis of cell lines CE490S and HtTA-1

		Gold particles					*Gold particles/μM²*		
Cell line	*Antiserum*	*C*	*N*	*NE In*	*NE Out*	*NE Total*	*C*	*N*	*NE*
CE490S	αCNN, αCNC	230	243	65	93	158	0.7	0.8	9.7
CE490S	αCNN	240	161	34	33	67	1.6	1.4	11.4
HtTA-1	αCNN, αCNC	230	152	10	15	25	1.0	0.8	3.1
CE490S	Pre-immune CNN, CNC	117	128	3	5	8	0.6	0.8	1.0

The number and position of gold particles were determined from 5–12 independent representative thin sections of cell line CE490S or parental cell line HtTA-1 that had been incubated with CAN-specific antisera or pre-immune sera, followed by 10 nm colloidal gold-labelled anti-rabbit IgG. C, cytoplasm; N, nucleoplasm; NE, nuclear envelope; IN, inner face; OUT, outer face

(Table 1). These data indicate that the signal at both sides of the nuclear envelope in the first experiment is not due to detection of other proteins than CAN.

The CAN sequence contains nucleoporin motifs

The C-terminal part of CAN contains 45 copies of a motif that is reminiscent of the recurrent motifs found in nucleoporins. Although they resemble the XFXFG and GLFG repeats (Panté and Aebi, 1993), they appear to vary and deviate from these consensus sequences, having only the FG dipeptide in common. Amino acid residues around this dipeptide show marked preferences and are depicted in Table 2. Of these preferred amino acids it is noteworthy that recurrent presence of glutamine residues at position +2 is also seen in GLFG repeats of NUP100 and NUP116 (Wente *et al.*, 1992). In addition to the C-terminal FG repeat, spanning one-fifth of the amino acid sequence, 11 copies of an FXF motif (mostly FSF, some included in the C-terminal FG repeat, see Figure 6A) are present. Three are situated at the amino-terminal side of the t(6;9) breakpoint. The function of the FXF and FG repeats is not known, therefore the significance of amino acid preferences remains to be established. In addition, two regions are present in the CAN protein, predicted to form coiled-coils (Lupas *et al.*, 1991; Sedgwick and Boder, 1991). The first coiled-coil region is located between amino acids 702 and 728, ending only a few amino acids N-terminal of a putative

Table 2 Amino acid composition of the C-terminal FG repeat in CAN

Position	*Most frequent amino acids*
−3	S, T, A, V, G
−2	S, T, P, G, F
−1	A, V, L, S, T
0	F
+1	G
+2	S, T, Q

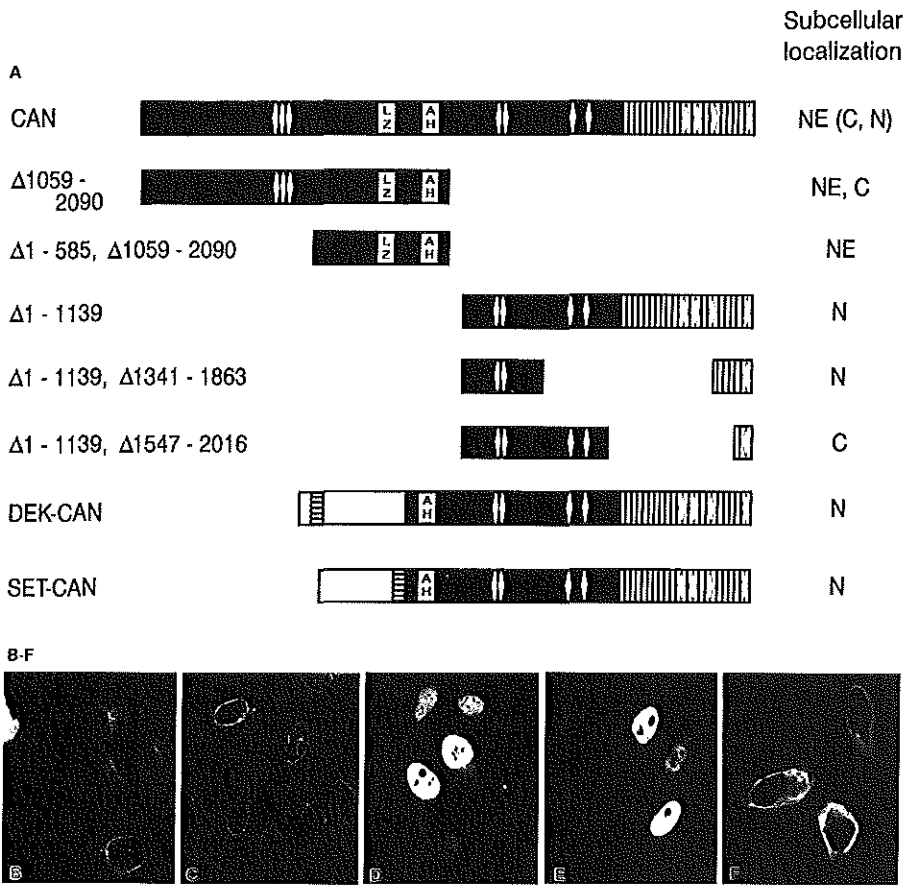


Figure 6 (A) Schematic representation of structural characteristics and subcellular localization of the CAN protein and deletion mutants, as well as of the DEK-CAN and SET-CAN fusion proteins. FG repeats are depicted as vertical bars, FXF repeats as diamonds. Horizontal stripes indicate acidic regions in the DEK and SET proteins. LZ: leucine zipper domain. AH: amphipathic helix. On the right the subcellular localization is given: NE, nuclear envelope; C, cytoplasm; N, nucleus. Localization of CAN between brackets signifies subcellular localization detected upon overexpression. (B–F) Immune fluorescence of mutant HA1-CAN proteins transiently expressed in HeLa cells, (B) Δ1059–2090, (C) Δ1-585, Δ1059–2090, (D) Δ1-1139, (E) Δ1-1139, Δ1341–1863, (F) Δ1-1139, Δ1547–2016

leucine zipper region (Figure 6A), previously described (Von Lindern *et al.*, 1992a). The second coiled-coil region is located between amino acids 860 and 887 and partly coincides with a second previously described putative amphipathic helix (Figure 6A).

Subcellular localization of CAN mutant proteins

Because the nuclear localization of normal CAN protein in some cells overexpressing CAN could either reflect a true aspect of the protein or an artifact, we decided to analyze deletion mutants to see if the observed distinct localizations of the protein can be attributed to functions of different domains in the CAN polypeptide. Mutant proteins were made, all carrying an amino-terminal HA1-tag for detection of protein in transfected HtTA-1 cells with the monoclonal antibody 12CA5. First, amino and carboxyl terminal truncated versions were tested, representing proteins comprising amino acids 1–1058 (Δ1059–2090) and 1140–2090 (Δ1–1139), respectively (Figure 6A). The Δ1059–2090 mutant protein is found in the nuclear membrane, and cytoplasm (Figure 6B), especially upon higher expression, similar to the full-length CAN protein. However, no aggregation in the cytoplasm was observed and, even under conditions of high overexpression, the protein never occurred in the nucleus. Additional deletion of amino acids 1–585 (Figure 6A) locates the protein almost exclusively to the nuclear envelope (Figure 6C). This observation indicates that sequences needed for association with the nuclear pore are contained within amino acids 586–1058. In sharp contrast, the Δ1–1139 mutant is always nuclear, whether expressed at low or high levels (Figure 6D). This shows that the carboxyl terminal half of the protein travels to the nucleus either by itself or complexed to factors that carry it along. Additional deletions superimposed on the Δ1–1139 mutant (Figure 6D–F) further map the area of the protein important for nuclear import, which is conferred by a region narrowed down to residues 1863–2090 (Figure 6A). Precise mapping of the amino acid sequences responsible for nuclear localization, as well as nuclear pore retention of CAN, will not be addressed in this paper and will be reported in another study.

Discussion

Our studies address the subcellular localization of the proteins DEK, CAN and SET and their putatively oncogenic derivatives DEK-CAN and SET-CAN, generated by gene rearrangements occurring in specific types of leukemia. All these proteins were found specifically in the nuclear compartment, except for CAN. Immunofluorescence experiments showed that CAN is predominantly located in a punctuated rim around the nucleus in cells expressing relatively small amounts of the protein. In addition, the C-terminal part of the protein contains amino acid repeats that show resemblance to those found in nucleoporins. Immuno-electron microscopic studies reveal that CAN is located in the nuclear envelope at the nuclear pores. These results identify CAN as a nucleoporin and are in agreement with earlier findings of Kraemer *et al.* (1994). The electron-microscopic data presented by these authors suggest that the CAN protein is exclusively present at the cytoplasmic side of the nucleopore complex; however, our studies show immunogold label at the inner and outer aspects of the nuclear envelope. The fact that we detected the nuclear component could be due to moderate overexpression of the protein, which may increase its appearance at the inner side of the nuclear envelope. Detection of CAN at the nuclear side may reflect a bona fide function of the protein that in normal HeLa cells is below the level of detection and, therefore, could not be noticed in the study of Kraemer *et al.* (1994). Alternatively, our result may be explained in terms of limits of resolution of the indirect immunogold-labelling technique used. Considering that each antibody molecule measures about 10 nm (Webster *et al.*, 1978), the distance between the antigen and the gold particle could be 24 nm or more (Tokuyasu, 1980; Murti *et al.*, 1985). This would make it difficult to determine from the position of the gold label, whether the antigen is at the inner or outer face of the nuclear membrane. However, since Kraemer *et al.* (1994) do not find gold-label at the nucleoplasmic site of the nuclear membrane using the same technique, this possibility is less likely.

A second indication of a nuclear function of CAN is the presence of a 227 amino acid region at the C-terminus of the protein containing sequences that act as a signal for nuclear localization, when sequences that associate CAN to the nuclear pore have been deleted. As this region has no homology to any of the known nuclear localization sequences, the routing mechanism of CAN to the nucleus remains to be defined. In addition, our deletion analysis positively identifies isolated domains in CAN that can confer different cellular localization to the protein. This modular structure suggests that if nuclear pore association could be regulated by posttranslational modification of CAN, it would determine the levels of nuclear component of the protein. Since CAN is found to be a phosphoprotein, an attractive model would be that its nuclear function may be regulated by phosphorylation of specific amino acid residues. At present we have no data supporting this idea, but experiments addressing this question are in progress.

Interestingly, HeLa cells highly overexpressing CAN display nuclear or cytoplasmic localization of the protein, in addition to its normal association with the nuclear envelope. As heterologous expression of the nucleoporin p62 is known to result in spill-over to the cytoplasm only (Starr *et al.*, 1990; Carmo-Fonseca *et al.*, 1991), nuclear accumulation does not seem to be a general feature of overexpressed nucleoporins, further supporting a possibly regulated nuclear role for CAN.

The presence of the fusion proteins DEK-CAN and SET-CAN in the nucleoplasm may be due to nuclear localization domains in DEK and SET, both nuclear proteins. In addition, loss of one or two predicted coiled-coil regions could be of importance in routing the fusion proteins to the nucleoplasm. Similar coiled-coil regions have been described in nucleoporins p62 (Starr *et al.*, 1990), NSP1 (Hurt, 1989), and NSP49 (Wimmer *et al.*, 1992), and these putative protein interaction domains have been suggested to play a role in anchoring nucleoporins to the nuclear pore complex (Hurt, 1990). This would agree with our deletion

analyses that show that sequences located between amino acid 586 and 1058 of CAN, including both predicted coiled-coil regions, are needed for retention in the nucleopore complex.

Since CAN is the common partner in DEK-CAN and SET-CAN, the oncogenic potential of these fusion proteins may be due to disruption or alteration of the normal cellular functions of the CAN protein. Our immunofluorescence studies indicate that DEK-CAN expression does not influence the subcellular localization of normal CAN suggesting that DEK-CAN acts in a dominant, rather than a dominant-negative, fashion. The dominant property of DEK-CAN and SET-CAN may be mediated by the constitutive presence of the carboxyterminal part of CAN in the nucleoplasm, permanently performing a nuclear function of CAN that normally would be highly controlled. Circumvention of this control may be the key feature of their oncogenic subversion.

The role of DEK and SET sequences in the fusion proteins remains to be determined but, since the translocation with the *dek* gene is recurrent, DEK is likely to provide additional qualities needed for oncogenic conversion. The observation that DEK is associated with metaphase chromosomes may be relevant and could indicate that DEK-CAN is directly involved in transcriptional regulation of certain genes.

Materials and methods

Expression constructs

Expression vector pCDXX is a derivative of pCD-X (Okayama and Berg, 1983) harboring a partial double pUC19 polylinker (*Pst*I to *Eco*RI, *Eco*RI to *Bam*HI) between the SV40 early promoter region and poly A signal (a gift of Dr N Heisterkamp, University of Southern California). pCDXX-Dek carries the *Eco*RI fragment of cDNA clone DK14 (Von Lindern *et al.*, 1992a). pCDXX-HA1-Can carries the entire open-reading frame of can, which was compiled of cDNA clones hXT23, hXT37, hXT54 and hXT65 (Von Lindern *et al.*, 1992a) as an *Eco*RI fragment. pCDXX-HA1-Can is complemented at the 5′ site with DNA encoding two copies of the flu virus HA1 epitope. The *Eco*RI fragment of pCDXX-Dek-Can is compiled from cDNA clones DK14, hXT37, hXT54, hXT65 and genomic clones MF1.7 (*dek*) and A11F7 (*can*) (Von Lindern *et al.*, 1992a). A 215 bp intron, consisting of 130 bp of 5′ icb-6 sequences (MF1.7) linked to 85 bp of 3′ icb-9 sequences (A11F7) was positioned at the *dek/can* fusion point. pCDXX-Set-Can is compiled from cDNA clones SE10 (Von Lindern *et al.*, 1992b), hXT37, hXT54, hXT65 and genomic clones Sg22 (Von Lindern *et al.*, 1992b) and A11F7. A 175 bp intron, consisting of 90 bp of the 5′ part of the last intron of *set* (Sg22) linked to 85 bp of 3′ icb-9 sequences (A11f7) was positioned at the *set/can* fusion point. Expression vector pSCTOP, a derivative of pSCT-GAL-X556 (Rusconi *et al.*, 1990), harbors the pBluescript polylinker (Stratagene, La Jolla, CA) from *Sal*I to *Sac*I between the human cytomegalovirus promoter and splice/poly(A) region. This vector expressed SET protein tagged with the HA1 epitope (pSCTOP-HA1-Set). UHD-Can contains the normal *can* cDNA cloned into the multiple cloning site of pUHD10-3 (Gossen and Bujard, 1992). pUHD10S is a derivative of pUHD10-3, in which the plasmid backbone is replaced by that of p6WtkCAT (Schöler *et al.*, 1989), thereby positioning two copies of the SV40 bi-directional transcriptional stop (2533 to 2770 of the SV40 genome) upstream of the tet operators, preventing readthrough from upstream promoters. pHA1-Dek-Can carries the HA1 tagged version of the *dek-can* semi-cDNA in pUHD10S. Deletion mutants of CAN were made using the HA1 tagged version of the *can* cDNA in pUHD10S. HA1-CanΔ1059–2090 was made by insertion of a double stranded oligo containing a stopcodon into the NheI site at position 3263 of the can cDNA sequence. The inserted sequence is 5′-GCTAGTTAATTAATTAACTAGC-3′. HA1-CanΔ1-585, Δ1059–2090 was made from HA1-CanΔ1059–2090 by fusion the filled-in *Nco*I site present in the artificial sequences following the HA1-tags to the *Hinc*II site at position 1847 in the *can* cDNA. HA1-CanΔ1–1139 was generated by an in-frame deletion using the *Nco*I site following the HA1 tags and the *Nco*I site at position 3507 of the *can* cDNA. HA1-CanΔ1-1139, Δ1341–1863 was made through an inframe deletion of HA1-Can Δ1–1139 using MseI sites at positions 4111 and 5680 in the *can* cDNA. HA1-CanΔ1-1139, Δ1547–2016 was made through an in-frame deletion of HA1-CanΔ1-1139, using *Pst*I sites at positions 4731 and 6141 of the *can* cDNA.

Cell culture and transfection

COS-1 (Gluzman, 1981), HeLA and HtTA-1 (Gossen and Bujard, 1992) cells were grown in DMEM containing 8% fetal calf serum. COS-1 cells were transiently transfected using the DEAE-dextran method on subconfluent 10 cm dishes as described (Meijer *et al.*, 1990). For transient transfection, 2×10^5 HeLa or HtTA-1 cells were transfected on microscope slides using the calcium phosphate precipitation method (Graham and Eb, 1973). For the generation of cell line CE490S, HtTA-1 cells were co-transfected with linearized UHD-Can plasmid and a linearized plasmid containing the hygromycin B resistance gene under control of the Herpes simplex virus tk promoter and Polyoma virus EPyF441 enhancer, pGEM-Hyg (Van Deursen *et al.*, 1991), using the calcium phosphate precipitation method. Twenty-four hours post-transfection, resistant cells were selected in 400 $\mu g\ ml^{-1}$ hygromycin B (Calbiochem, La Jolla, CA) in the presence of 2.5 $\mu g\ ml^{-1}$ tetracycline (FisherBiotech, Fair Lawn, NY). The same techniques were used for generation of cell line C4322, only the cotransfected plasmids were pHA1-Dek-Can and a plasmid containing the puromycin acetyl transferase (*pac*) gene (Vara *et al.*, 1986) under control of the rat β-actin promoter (Morgenstern and Land, 1990). Cells were selected in 0.5 $\mu g\ ml^{-1}$ puromycin (Sigma Co., St. Louis, MO), again in the presence of 2.5 $\mu g\ ml^{-1}$ tetracycline.

Antisera

Polyclonal antisera were obtained from rabbits (Flamish giants) injected with N- and C-terminal parts of the CAN protein, overproduced in *E. coli* BL21:DE3(pLysS) using the T7 polymerase system (Studier *et al.*, 1990). For N-terminal antiserum (αCNN), protein representing amino acids 164–804 of the CAN protein (Von Lindern *et al.*, 1992a) was used, for C-terminal antisera (αCNC), protein representing amino acids 960–1911. The t(6;9) breakpoint is located between amino acids 812 and 813. Antiserum against DEK (αDK) was raised against an N-terminally truncated (amino acids 68-375) DEK protein that was overproduced in *Spodoptera frugiperda* Sf21 insect cells infected with recombinant baculovirus (Smith *et al.*, 1983). Protein from *E. coli* and baculovirus-infected Sf21 cells was isolated using preparative SDS–PAGE and electro-elution.

Epitope tagging

N-terminally modified versions of the CAN and SET proteins were made by cloning a synthetic sequence at the 5′ end of each of the cDNAs, encoding two consecutive flu virus hemagglutinin (HA1) epitopes (YDVPDYASL) (Wilson *et al.*, 1984). For *can* and *dek-can*, the following sequence was placed 5′ of the start codon: 5′-GAATTC GCCGCCACC **ATG TAT GAC GTC CCA GAT TAC GCA AGT TTG CCC GGG TAT GAT GTT CCT GAT TAT GCT AGC CTC CCG GGT TAC GAT GTG CCC GAC AAT GCC TCT CTT CCT GGC GAG** CTC GCC ATG GGC AGC AGC *CAT CAT CAT CAT CAT CAC* AGC AGC GGC CTG GTG CCG CGC GGC AGC CAT ATT TCT CGT CGT GCA TCT GTT CAT ATG (codon 1 of CAN or DEK) 3′, while the *set* cDNA was complemented with the following sequence: 5′-GAATTCGCCGCCACC **ATG TAT GAC GTC CCA GAT TAC GCA AGT TTG CCC GGG TAT GAT GTT CCT GAT TAT GCT AGC CTC CCG GGT TAC GAT GTG CCC GAC AAT GCC TCT CTT CCT GGC GAG** (codon 12 of set) 3′. Nucleotides encoding the HA1 tag are marked in bold. In addition, to the CAN and DEK-CAN N-termini, a 6 × His tag (italics) and a heart muscle kinase recognition site (Li *et al.*, 1989) (underlined) were added.

Immunoprecipitation

Cells were labelled with 600 μCi [^{3}H]leucine (143 Ci mmol^{-1}, Amersham Corp., Arlington Heights, IL) for 16 h or with 600 μCi [^{32}P]orthophosphate (Amersham Corp.) for 5 h in leucine or phosphate free DMEM supplemented with 10% dialyzed fetal calf serum. Forty to 50 h after transfection, cells were rinsed twice with cold PBS and then lysed in RIPA buffer: 50 mM Tris-HCl pH 8; 1% Nonidet P-40; 0.5% deoxycholate; 0.1% SDS; 150 mM NaCl; 0.1% 2-mercaptoethanol and protease inhibitors (1 mM para-methyl sulfonyl fluoride (PMSF) and 0.02 trypsin inhibitor units (TIU) per ml of Aprotinin). In phosphorylation experiments phosphatase inhibitors (50 mM NaF, 10 mM Na_2MoO_4) were added. Radiolabelled proteins were immunoprecipitated using standard techniques.

Indirect immunofluorescence

Forty hours after transfection, cells attached to microscope slides were washed twice with PBS, fixed for 15 min in 3% paraformaldehyde in PBS, washed three times with PBS, and incubated for 10 min in 50 mM NH_4Cl in PBS. After washing three times with PBS, cells were permeabilized in 0.2% Triton X-100 in PBS for 10 min. Cells were then washed three times with PBS containing 1% nonfat milk (Korger Co., Cincinnati, OH) and incubated for 45 min with the first antibodies. Anti-CAN antisera were preabsorbed with 0.4 mg ml^{-1} bacterial competitor-protein in PBS/milk in order to reduce background. Bacterial competitor protein consisted of sonicated and boiled *E. coli* BL21:DE3(pLysS) cell extract. Anti-DEK and anti-CAN antisera were diluted 1/200 in PBS/milk. Monoclonal antibody 12CA5 (Boehringer-Mannheim Corp., Indianapolis, IN) was diluted to a concentration of 5 μg ml^{-1} in PBS/milk. Bound antibody was visualized using FITC- (Sigma Co., St. Louis, MO) or Texas Red- (United Stated Biochemical, Cleveland, OH) conjugated second antibodies. Images were obtained with either conventional (Figure 3) or confocal (Figure 4 and 6) immunofluorence microscopy. Confocal microscopy was performed using a BioRad MRC1000 Laser Scanning confocal microscope (BioRad, Hercules, CA).

Immuno-electron microscopy

Forty hours after release from tetracycline repression, subconfluent dishes of cell line CE490S were rinsed twice with PBS, scraped and fixed in suspension with 1% glutaraldehyde in PBS for 10 min at room temperature. After centrifugation for 10 min at 1000 *g*, the pellets were dehydrated three times for 15 min in 70% ethanol. This was followed by 15 min incubations with 2:1 and 1:1 mixtures of 70% ethanol and the hydrophilic resin LR White Hard Grade (London Resin Company Ltd., Basingstoke, England) at room temperature. Next, undiluted resin was added to the cell pellets and allowed to infiltrate three times for 20 min at 50°C. The resin was polymerized by addition of 0.15% (v/v) LR White 'accelerator' for 5 min at 0°C, followed by overnight incubation at room temperature *in vaccuo*. Thin sections were collected on nickel grids and blocked with 1% gelatin in PBS, and incubated for 1 h with a mixture of preabsorbed antiserum αCNN and αCNC, αCNN alone or respective pre-immune sera diluted 1:100 in a 1:1 mixture of PBS and immunobuffer (PBS supplemented with 500 mM NaCl, 0.05% Tween 20 and 0.1% fishgel). After extensive washes with immunobuffer, grids were incubated for 1 h in a suspension of 10 nm colloidal gold, coated with anti-rabbit antibody (Amersham Corp., Arlington Heights, IL), and diluted 1:20 in immunobuffer. After the final wash, the grids were contrasted by staining in 4% aqueous uranyl acetate for 1.5 min. Samples were visualized with a JEOL JEM-1200EX II electron microscope (Japanese Electron Optics USA, Inc, Peabody, MA). For quantitative analysis, gold particles were counted in 12 representative cells. The area of nucleus and cytoplasm was calculated by weight of paper cutouts of corresponding prints. The area of the nuclear envelope was calculated by multiplication of its total length by its average thickness, which was estimated to be 60 nm.

Acknowledgements

We thank Prof D Bootsma for continuous support and discussion, Eric Bonten for help with the baculovirus expression system, Janneke van Denderen and Reier Hoogendoorn for help in obtaining DEK and CAN antibodies, Rob Willemsen for technical support in initial immunofluorescence experiments. Monique van der Knaap for technical assistance, Jan van Deursen for critically reading and Peggy Burdick for editing the manuscript. We are grateful to Prof Dr H Bujard for providing us with cell line HtTA-1 and plasmid UHD10-3. These studies were supported in part by Cancer Center CORE Grant CA-21765 and by the American Lebanese Syrian Associated Charities (ALSAC) of St Jude Children's Research Hospital. A significant part of the work of authors MF, JB, SVB, AB and GC was performed at the Department of Cell Biology and Genetics, Erasmus University Rotterdam, The Netherlands. This work was supported by a grant of the Dutch Cancer Foundation.

References

Adachi Y, Pavlakis GN and Copeland TD. (1994). *J. Biol. Chem.*, **269**, 2258–2262.

Adriaansen HJ, *et al.* (1988). *Leukemia*, **2**, 136–140.

Carmo-Fonseca M, Kern H and Hurt EC. (1991). *Eur. J. Cell Biol.*, **55**, 17–30.

Delattre O, *et al.* (1992). *Nature*, **359**, 162–165.
Gluzman Y. (1981). *Cell*, **23**, 175–182.
Gossen M and Bujard H. (1992). *Proc. Natl. Acad. Sci. USA*, **89**, 5547–5551.
Graham FL and Eb AJ. (1973). *Virology*, **52**, 456–467.
Hurt EC. (1990). *J. Cell Biol.*, **111**, 2829–2837.
Hurt EC. (1989). *J. Cell Sci. Suppl.*, **12**, 243–252.
Kraemer D, Wozniak WR, Blobel G and Radu A. (1994). *Proc. Natl. Acad. Sci. USA*, **91**, 1519–1523.
Li BL, Langer JA, Schwartz B and Pestka S. (1989). *Proc. Natl. Acad. Sci. USA*, **86**, 558–562.
Lupas A, Van Dyke M and Stock J. (1991). *Science*, **252**, 1162–1164.
Meijer D, *et al.* (1990). *Nucl. Acids Res.*, **18**, 7357–7365.
Meyn MS, Lu-Kuo JM and Herzing LBK. (1993). *Am. J. Genet.*, **53**, 1206–1216.
Mogenstern JP and Land H. (1990). *Nucleic Acids Res.*, **18**, 1068.
Murti KG, Portner A, Troughton K and Deshpande K. (1985). *J. Electron Microsc. Tech.*, **2**, 139–146.
Okayama H and Berg P. (1983). *Mol. Cell. Biol.*, **3**, 280–289.
Panté N and Aebi U. (1993). *J. Cell Biol.*, **122**, 977–984.
Rowley JD and Potter D. (1976). *Blood*, **47**, 705–721.
Rusconi S, *et al.* (1990). *Gene*, **89**, 211–221.
Schöler HR, *et al.* (1989). *EMBO J.*, **8**, 2551–2557.
Sedgwick RP and Boder E. (1991). *Handbook Clin. Neurol.*, **16**, 347–423.
Sierakowski H, Williams KR, Szer IS and Szer W. (1993). *Clin. Exp. Immunol.*, **94**, 435–439.
Smith GE, Summers MD and Fraser MJ. (1993). *Mol. Cell. Biol.*, **3**, 2156–2165.
Soekarman D, *et al.* (1992a). *Blood*, **79**, 1–8.
Soekarman D, *et al.* (1992b). *Leukemia*, **6**, 489–494.
Sorensen PHB, *et al.* (1994). *Nature Genet.*, **6**, 146–151.
Starr CM, D'Onofrio M, Park MK and Hanover JA. (1990). *J. Cell. Biol.*, **110**, 1861–1871.
Studier FW, Rosenberg AH, Dunn JJ and Dubendorff JW. (1990). *Methods Enzymol*, **185**, 60–89.
Tokuyasu KT. (1980). *Histochem. J.*, **12**, 381–405.
Van Deursen J, *et al.* (1991). *Nucl. Acids Res.*, **19**, 2637–2643.
Vara JA, Portela A, Ortín J and Jiménez A. (1986). *Nucl. Acids Res.*, **14**, 4617–4624.
Von Lindern M, *et al.* (1992a). *Mol. Cell. Biol.*, **12**, 1687–1697.
Von Lindern M, *et al.* (1992b). *Mol. Cell. Biol.*, **12**, 3346–3355.
Von Lindern M, Poustka A, Lerach H and Grosveld G. (1990). *Mol. Cell. Biol.*, **10**, 4016–4026.
Webster RE, Osborn M and Weber K. (1978). *Exp. Cell Res.*, **117**, 47–61.
Wente SR, Rout MP and Blobel G. (1992). *J. Cell Biol.*, **119**, 705–723.
Wilson IA, *et al.* (1984). *Cell*, **37**, 767–778.
Wimmer C, *et al.* (1992). *EMBO J.*, **11**, 5051–5061.
Yarmush ML, Gates FT, Dreher KL and Kindt TJ. (1981). *J. Immunol.*, **126**, 2240–2244.
Young BD. (1992). *Bailliers Clin. Haematol.*, **5**, 791–964.

Chapter 5

Interaction of Cellular Proteins with the Leukemia Specific Fusion Proteins DEK-CAN and SET-CAN and Their Normal Counterpart, the Nucleoporin CAN

Maarten Fornerod, Judith Boer, Sjozèf van Baal, Hans Morreau[1] and Gerard Grosveld

Department of Genetics, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, Tennessee, USA, and [1]Department of Pathology, Leiden University Hospital, P.O. Box 9600, 2300 RC Leiden, The Netherlands.

Oncogene **13**, 1801-1809.

The recurrent chromosomal translocation (6;9) is associated with acute myeloid leukemia and results in expression of the DEK-CAN fusion protein. This oncoprotein consists of almost the entire DEK protein fused to the C-terminal two thirds of the CAN protein. In much the same way, CAN is fused to SET in a patient with acute undifferentiated leukemia, producing a SET-CAN fusion protein. Interestingly, CAN is associated with the nuclear pore complex (NPC), and we recently established its crucial role in nucleocytoplasmic transport processes and cell cycle progression. As a first step in the biochemical analysis of the oncogenic mechanism associated with translocation (6;9), we set out to identify proteins that interact with CAN and its fusion proteins. We found that two proteins specifically co-immunoprecipitate with CAN. One had a molecular mass of 88 kDa protein (CC88), and was determined to associate with the central region of CAN that contains several protein interaction motifs. A second protein of 112 kDa (CC112) was found to interact with the C-terminal nucleoporin-specific repeat of CAN, a region that is supposed to function in nucleocytoplasmic transport. CC112 also interacts with the DEK-CAN and SET-CAN fusion proteins. This finding suggests that CC112 may contribute an essential function to the leukemogenic effect of DEK-CAN and SET-CAN.

Introduction

The recurrent translocation (6;9)(p23;q34) is associated with human myeloid leukemia of diverse phenotype (Rowley & Potter, 1976; Soekarman *et al.*, 1992). It is often the sole karyotypic abnormality found in the leukemic cells, which suggests a causative role for this translocation in the disease. The translocation creates a *dek-can* fusion gene on the 6p-

chromosome (Von Lindern *et al.*, 1992a; Von Lindern *et al.*, 1990), that is predicted to encode a structurally identical 165 kDa DEK-CAN chimeric protein in all patients analyzed. This protein contains all but the 26 C-terminal amino acids of DEK linked to the C-terminal two-thirds of CAN. In addition, we found that a cryptic translocation or inversion of chromosome 9 in a patient with acute undifferentiated leukemia, created a chimeric protein comprising the same sequences of CAN but linked to N-terminal sequences of SET (Von Lindern *et al.*, 1992b).

The CAN protein was found to belong to a family of nuclear pore complex associated proteins called nucleoporins (Fornerod *et al.*, 1995; Kraemer *et al.*, 1994). This group of proteins is thought to play an important role in transport of RNA and protein across the nuclear membrane (for review see Görlich & Mattaj, 1996; Rout & Wente, 1994). In HeLa cells, CAN is found generally at the fibrillar, cytoplasmic extrusions of the nuclear pore complex (NPC) (Kraemer *et al.*, 1994; Panté *et al.*, 1994), but is also detected at the nuclear face upon over-expression (Fornerod *et al.*, 1995, J.B. and G.G., unpublished results). CAN is also referred to as NUP214 (Kraemer *et al.*, 1994) or p250 (Panté *et al.*, 1994).

Recently, we created a CAN knockout mutation in the mouse (van Deursen *et al.*, 1996) and found that CAN-/- embryos survive until day 4 of gestation by using maternal, oocyte-derived protein. At this point in development, the dwindling amounts of CAN become critical, resulting first in a G2 arrest of embryonic cells, followed by a detectable decrease in nuclear localization signal (NLS)-mediated protein import that coincides with a block in mRNA export. Although these data firmly establish a role for CAN in nucleocytoplasmic transport across the nuclear pore complex, they shed little light on CAN's contribution to leukemogenic transformation.

Presently the cellular function of the DEK protein is not known. It is a nuclear protein that may be associated with chromatin (Fornerod *et al.*, 1995). The SET protein is also a nuclear protein and was recently identified as a replication factor TAF-1 (Nagata *et al.*, 1995). It appears to be part of a growing superfamily of proteins, including the yeast nucleosome assembly factor NAP1 (Nagata *et al.*, 1995; Von Lindern *et al.*, 1992b), that specifically interact with B-type cyclins (Kellogg *et al.*, 1995). How these features should be interpreted in the context of leukemogenesis remains to be established. DEK has no sequence homology with SET, apart from highly acidic regions that are present in both proteins.

The primary amino acid sequence of CAN contains several distinct motifs including 11 FXF-sequence repeats scattered throughout the molecule, and a 35x repeated FG sequence that is confined to the C-terminus (Von Lindern *et al.*, 1992a). Similar repeat sequences appear in several vertebrate and yeast nucleoporins, (for review see Rout & Wente, 1994), and it has been suggested that these repeats are involved in protein-protein interactions that mediate substrate transport through the NPC (Chi *et al.*, 1995; Görlich *et al.*, 1995; Moroianu *et al.*, 1995; Radu *et al.*, 1995a). In fact, *in vitro* studies have shown that the 97 kDa import factor p97/importin-ß

subunit of the nuclear protein import complex binds to the repeat-containing nucleoporins NUP98, NUP153, NUP358/RanBP2, and CAN (Moroianu *et al.*, 1995; Radu *et al.*, 1995b). CAN also contains a coiled-coil region and a leucine zipper, situated at the N-terminal side of the translocation breakpoint, and a second coiled-coil region that flanks the breakpoint at the C-terminal side. These two domains are part of the central region of CAN that is responsible for nuclear envelope association (Fornerod *et al.*, 1995), and therefore could mediate NPC interaction. Surprisingly, mutants of CAN lacking these domains are directed to the nucleus. The sequences mediating this relocation are contained within the C-terminal 227 amino acids of the molecule (Fornerod *et al.*, 1995).

The DEK-CAN and SET-CAN fusion proteins retain one of the central coiled-coil domains and the FG repeat sequences. In contrast to CAN, these fusion proteins are permanently located in the nucleus and therefore we hypothesized that permanent relocation of CAN sequences to the nucleus could contribute an essential component to their leukemogenic activity. Hence, proteins that would physically interact with this portion of CAN may convey or contribute to the transforming potential of the fusion proteins.

Recently, a second nucleoporin was found to be targeted by a recurrent translocation in acute myeloid leukemia. This translocation (7;11) creates a fusion between the FXFG-repeat containing part of nucleoporin NUP98 and the homeobox-containing C-terminal part of HoxA9 (Borrow *et al.*, 1996; Nakamura *et al.*, 1996). This discovery suggests that the mode of action of the CAN and NUP98 derived fusion proteins could be related and emphasizes the need for detailed molecular analyses of leukemia-associated nucleoporins.

Here, as a further step to addressing the contribution of CAN to leukemogenesis, we define proteins that co-immunoprecipitate with CAN. We also identify the domains in CAN that interact with these proteins. Only one of the proteins appears to bind to DEK-CAN and SET-CAN and we believe this protein may be involved in the transformation process of myeloid precursor cells.

Results

Co-immunoprecipitation of CAN interacting proteins

We performed co-immunoprecipitation studies to identify CAN, DEK-CAN or SET-CAN interacting proteins, that could be important for their function. Our polyclonal CAN antisera (Fornerod *et al.*, 1995) only effectively immunoprecipitated the protein in the presence of SDS (data not shown), thereby dissociating pre-existing protein complexes. We overcame this problem by using previously generated expression plasmids encoding influenza virus hemagglutinin (HA1)-tagged versions of CAN and DEK-CAN, (HA1-CAN and HA1-DEK-CAN) (Fornerod *et al.*, 1995). HA1-CAN and HA1-DEK-CAN were transiently expressed in

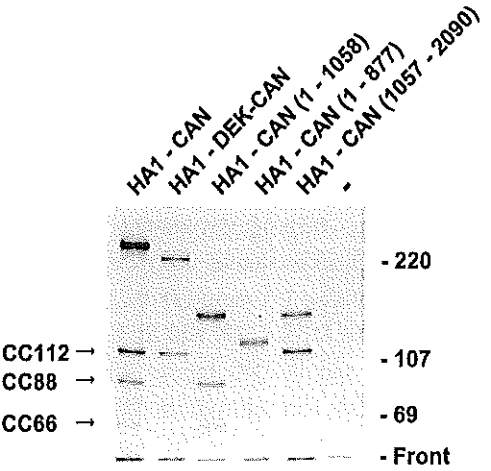


Figure 1 CAN co-immunoprecipitating proteins in HeLa derived HtTA-1 cells. Cells were transfected with expression plasmids encoding HA1 tagged versions of the CAN and DEK-CAN proteins or CAN deletion mutants as indicated above the lanes. Untransfected cells are represented the right lane (-). Following metabolic labelling with ^{3}H-leucine, proteins were immunoprecipitated using monoclonal antibody 12CA5 and analyzed on a 6% polyacrylamide gel. Co-precipitating

the HeLa derived cell-line HtTA-1 (Gossen & Bujard, 1992) and the proteins were subsequently immunoprecipitated with anti-HA1 monoclonal 12CA5 (Wilson *et al.*, 1984). As shown in Figure 1, two major proteins are co-precipitated with HA1-CAN from ^{3}H-leucine labeled cells; one of 112 and one of 88 kDa. A weaker band representing a protein of 66 kDa can also be observed. We named these proteins CC112, CC88 and CC66 (for CAN Co-precipitating protein) and found that only CC112 co-precipitated with HA1-DEK-CAN (Figure 1) and HA1-SET-CAN (not shown). These data suggest that CC112 interacts with the C-terminal part of CAN, whereas CC88 and CC66 require sequences that are either disrupted by the fusion or are no longer present in DEK-CAN and SET-CAN. To confirm this interpretation, we separately expressed HA1-tagged versions of the N-terminal half of CAN (aminoacids 1-1058), and the C-terminal half (amino acids 1059-2090). Indeed, the N-terminal half of CAN co-precipitated CC88 and CC66, while the C-terminal part of CAN co-precipitated only CC112 (Figure 1). No smaller co-precipitating proteins were detected in HtTA-1 cells (not shown).

To ensure that expression of these co-precipitating proteins was not limited to HtTA-1 cells, we performed a similar co-immunoprecipitation experiment using human U937 monoblastic cells, engineered to express HA1-CAN in a tetracycline dependent manner (cell line U937T-C7). We cultured these cells in the presence of different concentrations of tetracycline to establish distinct expression levels of HA1-CAN. As shown in Figure 2, CC112 and CC88 were also co-precipitated with HA1-CAN expressed in U937T-C7, indicating that, similar to CAN (Von Lindern *et al.*, 1992a), these proteins are expressed in multiple human cell types. The signal of ^{35}S-labelled CC88 was relatively weak, which we found is typical for ^{35}S-methionine labelling of this protein. Remarkably, the amount of co-precipitating CC88

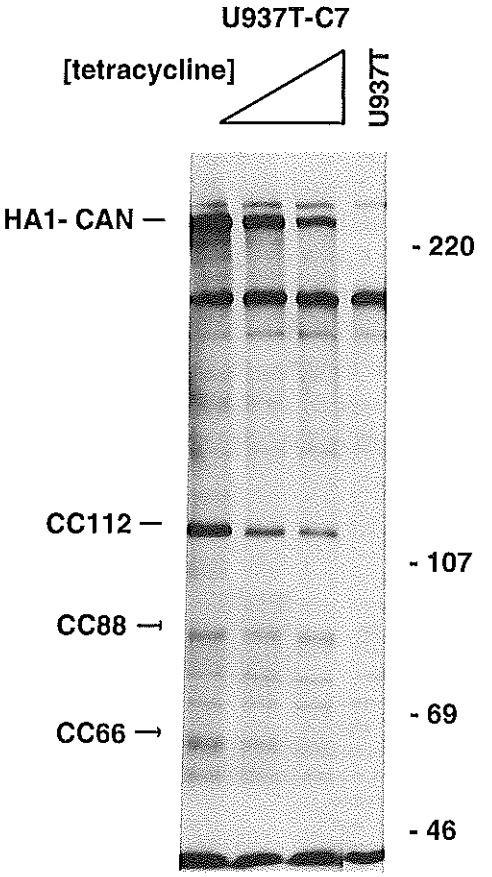


Figure 2 CAN co-immunoprecipitating proteins in human myeloid cell line U937T. HA1-CAN expression is tetracycline repressible in the stably transfected cell line U937T-C7. Cells were grown in 4, 6 or 10 ng/ml tetracycline (from left to right). Parental cell line U937T that does not express HA1-CAN is represented in the right lane. Following metabolic labelling with ^{35}S-methionine, proteins were immunoprecipitated using monoclonal antibody 12CA5 and analyzed on a 6% polyacrylamide gel. HA1-CAN and the co-precipitating proteins are indicated by arrows on the left. Molecular masses of standard proteins are indicated on the right (kDa).

remained more or less constant irrespective of the cellular HA1-CAN concentration. In contrast, the amount of co-precipitated CC112 increased proportionally to the amounts of HA1-CAN expression. These data suggest that there is a fixed amount of CC88 in the cell available for co-precipitation, whereas there is a surplus of CC112 of which normally only a portion is bound to CAN. A co-precipitating protein corresponding in size to CC66 in HeLa cells was also seen in U937T-C7 cells. Like CC112, the amount of CC66 increased proportionally with the amount of HA1-CAN.

Since CC112 and CC88 appeared to be cellular proteins that specifically interact with nucleoporin CAN, they could provide important new insights into the function of CAN, DEK-CAN and SET-CAN. We studied their interaction with CAN in detail and generated a collection of plasmids encoding HA1-CAN deletion mutants (see Materials & Methods). After transient transfection into HtTA-1 cells, ^{3}H-labelled proteins were immunoprecipitated with 12 CA5 and analyzed on 6% SDS-PAGE gels. Figure 3 summarizes the results of this analysis, which are further addressed in the paragraphs below.

CC112 co-precipitation

The deletion analysis shows that the smallest region in the C-terminal half of CAN that could mediate co-immunoprecipitation of CC112 consisted of the C-terminal 227 amino acids (Figure

3; CAN 1864-2090). This represents the C-terminal portion of the FG repeat of CAN. Interestingly, the N-terminal half of this repeat, represented by CAN1558-1839, was not able to co-precipitate CC112. This indicated that the FG repeat region of CAN is heterogenous in nature, an observation not made before for nucleoporin repeat regions. To map the CC112 binding region within the last 227 aminoacids, we tested additional N-terminal (CAN 1957-2090), C-terminal (CAN 1864-2052) and internal (CAN 1864-1912, 1984-2090) deletion mutants (Figure 3). None of these proteins was able to co-precipitate CC112, suggesting that strong CAN/CC112 interaction requires the entire C-terminal FG repeat region.

Since association with CC112 could represent an important function of CAN, DEK-CAN and SET-CAN, we investigated naturally occurring alternative forms of its binding region. Screening for alternatively spliced *can* mRNAs by RT-PCR, led us to identify a cDNA that represents a minor transcript in KG-1 and HeLa cells and lacks the exon representing nucleotides 5687- 5843 of the *can* mRNA, encoding amino acids 1865-1916 (data not shown).

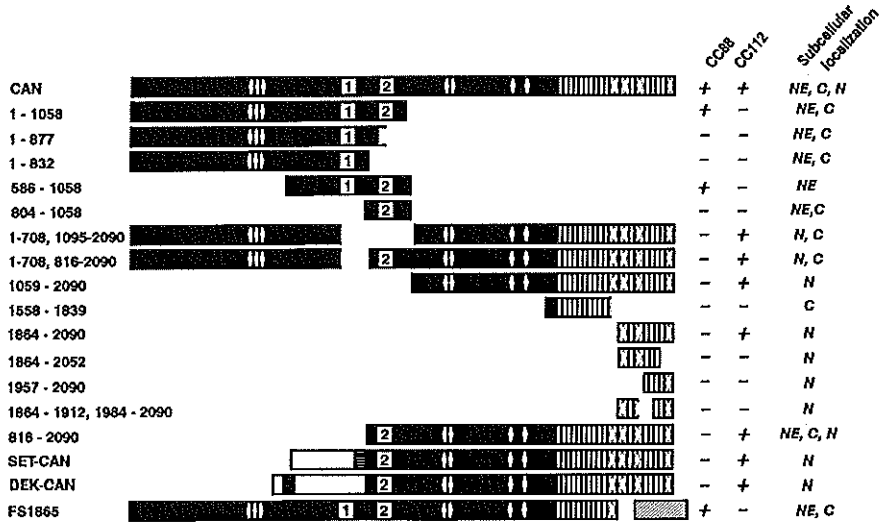


Figure 3 Mapping of domains in CAN mediating co-precipitation of CC88 and CC112 and subcellular localization. Bars represent CAN and CAN mutant proteins; numbers on the left represent amino acid boundaries; FS, frame-shift. The ability of mutants to co-immunoprecipitate CC88 or CC112 is indicated by pluses and minuses on the right hand side. Subcellular localization: NE, nuclear envelope; C, cytoplasm; N, nucleus. Multiple localizations appear in order of prominence. Predicted structural motifs are represented as follows: FG repeats as vertical bars, FXF repeats as diamonds, α-helical protein interaction domains as white boxes: (1) coiled-coil #1 and adjacent leucine zipper; (2) coiled-coil #2. The cross-hatched box represents the alternative reading frame in CAN FS1865. Horizontal stripes indicate acidic regions in the DEK and SET proteins.

At codon 1917, translation of the alternatively spliced mRNA changes to a different reading frame that remains open until position 6519, 154 nucleotides downstream of the normal stop codon. The 224 amino acid sequence encoded by the alternative reading frame shows no significant homology to known proteins, and does not contain any nucleoporin repeats. When this alternatively spliced form was introduced into the normal *can* cDNA, the expressed protein (CAN FS1865, see Figure 3) was unable to co-precipitate CC112, whereas CC88 precipitation was unaffected.

CC88 co-precipitation

Unlike CC112, co-precipitation of CC88 required the N-terminal half of CAN (Figure 1). CAN mutant 586-1058 was still able to co-precipitate CC88, whereas the mutant lacking 709-1094 (CAN1-708, 1095-2090) did not, further delineating CC88 interacting sequences to CAN's central region. The same region was previously identified to be important for association of CAN with the nuclear envelope and, by implication, the nuclear pore (Fornerod *et al.*, 1995). We therefore wished to map this region of interaction in more detail. The central domain of CAN contains three potential protein-protein interaction domains (Fornerod *et al.*, 1995; Von Lindern *et al.*, 1992a). A predicted N-terminal coiled-coil region and an adjacent leucine zipper are contained within amino acids 702-774, flanked by a second, more C-terminally located coiled-coil region formed by amino acids 860 - 887. CAN deletion mutants that lack the N-terminal coiled-coil and leucine zipper region, such as CAN1-708, 816-2090 and CAN804-1085 failed to co-precipitate CC88, establishing the importance of this region for CC88 interaction. The same was found for mutants that lack (CAN 1-832) or truncate (CAN 1-877, Figure 1) the second coiled coil region. These data indicate that multiple putative interaction domains contribute to CC88 binding and are needed for co-precipitation.

To further investigate whether the N-terminal coiled-coil, the leucine zipper and/or the C-terminal coiled-coil are involved in CC88 interaction, we introduced point mutations in these motifs using the CAN mutant 586-1085 background. In all cases, aliphatic amino acids in heptad repeat motifs of predicted α-helices were exchanged for helix-breaking prolines (Figure 4A). Since proteins had been metabolically labelled with ^{35}S-methionine and the number of methionines remained constant in these CAN mutants, quantitation of the ratio of radioactivity incorporated in immunoprecipitated CAN and CC88 directly reflected their binding affinity. As shown in Figure 4B, mutation of $I_{754} \rightarrow P$ strongly diminished co-precipitation of CC88, suggesting that the leucine zipper of CAN is important for CC88 binding. By contrast, no change in the amount of co-immunoprecipitating CC88 could be detected when prolines were introduced into the N-terminal ($I_{710} \rightarrow P$) or C-terminal ($V_{875} \rightarrow P$) coiled-coil regions. The double mutation $I_{710} \rightarrow P + V_{875} \rightarrow P$ did cause a marked decrease in CC-88 co-precipitation. This finding is in agreement with our deletion analysis; CC88 binding involves both the N- and

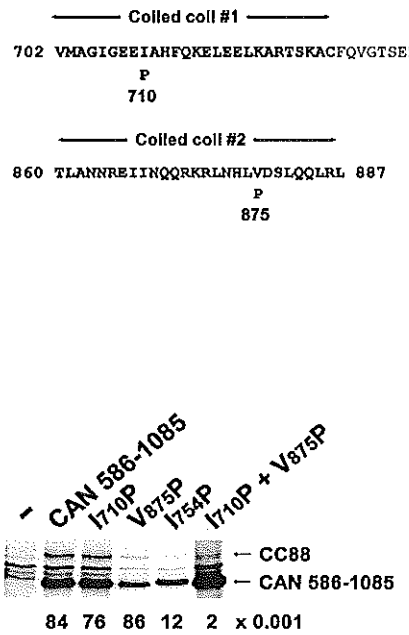


Figure 4 Effect of point mutations in predicted α-helical protein interaction domains of CAN on CC88 co-immunoprecipitation. (A) Amino acid sequence of the two predicted α-helical protein interaction domains flanking the t(6;9) breakpoint. Arrows are placed above coiled coil #1, the leucine zipper and coiled-coil #2. (Iso)leucines that form the leucine zipper are underlined. Point mutations introduced are indicated underneath the sequence. (B) HtTA-1 cells were transfected with expression plasmids encoding HA1-CAN586-1085 derivatives, carrying point mutations as indicated above the lanes. Untransfected cells are represented in the left lane (-). Following metabolic labelling with ^{35}S-methionine, proteins were immunoprecipitated using monoclonal antibody 12CA5 and analyzed on a 6% polyacrylamide gel. HA1-CAN586-1085 and CC88 are indicated by arrows on the right. The ratio of radioactivity present in CC88 (x 1000) over that of CAN586-1085 as determined by phosphorimager analysis is shown below each lane.

C-terminal α-helical protein interaction regions of CAN's central domain.

Subcellular localization of CAN mutants

The C-terminal FG-repeat region of CAN required for CC112 co-immunoprecipitation (amino acids 1864-2090) coincides with the region previously identified to confer nuclear localization of CAN mutants that do not associate with the nuclear pore (Fornerod *et al.*, 1995). Similarly, amino acids 586-1058 of CAN required for CC88 co-immunoprecipitation coincide with a region shown to be necessary and sufficient for nuclear pore association. To further investigate these apparent relations, we determined the subcellular localization of the CAN mutants described above by indirect immunofluorescence.

Expression of the N-terminal half of the C-terminal FG repeat (aminoacids 1558 -1839; Figure 3), that did not co-immunoprecipitate CC112, showed that this protein is located exclusively in the cytoplasm. This result identified the C-terminal 227 aminoacids of CAN as both necessary and sufficient for nuclear localization, and further sustained the correlation between nuclear localization and CC112 co-precipitation. However, deletion mutants within

the C-terminal, CC112-binding half of the FG repeat (Figure 3) were either completely or partially located to the nucleus, but did not co-precipitate CC112.

Immunofluorescence analysis of the central domain of CAN showed that mutants that have either the N- or the C-terminal predicted protein interaction domains intact, such as CAN804-1085 and CAN1-832, can be detected at the nuclear envelope, even if they do not co-precipitate CC88. Nuclear envelope association of mutants missing the C-terminal coiled-coil (CAN1-832) was decreased. Similarly, coiled-coil point mutant $V_{875}\rightarrow P$ displayed a marked decrease in nuclear envelope localization, and double mutant $I_{710}\rightarrow P + V_{875}\rightarrow P$ hardly showed any association at all (data not shown). In contrast, the single coiled-coil point mutation $I_{710}\rightarrow P$ did not noticeably decrease the protein's ability to associate with the nuclear envelope, nor did leucine zipper point-mutation $I_{754}\rightarrow P$. These data show that localization of CAN at the nuclear envelope does not correlate with CC88 co-precipitation, and they identify the C-terminal coiled-coil as the most important motif for nuclear pore association.

DEK-CAN and SET-CAN are nuclear proteins, in effect relocating the C-terminal part of CAN from the nuclear pore to the nucleoplasm (Fornerod *et al.*, 1995). Thus, we suspected that other sequences in the fusion proteins override the nuclear envelope interaction mediated by amino acids 813-1058. To test whether this is dependent on nuclear localization signals in DEK and SET, we expressed the CAN part (CAN816-2090) of the two fusion proteins alone. This mutant predominantly located to the nuclear envelope (Figure 3), demonstrating that DEK-CAN and SET-CAN depend on their DEK and SET moieties for relocation to the nucleus.

Discussion

As a first step to biochemically analyze the contribution of CAN sequences to the transforming capacity of the DEK-CAN and SET-CAN fusion proteins, we identified a 112 kDa protein (CC112) that associates with the C-terminal nucleoporin-repeat of CAN, DEK-CAN and SET-CAN. In addition, we found that the central region of CAN binds to an 88 kDa protein (CC88), that does not co-immunoprecipitate with the fusion proteins.

Using CAN mutants, we were able to locate regions in the CAN protein responsible for interaction with CC112 and CC88. Since these regions operate independently (see below), and CC112 and CC88 are by far the most prominent co-precipitating proteins, binding is likely to be direct. However, we cannot exclude that association of either of the two proteins is mediated by other proteins that escaped our attention, possibly because of small size. We noticed less consistent co-immunoprecipitation of a 66 kDa protein (CC66) in both HeLa and U937 derived cells. Because of this inconsistency, most likely caused by weaker binding to CAN and/or CC88, we decided to focus our analysis on the 88 and 112 kDa proteins.

CC112/CAN interaction

The region of CAN necessary and sufficient for CC112 co-precipitation consists of the C-terminal 227 amino acids. This region represents the distal half of CAN's 420 amino acid nucleoporin specific FG-repeat. We cannot identify significant differences in either FG density or inter-FG amino acid composition between these two halves of the FG repeat. It is therefore unclear what structural features of the distal part constitute the CC112 binding specificity. Identification of the FG repeat as a protein interaction domain per se suggests that its N-terminal part could specifically interact with other cellular proteins that may bind weaker than CC112, or that are less abundant in HeLa and U937 cells. In this respect, it is noteworthy that we isolated a cDNA for an alternatively spliced *can* mRNA encoding a C-terminally truncated FG-repeat. A CAN protein carrying this shortened repeat region could be functionally distinct since it displays a different specificity for associating factors, illustrated by the fact that it does not co-precipitate CC112.

The N-terminal half of the FG repeat alone localizes to the cytoplasm, whereas the C-terminal half localizes to the nucleus. This C-terminal domain of CAN bears no homology to known nuclear localization signals, such as a standard (Kalderon *et al.*, 1984) or bipartite NLS (Robbins *et al.*, 1988) or the M9 sequence of hnRNPA1 (Siomi & Dreyfuss, 1995). Thus, unless the C-terminal repeat of CAN carries an as yet unrecognized NLS-type sequence, transport to the nucleus must be mediated by an associating nucleophilic protein. An obvious candidate for this would be CC112, except for the apparently conflicting observation that additional N-terminal, C-terminal and internal deletion mutants of the C-terminal FG repeat still localize to the nucleus, but do not co-precipitate CC112. However, in these mutants, CC112 binding *in vivo* may be strong enough to mediate nuclear transport but too weak to withstand our co-immunoprecipitation conditions. Thus, resolving the question whether CC112 mediates the nuclear localization of C-terminal CAN awaits the molecular cloning of CC112. This would allow us to use more sensitive methods to probe for interaction between these proteins, such as a two-hybrid analysis in yeast.

Since CC112 interacts with CAN, which is part of the nuclear pore complex, CC112 is likely to be present at the nuclear pore as well. Our data strongly indicate that there is a molar excess of CC112 as compared to CAN in the cell. This additional CC112 may be bound to other nuclear pore components, in particular to those containing FG repeat sequences, or it may be located at additional sites in the cell. Since DEK-CAN and SET-CAN are nuclear proteins, we expect that CC112 is also located in the nucleus.

CC88/CAN interaction

A second co-immunoprecipitating protein, CC88, interacts with the central region of CAN containing three potential protein interaction domains: two coiled-coil regions and a leucine zipper (Fornerod *et al.*, 1995; Von Lindern *et al.*, 1992a). Coiled-coil domains are common motifs in nuclear pore components, such as NUP82 (Grandi *et al.*, 1995), Trp (Byrd *et al.*, 1994), NSP49 (Hurt, 1989), p62 (Starr *et al.*, 1990) and Nsp1 (Wimmer *et al.*, 1992). Their coiled-coil domains

appear to be involved in association of these proteins in nuclear pore subcomplexes (Grandi *et al.*, 1995) and could also be required for anchoring to structural components of the nuclear pore (Byrd *et al.*, 1994; Hurt, 1989). Our data show that the coiled-coil regions of CAN are prime candidates to mediate binding of this nucleoporin to the NPC, since point mutations in these motifs negatively influence CAN's nuclear envelope localization.

CC88 co-precipitation is drastically decreased by a single point mutation in the leucine zipper as well as by combined point mutations in both coiled-coils. These data indicate that strong interaction required for co-immunoprecipitation is mediated by all three α-helical interaction domains. This notion is also in agreement with our deletion analysis that shows that mutants that lack any of these domains are unable to co-precipitate CC88. We cannot exclude, however, the possibility that the point mutations affect normal folding of a larger region that could include other CC88 binding domains.

All CAN mutants that interact with CC88 are predominantly located at the nuclear envelope, and by implication at the nuclear pore. It is therefore likely that CC88 is part of the nuclear pore complex. Because overexpression of CAN does not increase the amount of co-precipitating CC88, the NPC may be its only location. This raises a question as to whether the apparent lack of CC88 interaction with DEK-CAN and SET-CAN could be the result of differential cellular compartmentalization. We find this explanation unlikely since CAN mutants that are located at the nuclear envelope (such as CAN 804-1058 or CAN 816-2090), and contain the same part of CAN's central region as DEK-CAN or SET-CAN, do not co-precipitate CC88 either. Thus, at present we have no indications that CC88 could contribute to the DEK-CAN or SET-CAN mediated leukemic process.

Concluding remarks

We have shown that CC112 interacts with DEK-CAN and SET-CAN. Insights into the functional significance of this 112 kDa protein for the leukemic process is obviously hindered by our lack of understanding of the function of DEK-CAN and SET-CAN themselves. However, we would like to present two hypothetical scenarios.

(i) The gene fusion between NUP98 and HoxA9 has recently been described in leukemic patients carrying translocation (7;11) (Borrow *et al.*, 1996; Nakamura *et al.*, 1996). Interestingly, the part of NUP98 that is predicted to be present in the NUP98/HoxA9 fusion protein consists of its nucleoporin-specific FXFG repeat, whereas sequences contributed by the transcription factor HoxA9 include its DNA binding domain. If NUP98-HoxA9 and DEK-CAN/SET-CAN work in a similar fashion, DEK and SET may contribute a DNA binding property to our fusion proteins. The observations that DEK co-localizes with metaphase chromosomes in mitotic cells (Fornerod *et al.*, 1995) and binds to double-stranded DNA *in vitro* (Titia de Lange, personal communication) could support this idea. SET has been shown to stimulate adenovirus replication *in vitro* (Nagata *et al.*, 1995), and thus could also be involved in DNA binding. Therefore, in analogy to many other

oncogenic nuclear fusion proteins (for review see Rabbitts, 1994), these nucleoporin fusions may act as mutant transcription factors, where DEK, SET or HoxA9 would mediate DNA binding to regulatory sequences of target genes, whereas CAN or NUP98 provide sequences essential for transcriptional disregulation. Thus proteins that associate with the nucleoporin portions of these fusion proteins, such as CC112, may provide an essential contribution.

(ii) Another, but not mutually exclusive possibility is that nucleoporin related fusion proteins deregulate cellular growth through changes in nucleo-cytoplasmic transport. A protein was recently described, that is involved in Rev-mediated export of HIV RNA from the nucleus (Bogard *et al.*, 1995; Friz *et al.*, 1995; Stutz *et al.*, 1995). This protein, called Rab or RIP, shows significant homology to the CC112-interaction domain of CAN. We find this homology fascinating, particularly in view of our observation that CAN is essential for nucleocytoplasmic transport of protein and mRNA (van Deursen *et al.*, 1996). In addition to the viral Rev protein, one cellular protein has previously been shown to interact with nucleoporin-specific repeats: import factor p97/importinβ. This protein is part of a cytoplasmic complex that mediates transport of NLS containing proteins from the cytoplasm to the nuclear pore (see for instance Görlich & Mattaj, 1996). In analogy to Rev and p97/importinβ, CC112 may be involved in transport processes between the nuclear pore and the nucleoplasm and/or cytoplasm, that may be disturbed by fusion proteins such as DEK-CAN and SET-CAN .

We hypothesize that further analysis of the CC112 protein may provide crucial information to improve our understanding of how DEK-CAN and SET-CAN deregulate growth of early hematopoietic progenitor cells. Further analysis of CC112 and CC88 may also shed more light on CAN's function in nucleocytoplasmic transport.

Materials and Methods

Expression constructs

All expression plasmids used in this study carry sequences encoding two copies of the influenza virus hemagglutinin (HA1) tag at the 5' end of their open reading frames (Fornerod *et al.*, 1995). All contructs with the exception of pSCTOP-HA1-Set-Can, make use of the Tet-VP16 responsive promoter (Gossen & Bujard, 1992). SET-CAN expression was directed by plasmid pSCTOP-HA1-Set-Can. This plasmid was constructed by replacing the *set* insert of pSCTOP-HA1-Set (Fornerod *et al.*, 1995) by a *Sac*I *set-can* fragment from anti-sense pCDXX-Set-Can (Fornerod *et al.*, 1995). DEK-CAN expression was directed by plasmid pHA1-Dek-Can (Fornerod *et al.*, 1995). pHA1-Can is similar to pCDXX-HA1-Can (Fornerod *et al.*, 1995), with the exception that the *Eco*RI cDNA insert is present in the pUHD10S vector (Fornerod *et al.*, 1995). Expression of deletion mutants CAN1-1058 and CAN586-1058 was directed by plasmids pHA1-CANΔ1059-2090 and pHA1-CANΔ1-585, and Δ1059-2090 respectively (Fornerod *et al.*, 1995). pHA1-CAN1-832 was created

by filling in the *Xba*I site of pHA1-CANΔ1059-2090 at position 2587 of the *can* cDNA (Von Lindern *et al.*, 1992a), creating a stop after codon 832. pHA1-CAN1864-2090 was derived from pHA1-CANΔ1-1139, Δ1341-1863 (Fornerod *et al.*, 1995) by deleting sequences encoding amino acids 1140-1340 using the *Nhe*I site flanking the HA1 encoding region and the *Msc*I site at position 5680 of the *can* cDNA. pHA1-CAN1864-2052 is a truncation derivative of this plasmid using *Xcm*I sites at position 6244 of the *can* cDNA and in pUHD10S. pHA1-CAN1957-2090 was also derived from pHA1-CAN1864-2090 by deleting sequences between the *Nco*I site flanking the HA1 tag coding region and the *Sty*I site at position 5758 of the *can* cDNA. pHA1-CAN1864-1912, 1984-2090 was created by an in frame deletion using the *Bam*HI sites at positions 5825 and 6044 of the *can* cDNA. pHA1-CAN1556-1839 was derived from pHA1-Can by ligation of the filled-in *Nde*I site flanking the HA1 tag coding region onto the *Spe*I site at position 4766 of the *can* cDNA and subsequent filling in of the HindIII site at position 5607. This creates a stop 5 codons downstream of codon 1839. pHA1-CAN 816-2090 was derived from pHA1-Can by deleting sequences upstream of the *Kas*I site at position 2538 of the *can* cDNA. Deletions in expression plasmids pHA1-CAN1-877, pHA1-CAN1-1085, pHA1-CAN804-2090 and pHA1-CAN1059-2090 were created using partial *E. coli* exonucleaseIII digestion using the Erase-a-base system (Promega, Madison, WI Henikoff, 1984), and was performed according to manufacturers's directions. For this purpose, two modifications of pHA1-Can were constructed. For N-terminal deletions, a *Not*I-*Sph*I linker was placed between sequences encoding the HA1 tag and the *Nde*I site at amino acid 1 of the *can* open reading frame. Removal of nucleotides originated from the *Nde*I site, while the 3' overhanging *Sph*I end protected upstream sequences from exo III digestion. Reading frames could be corrected by introducing adapters into the unique *Not*I site. For C-terminal deletions, a *Sal*I-*Sph*I-*Not*I linker was introduced 3' of the *can* cDNA, and removal of nucleotides started at the *Sal*I site, whereas the *Sph*I 3' overhanging end prevented downstream deletions. N- and C-terminal deletions could be linked using *Not*I sites, giving rise to internal CAN deletion mutants such as pHA1-CAN1-708, 1095-2090. pHA1-CAN586-1085 was created by an N-terminal truncation of pHA1-CAN1-1085 using a *Hinc*II site as described (Fornerod *et al.*, 1995). pHA1-CAN804-1058 was constructed by combining pHA1-CAN804-2090 and pHA1-CANΔ1059-2090. In a similar way, pHA1-CAN1-708, 816-2090 is the combination of pHA1-CAN1-708 and pHA1-CAN816-2090. Point mutations in pHA1-CAN586-1085 were generated in bacteriophage M13, using a commercial kit (Muta-Gene, BioRad, Hercules, CA) based on the Kunkel method (Kunkel, 1985). The following proline codons were introduced: amino acid 710(I→P): CCG, 754(I→P): CCT, 825(V→P): CCC.

Cell culture and transfection

HtTA-1 culture and transfection was performed as previously described (Fornerod *et al.*, 1995). Cell line U937T is a derivative of the human monoblast cell line U937 (Sundström & Nilsson, 1976) that expresses the tetracycline responsive chimeric transactivator Tet-VP16 (Gossen &

Bujard, 1992). U937T-C7 is stably transfected with pHA1-Can. A complete description and analysis of this cell line will be published elsewhere (J. B. and G. G., manuscript in preparation).

Immunoprecipitation and immunofluorescence

2 x 10^5 cells grown in 6 cm dishes were metabolically labelled for 6-16 h using 100 µCi ^{3}H-leucine (Amersham Corp., Arlington Heights, IL) or ^{35}S-methionine/cysteine *in vivo* labelling mix (DuPont NEN, Wilmington, DE) in 1.4 ml of leucine or methionine free DMEM, supplemented with 8% dialyzed fetal calf serum. Metabolic labelling was carried out in a closed container, in order to avoid radioactive contamination of the environment (Fornerod, 1996). Forty to 50 h after transfection, cells were cooled on ice, rinsed twice with ice cold PBS, scraped from the dishes and transferred to a microfuge tube. Cells were briefly spun down and lysed by pipet resuspension in 500 µl of NP40 lysis buffer (1% NP40; 50 mM Tris-HCl, pH8.0; 150 mM NaCl; 5 mM EGTA; 5 mM EDTA; 15 mM $MgCl_2$; 60 mM β-glycerolphosphate; 1 mM DTT; 0.1 mM $NaVO_4$; 0.1 mM NaF; 15 mM p-nitrophenylphosphate; 1.8 µg/ml aprotinin; 1 µg/ml leupeptin; 10 µg/ml soybean trypsin inhibitor; 0.1 mM benzamidine). Lysates were kept on ice for 10 min, gently vortexed, and centrifuged for 10 min in a microfuge at 4°C. The supernatant was then filtered through a 0.45 µm low-protein binding syringe filter (Supor Acrodisc, Gelman Sciences, Ann Arbor, MI), and precleared with 15 µl protein A sepharose. Precleared lysates were immunoprecipitated using 2 µg of monoclonal 12CA5 (BAbCo, Richmond, CA) bound to 7.5 µl protein A sepharose for one hour at 4°C. Beads were washed 4 times with 500 µl ice cold NP40 lysis buffer. Indirect immunofluorescence of transfected cells using anti-HA1 antibodies (12CA5) was carried out as previously described (Fornerod *et al.*, 1995).

PCR amplification

First strand cDNA was prepared from total RNA of the myeloid leukemia cell line KG-1 (Koeffler & Golde, 1978), using anti-sense primer CAN6237A (5'-AACCAGAAGTCTGTTGGGACAGTG-3') representing position 6237 - 6260 of the *can* cDNA (Von Lindern *et al.*, 1992a). DNA was subsequently PCR amplified using sense primer CAN5586SE (5'-CTTCTGGGTTCAGCTTTTGCCAAG-3') representing positions 5586-5609 and anti-sense primer CAN6209A (5'-AAGTGGGGGCATTCTGACTCGC-3') representing positions 6209-6230 of the *can* cDNA, annealing was at 60°C. Amplified DNA derived from alternative transcripts was sequenced and cloned into pHA1-Can using the *Hin*dIII and *Nru*I sites at positions 5607 and 6207 of the *can* cDNA, creating pHA1-CAN-FS1865.

Acknowledgements

We thank Dr Al Reynolds for the 12CA5 antibody, Dr Titia de Lange for communicating unpublished results, Dr Jan van Deursen for valuable discussions and critical reading of the

manuscript, and Peggy Burdick for expert secretarial assistance. These studies were supported in part by Cancer Center CORE Grant CA-21765 and by the American Lebanese Syrian Associated Charities (ALSAC) of St. Jude Children's Research Hospital. H.M. was supported by a travel grant of the Dutch Cancer Foundation.

References

Bogerd, H.P., Fridell, R.A., Madore, S. & Cullen, B.R. (1995). *Cell*, **82,** 485-494.

Borrow, J., Shearman, A.M., Stanton, V.P., Jr., Becher, R., Collins, T., Williams, A.J., Dube, I., Katz, F., Kwong, Y.L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J. & Housman, D.E. (1996). *Nat. Genet.*, **12,** 159-167.

Byrd, D.A., Sweet, D.J., Pante, N., Konstantinov, K.N., Guan, T., Saphire, A.C., Mitchell, P.J., Cooper, C.S., Aebi, U. & Gerace, L. (1994). *J. Cell Biol.*, **127,** 1515-1526.

Chi, N.C., Adam, E.J.H. & Adam, S.A. (1995). *J. Cell Biol.*, **130,** 265-274.

Fornerod, M. (1996). *BioTechniques*, **20,** 876-877.

Fornerod, M., Boer, J., van Baal, S., Jaeglé, M., von Lindern, M., Murti, K.G., Davis, D., Bonten, J., Buijs, A. & Grosveld, G. (1995). *Oncogene*, **10,** 1739-1748.

Fritz, C.C., Zapp, M.L. & Green, M.R. (1995). *Nature*, **376,** 530-533.

Görlich, D. & Mattaj, I. (1996). *Science*, **271,** 1513-1518.

Görlich, D., Vogel, F., Mills, A.D., Hartmann, E. & Laskey, R.A. (1995). *Nature*, **377,** 246-248.

Gossen, M. & Bujard, H. (1992). *Proc. Natl Acad. Sci. USA*, **89,** 5547-5551.

Grandi, P., Emig, S., Weise, C., Hucho, F., Pohl, F., Pohl, T. & Hurt, E.C. (1995). *J. Cell Biol.*, **130,** 1263-1273.

Henikoff, S. (1984). *Gene*, **28,** 351-359.

Hurt, E.C. (1989). *J. Cell Sci. Suppl.*, **12,** 243-252.

Kalderon, D., Roberts, B.L., Richardson, W.D. & Smith, A.E. (1984). *Cell*, **39,** 499-509.

Kellogg, D.R., Kikuchi, A., Fujii-Nakata, T., Turck, C.W. & Murray, A.W. (1995). *J. Cell Biol.*, **130,** 661-673.

Koeffler, H.P. & Golde, D.W. (1978). *Science*, **200,** 1153-1154.

Kraemer, D., Wozniak, R.W., Blobel, G. & Radu, A. (1994). *Proc. Natl Acad. Sci. USA*, **91,** 1519-1523.

Kunkel, T.A. (1985). *Proc. Natl Acad. Sci. USA*, **82,** 488-492.

Moroianu, J., Hijikata, M., Blobel, G. & Radu, A. (1995). *Proc. Natl Acad. Sci. USA*, **92,** 6532-6536.

Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A. & Matsumoto, K. (1995). *Proc. Natl Acad. Sci. USA*, **92,** 4279-4283.

Nakamura, T., Largaespada, D.A., Lee, M.P., Johnson, L.A., Ohyashiki, K., Toyama, K., Chen, S.J., Willman, C.L., Chen, I.M., Feinberg, A.P., Jenkins, N.A., Copeland, N.G. & Shaughnessy, J.D., Jr. (1996). *Nat. Genet.*, **12,** 154-158.

Panté, N., Bastos, R., McMorrow, I., Burke, B. & Aebi, U. (1994). *J. Cell Biol.*, **126,** 603-617.

Rabbitts, T.H. (1994). *Nature*, **372,** 143-149.

Radu, A., Blobel, G. & Moore, M.S. (1995a). *Proc. Natl Acad. Sci. USA*, **92,** 1769-1773.

Radu, A., Moore, M.S. & Blobel, G. (1995b). *Cell*, **81,** 215-222.

Robbins, J., Dilworth, S.M., Laskey, R.A. & Dingwall, C. (1988). *Cell*, **64,** 615-623.

Rout, M.P. & Wente, S.R. (1994). *Trends Biochem. Sci.*, **4,** 357-363.

Rowley, J.D. & Potter, D. (1976). *Blood*, **47,** 705-721.

Siomi, H. & Dreyfuss, G. (1995). *J. Cell Biol.*, **129,** 551-560.

Soekarman, D., Von Lindern, M., Daenen, S., de Jong, B., Fonatsch, C., Heinze, B., Bartram, C., Hagemeijer, A. & Grosveld, G. (1992). *Blood*, **79,** 1-8.

Starr, C.M., D'Onofrio, M., Park, M.K. & Hanover, J.A. (1990). *J.Cell Biol.*, **110,** 1861-1871.

Stutz, F., Neville, M. & Rosbash, M. (1995). *Cell*, **82,** 495-506.

Sundström, C. & Nilsson, K. (1976). *Int. J. Cancer*, **171,** 565-577.

van Deursen, J., Boer, J., Kasper, L. & Grosveld, G. (1996). *EMBO J.*, in press.

von Lindern, M., Fornerod, M., van Baal, S., Jaeglé, M., de Wit, T., Buijs, A. & Grosveld, G. (1992a). *Mol. Cell Biol.*, **12,** 1687-1697.

von Lindern, M., Poustka, A., Lerach, H. & Grosveld, G. (1990). *Mol. Cell Biol.*, **10,** 4016-4026.

von Lindern, M., van Baal, S., Wiegant, J., Raap, A., Hagemeijer, A. & Grosveld, G. (1992b). *Mol. Cell Biol.*, **12,** 3346-3355.

Wilson, I.A., Niman, H.L., Hougten, R.A., Cherenson, A.R., Connolly, M.L. & Lerner, R.A. (1984). *Cell*, **37,** 767-778.

Wimmer, C., Doye, V., Grandi, P., Nehrbass, U. & Hurt, E.C. (1992). *EMBO J.*, **11,** 5051-5061.

Chapter 6

The Human Homologue of Yeast CRM1 is in a Dynamic Subcomplex with CAN/Nup214 and a Novel Nuclear Pore Component Nup88

Maarten Fornerod[1], Jan van Deursen[1], Sjozef van Baal[1], Albert Reynolds[2], Donna Davis[3], K.Gopal Murti[3], Jack Fransen[4], and Gerard Grosveld[1].

Departments of [1]Genetics, [2]Tumor Cell Biology and [3]Virology & Molecular Biology, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, U.S.A., and [4]Department of Cell Biology and Histology, Faculty of Medical Sciences, Nijmegen, The Netherlands.

Accepted for publication in *The EMBO Journal*, Oxford University Press.

The oncogenic nucleoporin CAN/Nup214 is essential in vertebrate cells. Its depletion results in defective nuclear protein import, inhibition of messenger RNA export, and cell cycle arrest. We recently found that CAN associates with proteins of 88 and 112 kDa, which we have now cloned and characterized. The 88 kDa protein is a novel nuclear pore complex (NPC) component, which we have named Nup88. Depletion of CAN from the NPC results in concomitant loss of Nup88, indicating that the localization of Nup88 to the NPC is dependent on CAN binding. The 112 kDa protein is the human homologue of yeast CRM1, a protein known to be required for maintenance of correct chromosome structure. This human CRM1 (hCRM1) localized to the NPC as well as to the nucleoplasm. Nuclear overexpression of the FG-repeat region of CAN, containing its hCRM1-interaction domain, resulted in depletion of hCRM1 from the NPC. In *CAN -/-* mouse embryos lacking CAN, hCRM1 remained in the nuclear envelope, suggesting that this protein can bind to other repeat-containing nucleoporins as well. Lastly, hCRM1 shares a domain of significant homology with importinβ, a cytoplasmic transport factor that interacts with nucleoporin repeat regions. We propose that hCRM1 is a soluble nuclear transport factor that interacts with the NPC.

Keywords: CAN/Nup214 / CSE1 / Importinβ / Nuclear Pore Complex / Nucleocytoplasmic transport.

INTRODUCTION

The nuclear pore complex (NPC) is a ~125 megadalton complex embedded in the nuclear envelope (NE) that mediates bidirectional nucleocytoplasmic traffic in eukaryotic cells (recently reviewed by Görlich and Mattaj, 1996; Panté and Aebi, 1994; Panté and Aebi, 1996; Simos and Hurt, 1995).

Although more than 30 NPC components have been isolated in both yeast and vertebrates, the interactions between the NPC and trafficking macromolecules are only recently beginning to be understood. The import of nuclear localization signal (NLS) carrying proteins into the nucleus is mediated by a heterodimeric receptor complex. The smaller subunit of this complex, named importinα, NRα or karyopherinα, directly binds to the NLS (Adam and Adam, 1994; Adam and Gerace, 1991; Görlich *et al.*, 1994). The larger subunit, named p97, importinβ, NRβ, or karyopherinβ, is thought to mediate docking to the NPC. Importinβ binds to several repeat containing nucleoporins *in vitro*, and certain nucleoporin repeats may act as the docking sites for the NLS-import complex (Chi *et al.*, 1995; Görlich *et al.*, 1995b; Moroianu *et al.*, 1995; Radu *et al.*, 1995b). After docking, the import complex translocates through the central pore of the NPC, and the import substrate is released into the nucleoplasm in an energy dependent manner, requiring the Ras-like GTPase Ran/TC4 (Melchior *et al.*, 1993; Moore and Blobel, 1993, for a recent review see e.g. Schlenstedt, 1996). Export of proteins and ribonucleoproteins (RNPs) from the nucleus is also an active process that uses some of the same factors involved in protein import, notably Ran/TC4 (Schlenstedt *et al.*, 1995) and importinα (Görlich *et al.*, 1996).

Within the NPC, several proteins interact in a genetic or physical manner. Copurification studies in yeast showed that the nucleoporin Nsp1 forms one complex with the nucleoporins Nup49, Nup57, and Nic96 (Grandi *et al.*, 1993), and forms a separate complex with Nup82 (Grandi *et al.*, 1995). Other yeast proteins that physically interact include are Srp1, Nup1, and Nup2 (Belanger *et al.*, 1994). Interestingly, Srp1 is the yeast homologue of importinα. Recently, a yeast complex has been identified that includes nucleoporins Nup84, Nup120, Nup85, and also Sec13, which is involved in the transport of proteins from the endoplasmic reticulum to the Golgi apparatus (Siniossoglou *et al.*, 1996). In higher eukaryotes, two protein subcomplexes have been identified, one containing nucleoporin p62, complexed with proteins of 58, 54 and 45 kDa (Guan *et al.*, 1995; Panté *et al.*, 1994), and the second containing p250, associated with a protein of 75 kD (Panté *et al.*, 1994). p62 is the metazoan homologue of yeast Nsp1, and p250 is probably identical to CAN/Nup214.

The CAN protein was originally identified through its involvement in two types of acute myeloid or undifferentiated leukemia (von Lindern *et al.*, 1992a; von Lindern *et al.*, 1992b). We have recently developed an *in vivo* approach to study the consequences of CAN-depletion in knock-out mouse embryos. Using this approach, we found that the absence of CAN leads

to simultaneous defects in nucleocytoplasmic transport and in cell cycle progression (van Deursen *et al.*, 1996). Previously, we identified a new CAN-containing complex that included proteins of 88 and 112 kDa (Fornerod *et al.*, 1996). The central region of CAN associates with the 88 kDa protein, most likely through coiled-coil interactions, whereas the 112 kDa protein interacts with part of CAN's nucleoporin-specific repeat region. Identification of these proteins by molecular cloning could improve our understanding of the function of CAN in the NPC. In addition, CC112 might be important in the leukemic process associated with DEK-CAN and SET-CAN, because it interacts with these leukemia-specific fusion proteins (Fornerod *et al.*, 1996).

Here, we report the cloning and characterization of these 88 and 112 kDa proteins. The 88 kDa protein is a new nuclear pore component that we name Nup88. The 112 kDa protein is the human homologue of yeast CRM1 and is located at the NPC and nucleus. We provide evidence that the human CRM1 protein binds multiple NPC components and moves between the nuclear pore and the nucleoplasm. We also identify a group of proteins that includes hCRM1, yeast CRM1, and importinβ, which may constitute a novel family of NPC-interacting transport factors.

RESULTS

Purification of CAN coprecipitating proteins

We showed recently that two proteins specifically coimmunoprecipitate with CAN, one of 88 kDa (CC88) and one of 112 kDa (CC112) (Fornerod *et al.*, 1996). To coprecipitate sufficient quantities of the 112 and 88 kDa proteins for micro-amino acid sequence analysis, we created stable cell lines that express an HA1-tagged CAN protein. To avoid toxic effects of high CAN expression (Fornerod *et al.*, 1995), we made use of the Tet-VP16 system (Gossen and Bujard, 1992) to repress HA1-CAN during the transient phases of transfection. Two independent, stably transfected cell clones, TTB6 and TTD2, were analyzed for proteins that coprecipitate with CAN. As shown in Figure 1A, both cell lines coprecipitated the expected 88 and 112 kDa proteins. We generated in much the same way C4322, expressing HA1-tagged DEK-CAN (Fornerod *et al.*, 1995). Only the 112 kDa protein coprecipitated from this cell line. To visualize the coprecipitating proteins on a silver stained gel (Figure 1A) we needed at least 10^7 cells per immunoprecipitation.

For micro-aminoacid sequence analysis, we scaled up the immunoprecipitation approximately 1000 fold, using cell line TTD2 (see Materials & Methods). Proteins from the preparation were separated by SDS-PAGE (Figure 1B), and the 88 and 112 kDa protein bands were excised from the gel. Quantities of a coprecipitating protein of 66 kDa (CC66, Fornerod

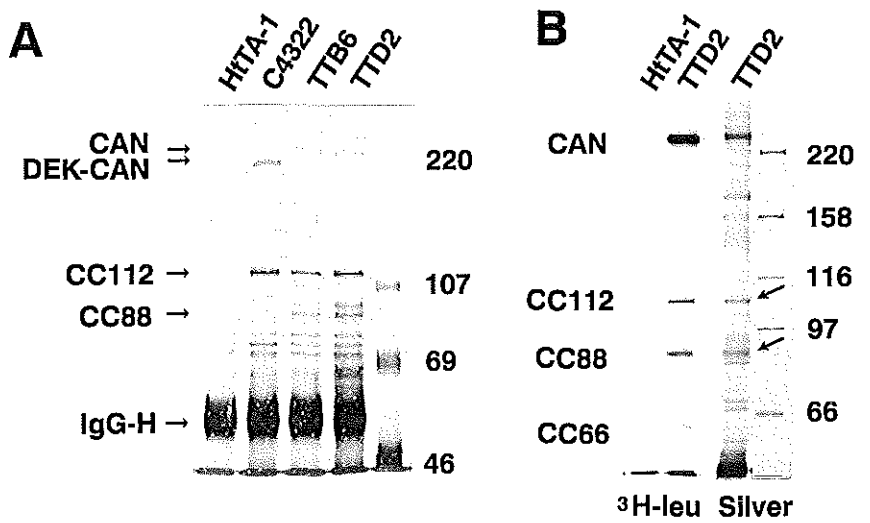


Figure 1 Immunopurification of CAN-associating proteins. (A) Proteins immunoprecipitated with monoclonal antibody 12CA5 from cell lines expressing an HA1-tagged version of CAN (TTB6 and TTD2), DEK-CAN (C4322), and the parental cell line (HtTA-1), electrophoretically separated on 6% polyacrylamide gels and visualized with silver staining. The positions of CAN and DEK-CAN, as well as those of the coprecipitating proteins CC112 and CC88, are indicated by arrows. A molecular weight standard is indicated on the right. IgG-H, immunoglobulin heavy chain. (B) An aliquot (0.1%) of the large scale immunopurification of CAN-associating proteins from the TTD2 cell line, run on a 6% polyacrylamide gel and silver stained. Arrows indicate the 112 and 88 kDa copurifying proteins, and molecular weight markers are shown in an adjacent lane (right panel). As a reference, HA1-CAN was coprecipitated from ^{3}H-leucine labeled TTD2 cells (left panel). The position of CAN and coprecipitating proteins CC112, CC88 and CC66 are indicated on the left.

et al., 1996) were insufficient for further analysis. Gelslices containing 82 and 48 pmol of CC88 and CC112 respectively were digested with trypsin *in situ*, and tryptic peptides were eluted, purified by preparative HPLC and then sequenced from the N-termini. Two sequences of 16 amino acids were obtained from CC88, and CC112 yielded one sequence of 7 amino acids (Table I). None of these sequences showed significant homology to known proteins; however, the two peptides derived from CC88 matched an uncharacterized human cDNA in the dBEST database (IMAGE clone 179414, Genbank Accession number H50498). Similarly, the peptide derived from CC112 matched a cDNA in the TIGR database (Clone HTTEU26, Human Genome Science, Rockville, MD). Interestingly, the amino acid sequence of the putative reading frame of this cDNA clone showed significant homology to a yeast protein of

Protein	Peptide/cDNA	Amino acid sequence
Nup88	CC88 peptide 50	-GPSGGGEEPAL(S)QYQ(R)
	c DNA IMAGE 179414	RGPSGGGEEPAL S QYQ R
	CC88 peptide 63rep20	-XQSPTEAEKPA(S)(S)(S/G)L(P/G)(K)
	c DNA IMAGE 179414	KNQSPTEAEKPA S S S L P S
hCRM1	CC112 peptide 63rep20	-LISGWVS(R)
	c DNA TIGR HTTEU26	KLISGWVS R

Table 1 Amino acid sequences of tryptic peptides derived from Nup88 and hCRM1 compared to virtual translations of expressed sequence tags found in computer databases. Trypsin hydrolyses peptide bonds at the C-terminal side of lysine (K) or arginine (R).

115 kDa, named CRM1. To determine whether these cDNAs were indeed derived from mRNAs encoding CC88 and CC112, full-length cDNA sequences were obtained from a human placenta cDNA library using clones 179414 and HTTEU26 as probes.

Sequence analysis of hCRM1 and NUP88

The complete cDNA putatively encoding CC112 had an open reading frame of 1071 amino acids and encoded a protein with a predicted molecular mass of 123 kDa (Figure 2A). This open reading frame showed high homology to *Saccharomyces cereviciae (S. cer.)* CRM1 (47% identity, 67% similarity) and to the *Schizosaccharomyces pombe (S. pombe)* homologue CRM1+ (52% identity, 69% similarity). We therefore named this protein hCRM1 (human CRM1). Further database searches revealed that the N-terminus of hCRM1 shared significant homology to the N-terminus of importinβ (Figure 2B). Importinβ is part of the nuclear protein import receptor and can bind CAN *in vitro* (Radu *et al.*, 1995a). In addition we found that a group of largely uncharacterized yeast and vertebrate proteins of similar size (110-120 kDa) shared this homology domain, that we propose to name the CRIME domain (CRm1, IMportinβ, Etcetera). The sequence divergence within the group was calculated according to Sneath and Sokal (1973)(Figure 2C).

The complete cDNA thought to encode CC88 had an open reading frame of 741 amino acids and a predicted molecular mass of 85 kDa (Figure 3A). Because an unrelated protein named Nup85 already exists, we have named this protein Nup88. Database searches revealed no significant homology of Nup88 to known proteins. However, the C-terminal sequences of Nup88 are predicted to form a coiled-coil (Lupas *et al.*, 1991 Figure 3B), an interaction domain often found in NPC proteins.

Interaction of Nup88 and hCRM1 with CAN

To further study the Nup88 and hCRM1 proteins, and to confirm their interaction with CAN,

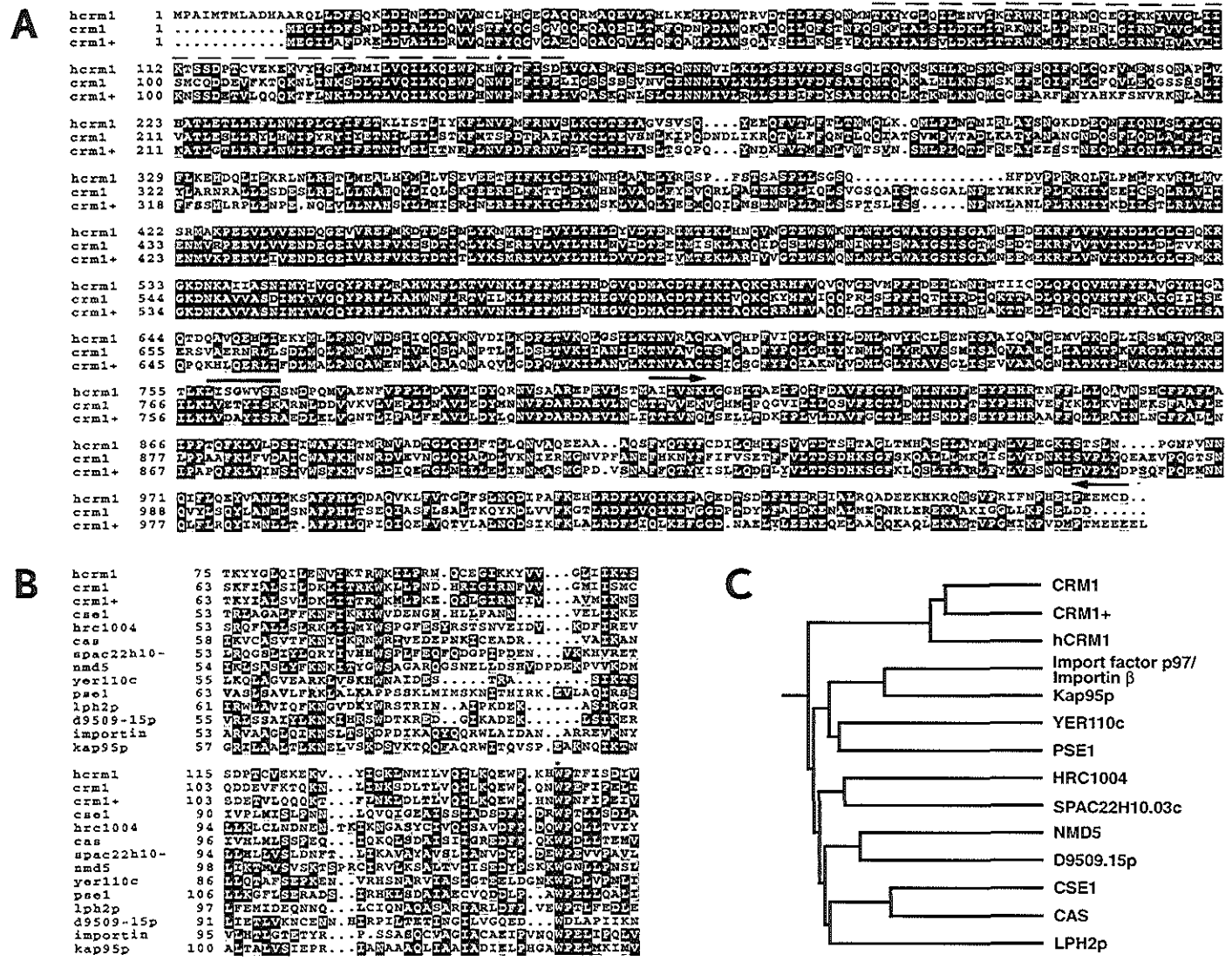

A
B
C
CRM1
CRM1+
hCRM1
Import factor p97/
Importin β
Kap95p
YER110c
PSE1
HRC1004
SPAC22H10.03c
NMD5
D9509.15p
CSE1
CAS
LPH2p

Figure 2 Amino acid sequence of hCRM1. (A) Comparison between hCRM1, *Saccharomyces cerevisiae* CRM1 (Adachi and Yanagida, 1989; Toda *et al.*, 1992), and *Schizosaccharomyces pombe* CRM1+ (Toda *et al.*, 1992). Identical and similar amino acids are boxed in black and gray respectively. The broken lines above the sequence denote the N-terminal homology domain and an asterisk indicates the conserved tryptophan. A bar indicates the position of peptide 63rep20. The amino acid sequences between the arrows were used to raise antibodies against the protein. (B) Comparison between hCRM1 amino acids 75-154 and similar N-terminal regions of *Saccharomyces cerevisiae* CRM1, CSE1 (Xiao *et al.*, 1993), Hrc1004 (Accession number S53939), Spac22H10.03c (Z69730), Nmd5 (P46970), Yer110c (P40069), Pse1 (Chow *et al.*, 1992), Lph2p (U43503), D9505.15p (U32274), Kap95p (Görlich *et al.*, 1995a), *Schizosaccharomyces pombe* CRM1+ and human CAS (Brinkmann *et al.*, 1995). The tryptophan that is conserved in all proteins is marked with an asterisk. Alignments were calculated using the program Clustal W with a gap penalty of 10 and a gap extension penalty of 0.05. (C) Dendrogram representing the sequence relationships between CRIME domain proteins. Relationships were calculated using the UPGMA algorithm (Sneath and Sokal, 1973) and were based on complete amino acid sequences.

we produced rabbit polyclonal antisera against amino acid 509-741 of Nup88 and amino acid 805-1071 of hCRM1 (Figures 2 and 3). These regions excluded the peptide sequences used to identify the Nup88 and hCRM1 cDNAs.

To confirm that isolated cDNA sequences indeed encoded the 88 and 112 kDa CAN-associating proteins, we tested whether affinity purified α-Nup88 and α-hCRM1 antisera could detect these proteins in CAN immunoprecipitates. Using IP-western blot analysis (Figure 4A), we found that the 112 kDa protein that was immunopurified from TTD2 or C4322 cells was recognized by antibodies against bacterially produced hCRM1. Similarly, the 88 kDa protein that was immunopurified from TTD2 cells reacted with antibodies against bacterially produced Nup88. The α-Nup88 antibodies also recognized a less abundant coprecipitating protein of ~90 kDa. This protein possibly represents a post-translationally modified form of Nup88. These results provide evidence that the *Nup88* and *hCRM1* genes encode CC88 and CC112, respectively.

We next assessed the cellular specificity of the affinity purified α-Nup88 and α-hCRM1 antisera by using them to immunoprecipitate proteins from ^{3}H-leucine labeled HtTA-1 whole cell extracts (Figure 4B). The antisera to hCRM1 and Nup88 specifically immunoprecipitated proteins of the correct sizes, relative to the proteins that coprecipitate with CAN. This indicates that the affinity purified antisera to hCRM1 and Nup88 are monospecific and suitable reagents with which to further characterize the proteins.

When we tested whether the hCRM1 and Nup88 specific antibodies would coprecipitate CAN, no precipitation was found. It is conceivable that these antibodies interfere with CAN association. For instance in case of Nup88, the serum may include antibodies against the predicted protein interaction domain (see Figure 3). Therefore, we linked the C-terminal 374 amino acids of Nup88, which contain this domain, to an HA1 tag and transiently expressed the product in HtTA-1 cells. Using monoclonal 12CA5 to the HA1 epitope, HA1-Nup88(368-741) coprecipitated a protein of the size of hCRM1 (Figure 4B) that reacted with anti-hCRM1

A

```
  1 MAAAEGPVGD GELWQTWLPN HVVFLRLREG LKNQSPTEAE KPASSSLPSS
 51 PPPQLLTRNV VFGLGGELFL WDGEDSSFLV VRLRGPSGGG EEPALSQYQR
101 LLCINPPLFE IYQVLLSPTQ HHVALIGIKG LMVLELPKRW GKNSEFEGGK
151 STVNCSTTPV AERFFTSSTS LTLKHAAWYP SEILDPHVVL LTSDNVIRIY
201 YLREPQTPTN VIILSEAEEE SLVLNKGRAY TASLGETAVA FDFGPLAAVP
251 KTLFGQNGKD EVVAYPLYIL YENGETFLTY ISLLHSPGNI GKLLGPLPMH
301 PAAEDNYGYD ACAVLCLPCV PNILVIATES GMLYHCVVLE GEEEDDHTSE
351 KSWDSRIDLI PSLYVFECVE LELALKLASG EDDPFDSDFS CPVKLHRDPK
401 CPSRYHCTHE AGVHSVGLTW IHKLHKFLGS DEEDKDSLQE LSTEQKCFVE
451 HILCTKPLPC RQPAPIRGFW IVPDILGPTM ICITSTYECL IWPLLSTVHP
501 ASPPLLCTRE DVEVAESPLR VLAETPDSFE KHIRSILQRS VANPAFLKAS
551 EKDIAPPPEE CLQLLSRATQ VFREQYILKQ DLAKEEIQRR VKLLCDQKKK
601 QLEDLSYCRE ERKSLREMAE RLADKYEEAK EKQEDIMNRM KKLLHSFHSE
651 LPVLSDSERD MKKELQLIPD QLRHLGNAIK QVTMKKDYQQ QKMEKVLSLP
701 KPTIILSAYQ RKCIQSILKE EGEHIREMVK QINDIRNHVN F 741
```

B

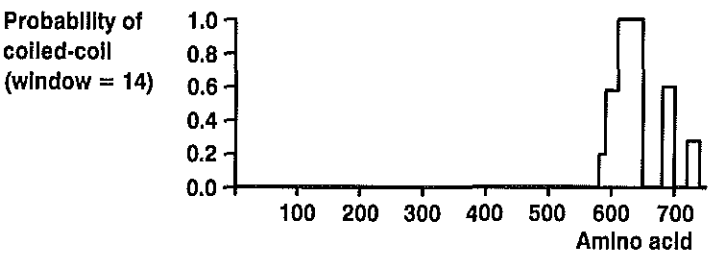


Figure 3 Sequence characteristics of Nup88. (A) Predicted amino acid sequence of the Nup88 protein. Bars above the sequence indicate peptides 50 and 67rep23 respectively. Solid lines show predicted coiled-coil regions. The amino acid sequences between the arrows were used to raise antibodies to Nup88. (B) Prediction of coiled-coil regions within Nup88; the program PEPCOIL was used which identifies potential coiled-coil regions of protein sequences based on the algorithm of Lupas *et al.* (1991).

antibodies in IP-western analysis (not shown). Since hCRM1 does not directly coprecipitate with Nup88 (Fornerod, 1996, Oncogene), hCRM1 most likely is coprecipitated via CAN. Indeed, the immunoprecipitate also contained a protein of ~220 kDa (Figure 4B), the size of CAN. This result indicates that the C-terminal part of Nup88 contains the CAN-interaction domain and confirms the existence of the CAN/hCRM1/Nup88 complex. Attempts to

precipitate this complex using an N-terminally HA1-tagged hCRM1 were unsuccessful, which could be due to the aberrant subcellular localization of this protein, that appeared to be exclusively nuclear (data not shown).

Subcellular localization of hCRM1 and Nup88

Because CAN localizes to the NPC, we anticipated that the proteins with which it interacts would also be present at that subcellular location. Moreover, the *S. pombe* and *S. cer.* CRM1 proteins are known to localize to the nucleus and particularly the nuclear periphery (Adachi and Yanagida, 1989). We addressed the subcellular localization of the hCRM1 and Nup88 protein by indirect immunofluorescence. Our monospecific α-hCRM1 antiserum showed a punctate rim staining around the nucleus (Figure 5A), which is a staining pattern characteristic for NPC proteins (Davis and Blobel, 1986). When focussed on the nuclear surface, a dotted staining pattern was observed (Figure 5B), also characteristic for NPC proteins. Moreover, a specific signal was present in the nucleoplasm (Figure 5A), with the nucleoli often staining stronger than the surrounding nucleoplasm. In the cytoplasm, the hCRM1-specific signal was restricted to small dots. Since these dots also stain with monoclonal antibody 414, directed against a common nucleoporin epitope (Davis and Blobel, 1986), they most likely represent annulate lamellae (data not shown). Antibodies to Nup88 localized exclusively to the nuclear envelope (Figure 5C).

To get a more detailed understanding of the subcellular localization of hCRM1, we examined this protein by using immunoelectron microscopy on LR-White embedded ultrathin sections of HtTA-1 cells. Most of the CRM1-specific gold label appeared in the nucleus, and the gold-density was highest at the nuclear envelope (data not shown). At the level of the nuclear pore, hCRM1 was present at both the cytoplasmic (Figure 5D and E) and nucleoplasmic face (Figure 5E and F) of the NPC. To confirm the localization at the level of the nuclear pore, we performed the same analysis on ultrathin cryosections of HtTA-1 cells. In such sections hCRM1-specific label was also found at both sides of the NPC, at approximately equal frequency (data not shown).

hCRM1 and Nup88 in CAN -/- embryos

Recently, we studied the phenotypic consequences of CAN depletion in early mouse embryos, homozygous for a knock-out mutation in the *CAN* gene (van Deursen *et al.*, 1996). Typically, depletion of CAN protein from maternal sources starts at around day 2.5 of gestation. By day 3.5, CAN is undetectable. To investigate hCRM1 and Nup88 localization in CAN-depleted embryos, we immunostained *CAN-/-* and wild-type blastocysts that were cultured *in vitro* for 18 hours, with affinity purified α-hCRM1 and α-Nup88 antibodies. Indirect immunofluorescence on wild-type embryos showed staining for hCRM1 within the nucleus and

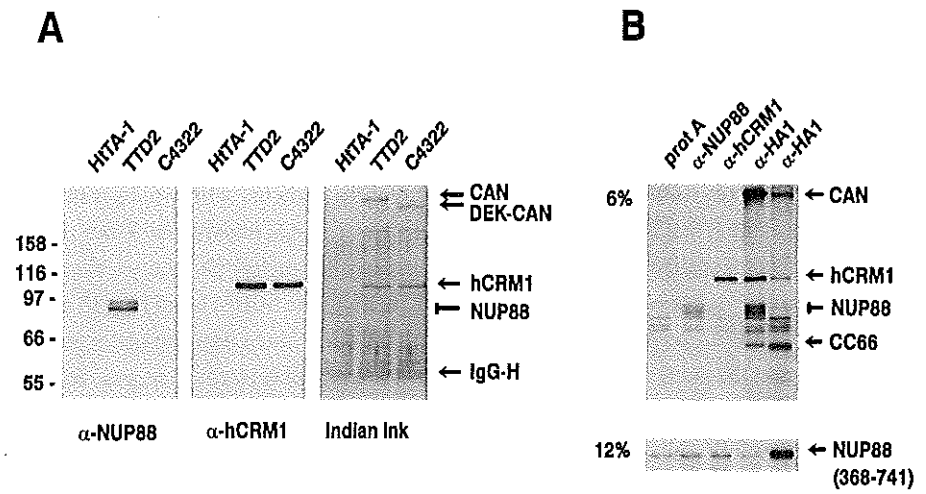


Figure 4 Interaction of CAN, hCRM1, and Nup88. (A) Western blot analysis of proteins that coprecipitated with HA1-tagged CAN or DEK-CAN, using monoclonal 12CA5, from cell lines TTD2 and C4322, as indicated above the lanes. The parental cell line HtTA-1 served as a negative control. Blots were stained with indian ink and incubated with antibodies to Nup88 or hCRM1, as indicated below the blots. Positions of CAN, DEK-CAN, hCRM1, and Nup88 are denoted on the right; IgG-H, immunoglobulin G heavy chain. Molecular weight standards are indicated on the left. (B) Immunoprecipition proteins from ^{3}H-leucine-labeled HtTA-1 cells with α-Nup88 or α-hCRM1 antibodies, as indicated above the lanes. Protein A sepharose (prot A) served as a negative control. Anti-HA1 antibody 12CA5 was used to immunoprecipitate ^{3}H-leucine labeled proteins from HtTA-1 cells that transiently expressed HA1-CAN (second lane from right) or HA1-Nup88(368-741) (right lane). Positions of CAN, hCRM1, Nup88, CC66, and Nup88(368-741) are indicated on the right. The polyacrylamide gel percentages are indicated on the left.

at the nuclear rim (Figure 6B, n=6), comparable to the staining pattern in human HtTA-1 cells. To confirm that the nuclear envelope was stained specifically, we performed a costaining with monoclonal antibody 414 (Davis and Blobel, 1986), that recognizes a group of nucleoporins (Figure 6A and D). CAN-depleted embryos also showed a clear nuclear rim staining and a nuclear signal (Figure 6C, n=5). This result suggests that NPC association of hCRM1 is not dependent on CAN, and presumably can be mediated by other NPC components. Intriguingly, every cell of the mutant embryo showed prominent staining of the nucleolus, which was not observed in wild-type embryos.

Nup88 was detected at the nuclear envelope in the wild-type embryos (Figure 6E, n=4). This staining was absent or barely detectable in cells of *CAN-/-* embryos (Figure 6F: n=4), indicating that the interaction of Nup88 at the NPC is CAN-dependent.

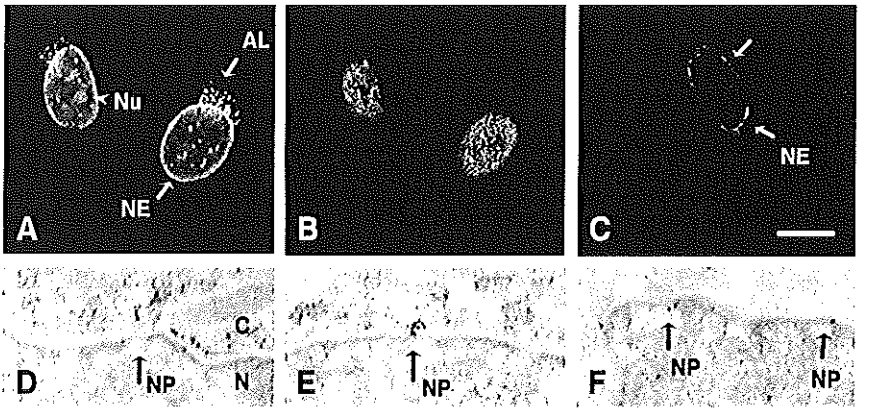


Figure 5 Subcellular distribution of hCRM1 and Nup88 in HtTA-1 cells. **A-C**, indirect immunofluorescence with α-hCRM1 (A and B) or α-Nup88 (C) antibodies. Panels A and B show the same cell, focused across (A) or on top of (B) the nucleus. Arrows indicate the nuclear envelope (NE) and structures resembling annulate lamellae (AL); an arrowhead marks a nucleolus (Nu) in panel A. **D-F**, immunoelectron microscopic localization of hCRM1. Three cross sections through the nuclear envelope are shown. The nuclear membrane is negatively stained. Gold particles decorate nuclear pores, indicated by arrows, at the cytoplasmic face (D and E) and at the nuclear face (E and F). NP, nuclear pore; C, cytoplasm; N, nucleoplasm. The bar is 6 μm in A-C, 150 nm in D-F.

hCRM1 displays dynamic behavior

We have shown in HtTA-1 cells that hCRM1 appears in the nucleoplasm, the NPC, and in nucleoli. We then examined whether this localization represents a static situation, or if there is trafficking of hCRM1 between these cellular compartments.

In previous studies, we found that the C-terminal part of CAN is located in the nucleus when expressed by itself (Fornerod *et al.*, 1995). Moreover, this part of CAN includes the hCRM1 binding domain (Fornerod *et al.*, 1996). If hCRM1 moves between the nuclear pore and the nucleoplasm, the presence in the nucleus of the CAN domain for hCRM1 binding might disturb hCRM1 intracellular routing. We therefore expressed the hCRM1-binding domain of CAN (the C-terminal amino acids 1864-2090) transiently in HtTA-1 cells. By using indirect immunofluorescence, we could detect the transfected protein with monoclonal 12CA5 (Figure 7A), while, in the same cells, we could monitor endogenous hCRM1 localization by using α-hCRM1 antibodies (Figure 7B). Expression of C-terminal CAN caused hCRM1 accumulation in the nucleus and its disappearance from the nuclear envelope, as we could verify by a double labeling experiment with monoclonal 414, that specifically stains the nuclear envelope (data not shown). These results suggest that the hCRM1-binding domain of CAN is

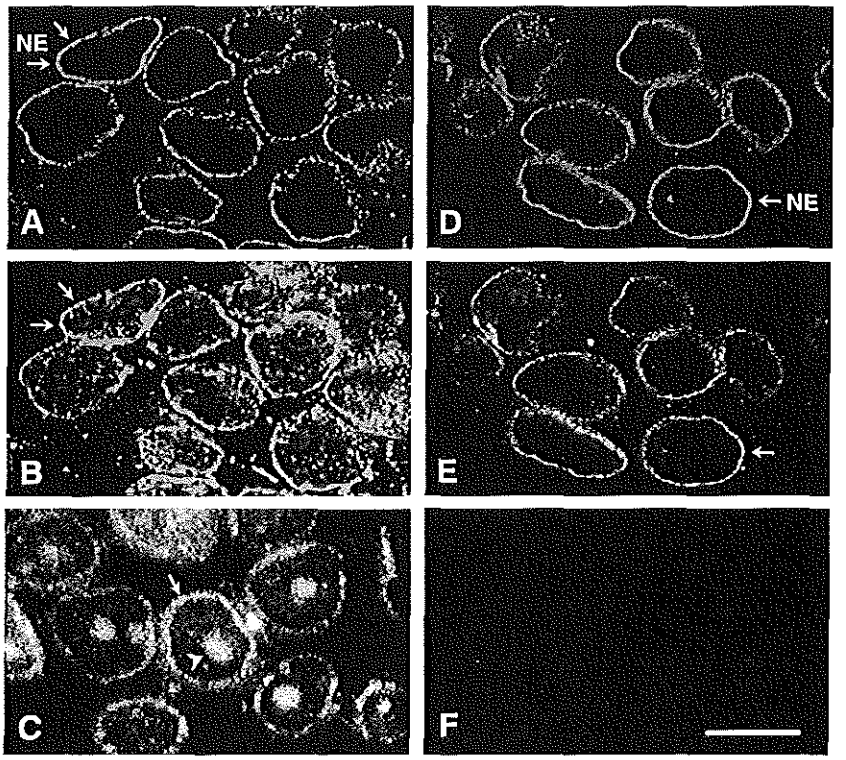


Figure 6 Subcellular localization of hCRM1 (B and C) and Nup88 (E and F) in wild-type (B and E) or *CAN-/-* (C and F) blastocysts. The specificity of nuclear envelope staining is shown by colocalization of wildtype embryos with monoclonal 414 (A and D). Only part of the embryo is shown. Arrows point to the nuclear envelope, an arrowhead denotes a nucleolus in panel C; NE, nuclear envelope. The bar represents 10 μm.

able to titrate hCRM1 from the NPC, and that the presence of hCRM1 at both the nuclear pore and in the nucleoplasm is a result of a dynamic exchange. However, if the turn-over time of hCRM1 is relatively short, nuclear hCRM1 accumulation could also be explained by newlysynthesized hCRM1 getting trapped in the nucleoplasm by C-terminal CAN. We therefore determined the half life of hCRM1 by pulse chase experiments, and found it to be approximately 24 hours (data not shown). This means that at least part of the hCRM1 protein that accumulates in the nucleus was originally located at the nuclear envelope.

As stated earlier, hCRM1 is present very prominent in the nucleoli of CAN-depleted embryos. Such accumulation could indicate that the phenotypic effects of CAN-depletion

include disruption of processes within the nucleolus. In an attempt to mimic such disturbances, we cultured HtTA-1 cells for 45 minutes in the presence of 0.04 μg/ml actinomycin D, a compound that, at this concentration, specifically inhibits RNA polymerase I dependent transcription (Perry and Kelley, 1970). Cells cultured in the presence of actinomycin D distinctly accumulated hCRM1 in their nucleoli (Figure 7D). Higher concentrations of actinomycin D (5 μg/ml), which also affect RNA polymerase II dependent transcription, had the same effect on hCRM1 localization (not shown). These results suggest that the nucleolus may be part of normal hCRM1 routing, and encourage the design of studies into the role of hCRM1 in nucleolar function and nucleolar/NPC trafficking.

DISCUSSION

The oncogenic nucleoporin CAN/Nup214 forms an NPC subcomplex with proteins of 88 and 112 kDa (Fornerod *et al.*, 1996). We immunopurified these two proteins and cloned their cDNAs via peptide sequencing.

The 88 kDa protein is a novel nuclear pore component, which we have named Nup88. This protein may be identical to p75, a protein previously shown to copurify with CAN from rat liver extracts (Panté *et al.*, 1994), although the difference in molecular weight seems to be considerable. Nup88 has no sequence homology to known proteins, but its C-terminus contains sequences that are predicted to form a coiled-coil domain. Predicted coiled-coil regions have been found in several other nuclear pore proteins, including CAN (for review see Panté and Aebi, 1994), and are thought to mediate interactions within NPC subcomplexes. Previously, we showed that mutations in CAN's coiled-coiled regions inhibit CAN interaction with the 88 kDa protein identified here as Nup88 (Fornerod *et al.*, 1996). This result suggests that the Nup88/CAN interaction is coiled-coil mediated. The position of the coiled-coil region of Nup88 is similar to that of *S. cer.* Nup82p, a protein that, if mutated, causes mRNA export defects (Hurwitz and Blobel, 1995), as does mutation of CAN (van Deursen *et al.*,1996). However, the sequence homology between these proteins is marginal (data not shown), and it remains to be determined whether Nup88 could be the functional homologue of yeast Nup82p.

CAN-/- mouse embryos that lack CAN, have no detectable Nup88 at their nuclear envelopes. Therefore, the presence of Nup88 at the nuclear pore depends on its physical interaction with CAN. This implies that the phenotypic effects of CAN elimination, which include G2 arrest and changes in nucleocytoplasmic trafficking (van Deursen *et al.*, 1996), may in part be caused by Nup88 depletion from the NPC.

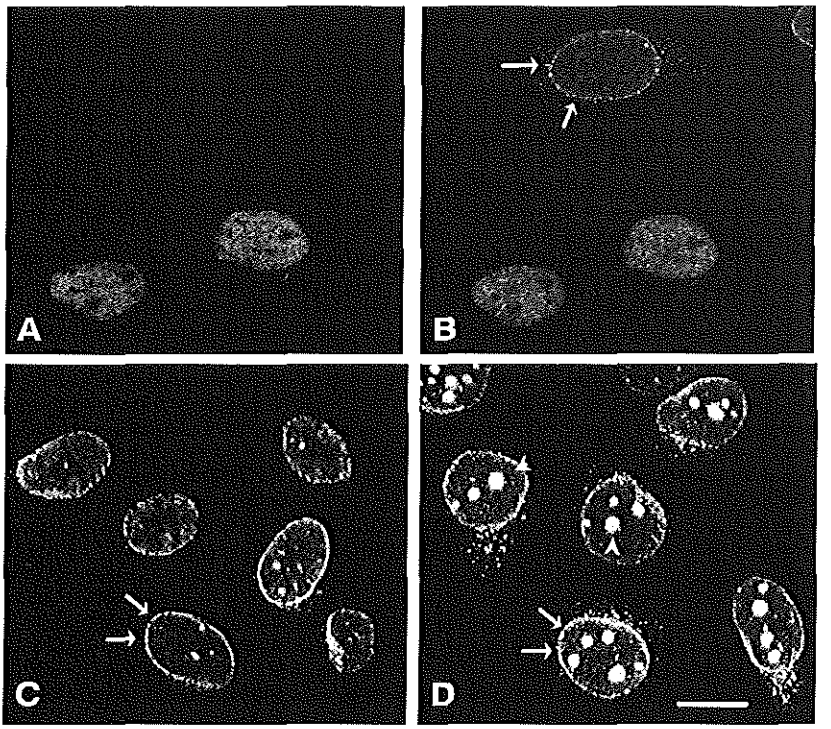


Figure 7 hCRM1 displays dynamic behavior. **A-B**, Double-stained HtTA-1 cells, transiently expressing the hCRM1 binding region of CAN, visualized with monoclonal 12CA5 (A). Endogenous hCRM1 is detected with α-hCRM1 antibodies (B). Arrows indicate hCRM1 nuclear envelope staining in an untransfected cell. **C-D**, hCRM1 protein localized by indirect immunofluorescence in normal HtTA-1 cells cultured for 45 minutes in the absence (C) or presence of (D) 0.04 μg/ml actinomycin D. Arrows indicate nuclear envelope staining, arrowheads point to nucleoli. The bar represents 8 μm.

Surprisingly, the 112 kDa protein appeared to be the human homologue of *S. cer.* and *S. pombe* CRM1, proteins not previously implicated in nucleocytoplasmic transport. The *S. pombe crm1+* (chromosome region maintenance) gene was first identified as a mutated gene in certain cold-sensitive strains that display deformed chromosomes at the restrictive temperature (Adachi and Yanagida, 1989). Furthermore, mutations in the *crm1* gene cause deregulation of the transcription factor *pap1* (the budding yeast homologue of human AP1) (Toda *et al.*, 1992), and can lead to multi-drug resistance (Nishi *et al.*, 1994; Turi *et al.*, 1994). Mutations in yeast genes involved in nucleocytoplasmic trafficking can, apart from transport defects, lead to similar pleiotropic effects, as illustrated by the yeast homologue of the GTPase

Ran/TC4, and its exchange factor RCC1 (Forrester *et al.*, 1992 6, 1914; Kadowaki *et al.*, 1993), and by Nup85p (Goldstein *et al.*, 1996). It is therefore conceivable that a transport defect may be responsible for the *crm1* phenotype. The yeast CRM1 proteins have been localized to the nucleus and are particularly prominent at the nuclear periphery (Adachi and Yanagida, 1989). This, together with the high homology between yeast and human CRM1 suggests that also in yeast, CRM1 may strongly associate with repeat-containing nucleoporins.

The hCRM1 protein is identical to the 112 kDa protein that interacts with DEK-CAN and SET-CAN, two nuclear fusion proteins associated with acute myeloid and undifferentiated leukemia respectively (Fornerod *et al.*, 1996). Because hCRM1 is not related to any proteins known to be involved in oncogenic transformation, its possible role in leukemogenesis remains to be determined. However, hCRM1 could be part of a novel pathway, via which nuclear pore components contribute to leukemogenesis.

Is hCRM1 a novel transport factor?

Several lines of evidence support the idea that hCRM1 could be a transport factor that dynamically interacts with the NPC.

First, the dual subcellular localization of hCRM1 to the nucleus and to the NPC suggests that this protein can travel between the two compartments. To test this, we overexpressed the hCRM1-binding domain of CAN, which is located in the nucleus and not at the NPC. If hCRM1 permanently binds to CAN, expression of its binding domain in the nucleus would have no effect. If, on the other hand, hCRM1 is periodically released from the NPC into the nucleus, the presence of an excess binding domain could sequester the hCRM1 in the nucleoplasm and lead to a gradual disappearance of hCRM1 from the nuclear envelope. We found that, under these conditions, hCRM1 was completely absent from the nuclear envelope and was only present in the nucleoplasm. Since we showed that the half life of hCRM1 is long, this suggests that hCRM1 can move from the NPC to the nucleoplasm. In addition, we have shown that repression of RNA polymerase I dependent transcription causes accumulation of hCRM1 in the nucleolus. Although we don't understand the mechanism causing this effect, it does suggest that hCRM1 routing involves the nucleolus.

Second, the nuclear envelopes of cells from CAN-depleted mouse embryos contain hCRM1. This suggests that hCRM1 can bind to NPC components other than CAN. We previously demonstrated that hCRM1 interacts with the C-terminal half of CAN's nucleoporin repeat region (Fornerod *et al.*, 1996). This repeat of CAN has significant homology to repeats of several other nucleoporins, including Nup98, Nup153 and p62. Thus, hCRM1 may interact with repeat regions of these, or other, yet unknown, vertebrate nucleoporins. In agreement with this is our observation that hCRM1 is present at the nuclear as well as the cytoplasmic face of the NPC, while CAN is only present at the cytoplasmic side (Kraemer *et al.*, 1994). Therefore, good candidates to mediate additional nuclear NPC association of hCRM1 are Nup98 and

Nup153, that both reside at the nuclear face of the NPC (Radu *et al.*, 1995b; Sukegawa and Blobel, 1993).

Third, hCRM1 shares a region of significant homology with importinβ. This factor physically interacts with nucleoporin-specific repeat regions (Moroianu *et al.*, 1995; Radu *et al.*, 1995b) and can bind CAN in ligand blot assays (Radu *et al.*, 1995a). This suggests that hCRM1 and importinβ may interact with the NPC by the same mechanism. Importinβ forms part of a cytoplasmic transport complex that mediates protein import into the nucleus. Two molecular mechanisms have been proposed for the translocation of this protein-import complex through the NPC (Görlich and Mattaj, 1996; Nehrbass and Blobel, 1996; Rexach and Blobel, 1995). Both models propose a stepwise binding and release of the importinβ component of the complex to and from nucleoporin repeats. Because the different nucleoporins localize to specific sites along the NPC, the transport direction of the complex is proposed to be established via an increased binding affinity of importinβ for nucleoporin repeats towards the nucleus. Following these models, the more cytoplasmically located CAN would have a relatively weak affinity for importinβ, which is in agreement with its absence in our CAN coimmunoprecipitation experiments. In contrast, hCRM1 appears to have a high affinity for CAN. This suggests that if hCRM1 interacts with nucleoporin-repeats in an importinβ-like fashion, it could move in the opposite direction, i.e., from the nucleus to the cytoplasmic face of the NPC.

In addition to importinβ and its yeast homologue Kap95p, we found nine other proteins that share the N-terminal CRIME domain. The majority of these proteins came from hypothetical open reading frames identified as part of the *S. cer.* genome sequencing project. However, CSE1 has been identified as an essential yeast protein, and its mutation results in a chromosome segregation defect (Xiao *et al.*, 1993). Moreover, it was reported that the *cse1* phenotype can be suppressed by high expression of Srp1 (Belanger *et al.*, 1994), the yeast importinα homologue that interacts with the nucleoporins Nup1 and Nup2. Thus, CSE1 is the third otherwise unrelated protein that shares the N-terminal domain and is implicated in NPC interaction. Therefore, this N-terminal homology domain may define a new group of NPC-interacting transport factors, and it will be interesting to test whether it is this domain that mediates interaction with nucleoporin repeat sequences.

MATERIALS AND METHODS

Cell culture and transfection

HtTA-1 cells (Gossen and Bujard, 1992) were cultured as described (Fornerod *et al.*, 1995). In some experiments Actinomycin D_1 (Boehringer Mannheim, Indianapolis, IN) or cycloheximide (Sigma, St. Louis, MO) were added to the culture medium. Cell lines TTD2 and

TTB6, which express HA1-CAN under the control of a tetracycline dependent promoter, were created by cotransfecting HtTA-1 cells with *Ssp*I linearized plasmid pHA1-CAN (Fornerod *et al.*, 1995) and *Sca*I linearized pJΩ6Puro at a molar ratio of 20:1. Puromycin resistant clones were selected as described (Fornerod *et al.*, 1995). TTD2 expressed HA1-CAN predominantly in the nuclear envelope in the absence of tetracycline. Under these conditions, the cell line showed normal growth characteristics for multiple passages. The HA1-DEK-CAN expressing cell line C4322 has been described previously, as has HtTA-1 transient transfection (Fornerod *et al.*, 1995). Plasmid pHA1-Nup88(368-741) was created by placing sequences encoding two copies of the influenza virus HA1 tag (Fornerod *et al.*, 1995) at the 5' side of codons 368-741 of the *Nup88* cDNA.

Immunopurification

Approximately 10^{10} TTD2 cells cultured on five hundred 15-cm dishes were rinsed once with PBS, scraped in PBS and spun down at 2000*g* in 50 ml tubes for 10 minutes at 4°C. Cell pellets, in total weighing 40 grams (wet weight) were frozen at -80°C until further processing. TTD2 cell aliquots (2 grams each) were transferred to 15 ml tubes, lysed in 8 ml ice-cold NP-40 lysis buffer (1% NP40; 50 mM Tris-HCl, pH8.0; 150 mM NaCl; 5 mM EGTA; 5 mM EDTA; 15 mM $MgCl_2$; 60 mM β-glycerolphosphate; 1 mM DTT; 0.1 mM $NaVO_4$; 0.1 mM NaF; 15 mM p-nitrophenylphosphate; 1.8 μg/ml aprotinin; 1 μg/ml leupeptin; 10 μg/ml soybean trypsin inhibitor; 0.1 mM benzamidine), and filtered through .45 μm cellulose acetate membranes. Lysates were then precleared for 30 minutes with 1 ml packed Sepharose CL-4B (Pharmacia). HA1-CAN was immunoprecipitated by rotating the cleared lysates twice for 1 hour with 0.4 mg of monoclonal antibody 12CA5 (BAbCo, Newport, CA) covalently linked to 0.2 ml packed CNBr-activated Sepharose CL-4B beads (Pharmacia). The beads were washed 4 times with 8 ml NP-40 lysis buffer and then once with PBS. Proteins were eluted from the Sepharose beads by subsequent batchwise elutions with 0.4, 0.4, and 0.2 ml 0.5% SDS, and concentrated in a Speed-Vac vacuum excicator to 250μl. Proteins in the eluates were subsequently precipitated with 5 volumes acetone at room temperature and spun down at 18,000 *g* for 10 minutes. The pellets were suspended in 25 μl solubilization buffer (10% SDS; 100 mM $MgCl_2$; 50 mM Tris-HCl (pH6.8); 0.1% bromophenol blue, 10% glycerol, 50 mM dithiothreitol) by vortexing and heating to 90°C. Fifty microliters of this fraction, which represents proteins purified from 4 grams of TTD2 cells, were loaded in 7 mm wide slots and separated on a 0.75 mm 6% polyacrylamide gel. Mock samples containing solubilization buffer alone were run in adjacent lanes to prevent the protein bands from fanning out. After electrophoresis, proteins were stained with Coomassie brilliant blue R250 (Bio-Rad, Hercules, CA), and excised from the gel. The protein in the gelslices was quantitated by laser desorbtion mass spectrometry (Williams *et al.*, 1996). In gel trypsin digestion, reversed phase HPLC, and amino acid sequencing were subsequently performed as described (Williams and Stone, 1995).

cDNA cloning

cDNA clone IMAGE179414, encoding Nup88 peptides 50 and 67rep23, was obtained from Research Genetics (Huntsville, AL). cDNA clone HTTEU26, encoding hCRM1 peptide 63rep20 was obtained from Human Genome Sciences (Rockville, MD). Full-length cDNAs were obtained from a human placenta cDNA library (Hu2002B#29203, Clonetech, Palo Alto, CA) and sequenced according to the established methodology (Sambrook *et al.*, 1989).

Antibodies

Polyclonal antisera against hCRM1 and Nup88 were raised in rabbits by using recombinant protein produced in *E. coli* BL21:DE3(pLysS) (Studier *et al.*, 1990). hCRM1 amino acids 805-1071 and Nup88 amino acids 509-741 were expressed as glutathione-S-transferase fusion proteins (Smith and Johnson, 1988) and were purified by using preparative SDS-PAGE and electro-elution. The antisera were then affinity purified by using these bacterial proteins immobilized on PVDF membrane as described previously (van Deursen *et al.*, 1996).

Immunoprecipitation and Western blotting

HA1-tagged protein was immunoprecipitated from ^{3}H-leucine-labeled cells as described (Fornerod *et al.*, 1996). It is important to note that NP-40 lysis buffer extracts both cytoplasmic and nuclear proteins, as shown previously (Fornerod *et al.*, 1996). hCRM1 and Nup88 were immunoprecipitated by the same procedure and by using purified antisera to the respective proteins at a dilution of 1 in 100. Unlabeled HA1-tagged protein was immunoprecipitated from 1×10^{7} TTD2 or C4322 cells with 5 μg of monoclonal antibody 12CA5 (Wilson *et al.*, 1984). Following gel electrophoresis, proteins were either silver stained (Ansorge, 1985) or blotted onto PVDF membrane (Millipore, Bedford, MA). Blots were stained with indian ink (Hancock and Tsang, 1983), blocked overnight in PBS containing 1% non-fat milk and 1% BSA, and then incubated for three hours with affinity purified α-hCRM1 or α-Nup88 antisera diluted to 1 in 100 in PBS/1% non-fat milk. Bound antibody was visualized by adding a horse-radish peroxidase- labeled secondary antibody (Jackson Laboratories, West Grove, PA), and using a commercial chemiluminescence kit (Dupont NEN, Boston, MA). The blots were washed in between antibody incubations 5 times for 5 minutes each in PBS/0.05% Tween-20.

Indirect immunofluorescence and immunoelectron microscopy

HtTA-1 cells were fixed and permeabilized as described (Fornerod *et al.*, 1995) and immunostained with affinity purified α-hCRM1 antibodies diluted to 1 in 90, α-Nup88 antibodies diluted to 1 in 30, monoclonal 414 (BAbCo, Richmond, CA) at 5 μg/ml, or monoclonal 12CA5 at 2 μg/ml. Embryos from CAN+/- heterozygous intercrosses were collected 3.5 days postcoitum and cultured *in vitro* for approximately 18 hours as described (van Deursen *et al.*, 1996). CAN-/- embryos were identified at the onset of blastocoel

contraction and fixed, alongside normal embryos (van Deursen *et al.*, 1996). These embryos were then immunostained overnight with purified α-hCRM1 or α-Nup88 antibodies, both diluted to 1 in 30, or monoclonal 414 (5 μg/ml), and images were collected by confocal laser scanning microscopy on a BioRad MRC1000 (BioRad, Hercules, CA) using a 40x oil objective.

HtTA-1 cells were examined by immunoelectron microscopy as previously described (Fornerod *et al.*, 1995), except that cells were embedded in LR-White normal grade rather than in hard grade (London Rasin Company Ltd., Basingstoke, England). Ultrathin cryosections were made and immunolabeled as described (Fransen *et al.*, 1991). Sections were labeled with affinity purified α-hCRM1 antibodies diluted to 1 in 10.

Acknowledgements

We thank Kathy Stone, Edward Papacoda and Dr. Kenneth Williams of the W.M. Keck foundation for micro amino acid sequencing analyses, Sharon Frase and Dr. Andrea J. Elberger for use of the Confocal Laser Scanning Facility, UT Memphis (funded by PHS Grant CLSM 1S10RR08385), Judith Boer for helpful discussions and critical reading of the manuscript, Huib Croes and Mietske Wijers for immunoelectron microscopy studies, Dr. Sue Vallance for scientific editing, and Peggy Burdick for secretarial assistance. These studies were supported in part by Cancer Center CORE Grant CA-21765 and by the Associated Lebanese Syrian American Charities (ALSAC) of St. Jude Children's Research Hospital.

Note added in proof: The sequences described in this publication will be available from the EMBL database under accession numbers Y08614 (hCRM1) and Y08612 (Nup88). The sequence of an alternative 3' untranslated region of Nup88 mRNA is available under accession number Y08613.

References

Adachi, Y. and 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 at its perphery. *J. Cell Biol.*, **108**, 1195-1207.

Adam, E.J.H. and Adam, S.A. (1994) Identification of cytosolic factors required for nuclear location sequence-mediated binding to the nuclear envelope. *J. Cell Biol.*, **125**, 547-555.

Adam, S.A. and Gerace, L. (1991) Cytosolic proteins that specifically bind nuclear localization signals are receptors for nuclear import. *Cell*, **66**, 837-847.

Ansorge, W. (1985) Fast and sensitive detection of protein and DNA bands. *J. Biochem. Biophys. Methods*, **11**, 13-20.

Belanger, K.D., Kenna, M.A., Wei, S. and Davis, L.I. (1994) Genetic and physical interactions between Srp1p and nuclear pore complex proteins Nup1p and Nup2p. *J. Cell Biol.*, **126**, 619-630.

Brinkmann, U., Brinkmann, E., Gallo, M. and Pastan, I. (1995) Cloning and characterization of a cellular apoptosis susceptibility gene, the human homologue to the yeast chromosome segregation gene *CSE1*. *Proc. Natl Acad. Sci. USA*, **92**, 10427-10431.

Chi, N.C., Adam, E.J.H. and Adam, S.A. (1995) Sequence and characterization of cytoplasmic nuclear protein import factor p97. *J. Cell Biol.*, **130**, 265-274.

Chow, T.Y.-K., Ash, J.J., Dignard, D., Thomas, D.Y. (1992) Screening and identification of a gene, *PSE1*, that affects protein secretion in *Saccharomyces cerevisiae*. *J. Cell Sci.* **101**, 709-719.

Davis, L.I. and Blobel, G. (1986) Identification and characterization of a nuclear pore complex protein. *Cell*, **45**, 699-709.

Fornerod, M., Boer, J., van Baal, S., Jaeglé, M., Von Lindern, M., Murti, K.G., Davis, D., Bonten, J., Buijs, A. and Grosveld, G. (1995) Relocation of the carboxyterminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. *Oncogene*, **10**, 1739-1748.

Fornerod, M., Boer, J., van Baal, S., Morreau, H. and Grosveld, G. (1996) Interaction of cellular proteins with the leukemia specific fusion proteins DEK-CAN and SET-CAN and their normal counterpart, the nucleoporin CAN. *Oncogene*, in press.

Forrester, W., Stutuz, F., Rosbash, M. and Wickens, M. (1992) Defects in mRNA 3'-end formation, transcription initiation, and mRNA transport associated with the yeast mutation prp20: Possible coupling of mRNA processing and chromatin structure. *Genes Dev.*, **6**, 1914-1926.

Fransen, J.A.M., Hauri, H.P., Ginsel, L.A. and Naim, H.Y. (1991) Naturally occuring mutations in intestinal sucrase-isomaltase provide evidence for the existence of an intracellular sorting signal in the isomaltase subunit. *J. Cell Biol.*, **115**, 45-57.

Goldstein, A.L., Snay, C.A., Heath, C.V. and Cole, C.M. (1996) Pleiotropic nuclear defects associated with a conditional allele of the novel nucleoporin Rat9p/Nup85p. *Mol. Biol. Cell*, **7**, 917-934.

Görlich, D., Draft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R.A., Mattaj, I.W. and Izaurraide, E. (1996) Importin provides a link between nuclear protein import and U snRNA export. *Cell*, in press.

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

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

Görlich, D., Prehn, S., Laskey, R.A. and Hartmann, E. (1994) Isolation of a protein that is essential for the first step of nuclear protein import. *Cell*, **79**, 767-778.

Görlich, D., Vogel, F., Mills, A.D., Hartmann, E. and Laskey, R.A. (1995b) Distinct functions for the two importin subunits in nuclear protein import. *Nature*, **377**, 246-248.

Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammelian cells by tetracyclin-responsive promoters. *Proc. Natl Acad. Sci. USA*, **89**, 5547-5551.

Grandi, P., Doye, V. and Hurt, E.C. (1993) Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. *EMBO J.*, **12**, 3061-3071.

Grandi, P., Emig, S., Weise, C., Hucho, F., Pohl, F., Pohl, T. and Hurt, E.C. (1995) A novel nuclear pore protein Nup82p which specifically binds to a fraction of Nsp1p. *J. Cell Biol.*, **130**, 1263-1273.

Guan, T., Muller, S., Klier, G., Panté, N., Blevitt, J.M., Haner, M., Paschal, B., Aebi, U. and Gerace, L. (1995) Structural analysis of the p62 complex, an assembly of O-linked glycoproteins that localizes near the central gated channel of the nuclear pore complex. *Mol. Biol. Cell*, **6**, 1591-1603.

Hancock, K. and Tsang, V.C.W. (1983) India ink staining of proteins on nitrocellulose paper. *Anal. Biochem.*, **133**, 157-162.

Hurwitz, M.E. and Blobel, G. (1995) NUP82 is an essential yeast nucleoporin required for Poly(A)+ RNA export. *J. Cell Biol.*, **130**, 1275-1281.

Kadowaki, T., Goldfarb, D., Spitz, L.M., Tartakoff, A.M. and Ohno, M. (1993) Regulation of RNA processing and transport by a nuclear guanine nucleotide release protein and members of the Ras superfamily. *EMBO J.*, **12**, 2929-2937.

Kraemer, D., Wozniak, R.W., Blobel, G. and Radu, A. (1994) The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. *Proc. Natl Acad. Sci. USA*, **91**, 1519-1523.

Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162-1164.

Melchior, F., Paschal, B., Evans, J. and Gerace, L. (1993) 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. [published erratum appears in *J. Cell Biol.* **124**, 217 (1994).]

Moore, M.S. and Blobel, G. (1993) The GTP-binding protein Ran/TC4 is required for protein import into the nucleus. *Nature*, **365**, 661-663.

Moroianu, J., Hijikata, M., Blobel, G. and Radu, A. (1995) Mammalian karyopherin $\alpha_1\beta$ and $\alpha_2\beta$ heterodimers: α_1 or α_2 subunit binds nuclear localization signal and β subunit interacts with peptide repeat-containing nucleoporins. *Proc. Natl Acad. Sci. USA*, **92**, 6532-6536.

Nehrbass, U. and Blobel, G. (1996) Role of the nuclear transport factor p10 in nuclear import. *Science*, **272**, 120-122.

Nishi, K., Yoshida, M., Fujiwara, D., Nishikawa, M., Horinouchi, S. and 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.

Panté, N. and Aebi, U. (1994) Toward the molecular details of the nuclear pore complex. *J. Struct. Biol.*, **113**, 179-189.

Panté, N. and Aebi, U. (1996) Toward the molecular dissection of protein import into nuclei. *Curr. Opin. Cell Biol.*, **8**, 397-406.

Panté, N., Bastos, R., McMorrow, I., Burke, B. and Aebi, U. (1994) Interactions and three-dimensional localization of a group of nuclear pore complex proteins. *J. Cell Biol.*, **126**, 603-617.

Perry, R.P. and Kelley, D.E. (1970) Inhibition of RNA synthesis by actinomycin D: Characteristic dose-response of different RNA species. *J. Cell Physiol.*, **76**, 127-139.

Radu, A., Blobel, G. and Moore, M.S. (1995a) Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. *Proc. Natl Acad. Sci. USA*, **92**, 1769-1773.

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

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

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York.

Schlenstedt, G. (1996) Protein import into the nucleus. *FEBS Lett.*, **389**, 75-79.

Schlenstedt, G., Wong, D.H., Koepp, D.M. and Silver, P.A. (1995) Mutants in a yeast Ran binding protein are defective in nuclear transport. *EMBO J.*, **14**, 5367-5378.

Simos, G. and Hurt, E.C. (1995) Nucleocytoplasmic transport: Factors and mechanisms. *FEBS Lett.*, **369**, 107-112.

Siniossoglou, S., Wimmer, C., Rieger, M., Doye, V., Tekotte, H., Weise, C., Emig, S., Segref, A. and Hurt, E.C. (1996) A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell*, **84**, 265-275.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*, **67**, 31-40.

Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy*. W.H. Freeman and Company, San Francisco, CA.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60-89.

Sukegawa, J. and Blobel, G. (1993) A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. *Cell*, **72**, 29-38.

Toda, T., Shimanuki, M., Saka, Y., Yamano, H., Adachi, Y., Shirakawa, M., Kyogoku, Y. and Yanagida, M. (1992) Fission yeast pap1-dependent transcription is negatively regulated by an essential nuclear protein, crm1. *Mol. Cell. Biol.*, **12**, 5474-5484.

Turi, T.G., Webster, P. and Rose, J.K. (1994) Brefeldin a sensitivity and resistance in *Schizosaccharomyces pombe. J. Biol. Chem.*, **269**, 24229-24236.

van Deursen, J., Boer, J., Kasper, L. and Grosveld, G. (1996) G_2 arrest and impaired nucleocytoplasmic transport in mouse embryos lacking the proto-oncogene *CAN/Nup214. EMBO J.*, in press.

von Lindern, M., Fornerod, M., van Baal, S., Jaeglé, M., de Wit, T., Buijs, A. and Grosveld, G. (1992a) The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA. *Mol. Cell. Biol.*, **12**, 1687-1697.

von Lindern, M., van Baal, S., Wiegant, J., Raap, A., Hagemeijer, A. and Grosveld, G. (1992b) *can*, a putative oncogene associated with myeloid leukemogenesis, can be activated by fusion of its 3' half to different genes: characterization of the set gene. *Mol. Cell. Biol.*, **12**, 3346-3355.

Williams, K.R., Samandar, S.M., Stone, K.L., Saylor, M. and Rush, J. (1996) Matrix-assisted laser desorption ionization mass spectrometry as a complement to internal protein sequencing. In Walker, J.M. (ed.), *The protein protocols handbook*. Humana Press, Totowa, NJ, pp. 541-555.

Williams, K.R. and Stone, K.L. (1995) In gel digestion of SDS-PAGE separated proteins: Observations from internal sequencing of 25 proteins. In Crabb, J. (ed.), *Techniques in protein chemistry*. Academic Press, San Diego, CA, pp. 143-152.

Wilson, I.A., Niman, H.L., Hougten, R.A., Cherenson, A.R., Connolly, M.L. and Lerner, R.A. (1984) The structure of an antigenic determinant in a protein. *Cell*, **37**, 767-778.

Xiao, Z., McGrew, J.T., Schroeder, A.J. and Fitzgerald-Hayes, M. (1993) *CSE1* and *CSE2*, two new genes required for accurate mitotic chromosome segration in *Saccharomyces cerevisiae. Mol. Cell. Biol.*, **13**, 4691-4702.

Chapter 7

Chromosomal Localization of Genes Encoding CAN/Nup214-interacting Proteins — *hCRM1* localizes to 2p16, whereas *Nup88* localizes to 17p13 and is physically linked to *SF2p32*.

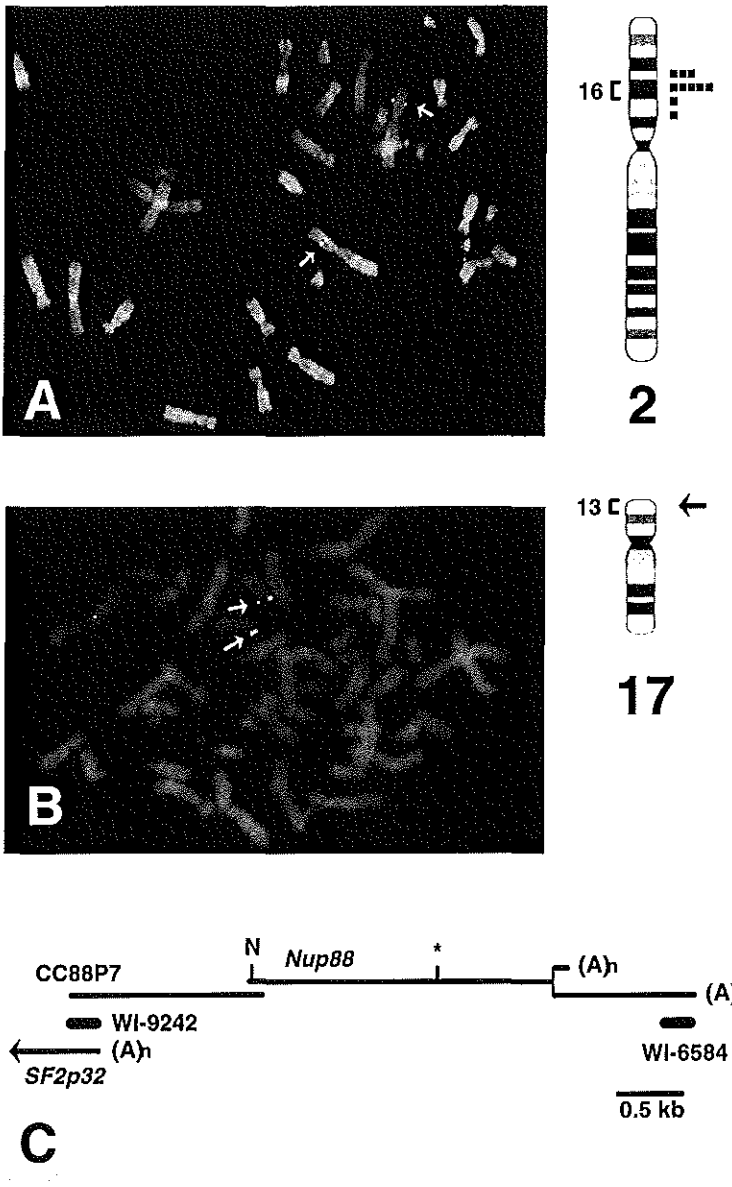
Maarten Fornerod[1], Sjozef van Baal[1], Virginia Valentine[2], David N. Shapiro[2], and Gerard Grosveld[1].

Departments of [1]Genetics and [2]Experimental Oncology, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN38105, U.S.A.

Submitted for publication.

The *CAN/Nup214* proto-oncogene, located on human chromosome 9q34, is involved in two different chromosomal rearrangements. Translocation (6;9)(p23;q34) defines a specific subtype of acute myeloid leukemia, and creates a chimeric *DEK-CAN* gene (13, 14). A rearrangement involving chromosomal band 9q34 juxtaposes *CAN* to *SET* and gives rise to *SET-CAN*, a fusion gene originally detected in a patient with acute undifferentiated leukemia (15). The CAN/Nup214 protein is a nuclear pore complex (NPC) component, whose depletion results in defective nucleocytoplasmic transport and cell cycle arrest (12). However, its contribution to leukemogenesis is not yet understood. Recently, we identified, cloned and characterized two proteins that physically interact with CAN, named Nup88 and hCRM1 (3, 4). The Nup88 protein binds to the central region of CAN, which contains several protein interaction domains. Because this region is disrupted in DEK-CAN and SET-CAN, Nup88 does not associate with the fusion proteins. hCRM1 interacts with the C-terminal nucleoporin-specific region that is common to CAN and its derived fusion proteins (3). Therefore, hCRM1 could contribute an essential function to the leukemogenic effects of DEK-CAN and SET-CAN. Proteins that interact with oncogene products are frequently involved in oncogenesis themselves. To address the possibility that *Nup88* or *hCRM1* could be targeted for cytogenetic alterations, we determined the chromosomal localization of these two genes, using fluorescent *in situ* hybridization (FISH)(11).

For *hCRM1* (Accession# Y08614), a probe representing nucleotides 1830-4223 of the *hCRM1* cDNA was labeled with dioxigenin-11-dUTP (Boehringer Mannheim, Indianapolis, IN) by nick translation. The labeled probe was preannealed with sheared human DNA and


A
16
2
B
13
17
C
CC88P7
N
Nup88
*
(A)n
(A)n
WI-9242
(A)n
SF2p32
WI-6584
0.5 kb

hybridized to normal metaphase chromosomes derived from PHA-stimulated peripheral blood lymphocytes in a solution containing 50% formamide, 10% dextran sulphate, and 2X SSC. Specific hybridization signals were detected by incubating the hybridized slides with fluorescein-conjugated sheep antibodies to digoxigenin. The chromosomes were counterstained with 4,6-diamino-2-phenylindole (DAPI). The probe detected a specific signal on the short arm of chromosome 2 (Figure 1A). Fractional length measurements of 10 specifically hybridized chromosomes 2 indicated that hCRM1 was located at 40% of the distance from the centromere to the telomere of chromosome arm 2p, an area which corresponds to chromosome band 2p16 (Figure 1A). This region has not been implicated in recurrent leukemia-associated aberrations so far (10).

cDNAs encoding Nup88 were derived from the dBEST database and from screening a human placental cDNA library (4). By combining sequence information from these two sources, we constructed two overlapping cDNA contigs that diverged at their 3' ends (Figure 1C). Both cDNAs, measured 2359 (Accession# Y08612) and 3483 bp (Y08613) respectively, and were colinear until nucleotide 2323, which is 3' of the proposed stop codon. Northern blot analysis showed a major transcript of ~2.5 kb expressed in every tissue analyzed, but particularly in testes, and a minor transcript of ~3.5 kb, that was most abundant in brain (data not shown). Interestingly, the 3' UTR of the longer cDNA contained the sequence-tagged site (STS) WI-6584 (Figure 1C), which was previously localized by radiation mapping to 38.4 cR (approximately 10 cM) from the top of the chromosome 17 linking group (7). WI-6584 lies approximately 11 cR (approximately 3 cM) telomeric from D17S938, an anonymous marker that has been cytogenetically mapped to 17p13.1 (9). Therefore, both WI-6584 and *Nup88* must be located on 17p13, most likely in the proximal subbands (17p13.1-13.2). A phage representing a partially spliced 5' *Nup88* mRNA (CC88P7) was isolated from a placental cDNA library and overlapped with STS WI-9242 (Figure 1C). This marker has been unambiguously mapped to the same cosmid (961-C-4) as WI-6584 (7), confirming *Nup88*'s linkage to this region of chromosome 17. STS WI-9242 is part of the 3' UTR of a gene encoding the p32 subunit of pre-mRNA splicing factor 2, *SF2p32* (6)(Figure 1C), establishing physical linkage

Figure 1 Chromosomal localization of *hCRM1* and *Nup88*. **(A-B)** *In situ* hybridization of normal human metaphase chromosomes with hCRM1 (A) or Nup88 (B) cDNA probes. Arrows indicate specific hybridization signals. Ideograms to the right, representing G-banded chromosomes 2 and 17, show the distribution of hybridization signals (A) or indicate the localization of the gene (B). (C) Physical linkage of the *Nup88* cDNA with sequence-tagged sites WI-6584 and WI-9242, which are located near to the top of the chromosome 17 linkage group. Lines represent the two *Nup88* cDNAs that are differentially polyadenylated and overlap at the 5' site with phage CC88P7. The 5' end of CC88P7 overlaps with the 3' end of *SF2p32*. Bars denote sequence identity to WI8242 or WI-6584. A(n), polyadenylation site; N, *NotI*; *, translational stop of *Nup88* cDNA.

between this gene and *Nup88*. Since CC88P7 is derived from a cDNA library and is partially spliced, we do not know whether its size (1.5 kb) represents the actual physical distance between the *Nup88* and *SF2p32* loci. To corroborate the radiation mapping data, we localized *Nup88* on metaphase chromosomes by FISH using a 2.4 kb *Nup88* cDNA probe. We detected a specific signal on the short arm of a chromosome that, based on size and morphology, is consistent with chromosome 17 (Figure 1B). Also the FISH location agrees with a physical location to p13. A number of genetic disorders have been mapped to the proximal part of 17p13, including three retinal disorders: Leber's congenital amaurosis (2), autosomal dominant progressive cone dystrophy (CORD5) (1), and autosomal dominant retinitis pigmentosa (5). Infantile nephrophatic cystinosis has also been linked to this chromosomal area (8). Further studies are required to determine whether *Nup88* or *SF2p32* might have roles in these disorders.

Acknowledgements

We thank Dr. Licia Selleri for helpful discussions, Dr. Sue Valance for scientific editing, and Peggy Burdick for secretarial assistance. These studies were supported by Cancer CORE Grant CA-21765, and by the Associated Lebanese Syrian American Charities (ALSAC) of St. Jude Children's Research Hospital.

References

1. Balciuniene, J., K. Johansson, O. Sandgren, L. Wachtmeister, G. Holmgren, and K. Forsman (1995). A Gene for Autosomal Dominant Progressive Cone Dystrophy (CORD5) Maps to Chromosome 17p12-p13. *Genomics* **30**: 281-286.
2. Camuzat, A., J. M. Rozet, H. Dollfus, S. Gerber, I. Perrault, J. Weissenbach, A. Munnich, and J. Kaplan (1996). Evidence of Genetic heterogeneity of Leber's Congenital Amaurosis (LCA) and Mapping of LCA1 to Chromosome 17p13. *Hum. Genet.* **97**: 798-801.
3. Fornerod, M., J. Boer, S. van Baal, H. Morreau, and G. Grosveld (1996). Interaction of cellular proteins with the leukemia specific fusion proteins DEK-CAN and SET-CAN and their normal counterpart, the nucleoporin CAN. *Oncogene* **13**: 1801-1809.
4. Fornerod, M., J. van Deursen, S. van Baal, A. Reynolds, D. Davis, K. G. Murti, and G. Grosveld (1996). The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. *EMBO J.*, in press.
5. Goliath, R., Y. Shugard, P. Janssens, J. Weissenbach, P. Beighton, R. Ramasar, and J. Greenberg (1995). Fine Localization of the Locus for Autosomal Dominant Retinis

Pigmentosa on Chromosome 17p. *Am. J. Hum. Genet.* **57**: 962-965.

6. Honore, B, P. Madsen, H. H. Rasmussen, J. Vandekerckhove, J. E. Celis and H. Leffers (1993). Cloning and Expression of a cDNA covering the Complete Coding Region of the P32 Subunit of Human pre-mRNA Splicing Factor SF2. *Gene* **134**: 283-287.
7. Hudson, T., L. Stein, S. Gerety, J. Ma, A. Castle, J. Silva, D. Slonim, R. Baptista, L. Kruglyak, *et al.* (1995). An STS-based map of the human genome. *Science* **270**: 1945-1954.
8. Jean, G., A. Fuchshuber, M. M. Town, O. Gribouval, J. A. Schneider, M. Broyer, W. van 't Hoff, P. Niaudet, and C. Antignac (1996). High-Resolution Mapping of the Gene for Cysteinosis, Using Combined Biochemical and Linkage Analysis. *Am. J. Hum. Genet.* **58**: 535-543.
9. Kojis, T. L., C. Heinzmann, P. Flodman, J. T. Ngo, R. S. Sparkes, M. A. Spence, J. B. Bateman, and J. R. Heckenlively (1996). Map Refinement of Locus RP13 to Human Chromosome 17p13.3 in a Second Family with Autosomal Dominant Retinis Pigmentosa. *Am. J. Hum. Genet.* **58**: 347-355.
10. Mitelman, F. (1994). Catalog of Chromsome Aberrations in Cancer. Vol. 4. Wiley-Liss Publications, New York.
11. Lichter, P, C.J. Tang, K. Call, G. Hermanson, G.A. Evans, D. Housman, and D.C. Ward (1990). High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247**: 64-69.
12. van Deursen, J., J. Boer, L. Kasper, and G. Grosveld (1996). G_2 arrest and impaired nucleocytoplasmic transport in mouse embryos lacking the proto-oncogene *CAN/Nup214*. *EMBO J.*, in press.
13. von Lindern, M., M. Fornerod, S. van Baal, M. Jaeglé, T. de Wit, A. Buijs, and G. Grosveld (1992). The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA. *Mol. Cell. Biol.* **12**: 1687-1697.
14. von Lindern, M., A. Poustka, H. Lerach, and G. Grosveld (1990). The (6;9) chromosome translocation, associated with a specific subtype of acute nonlymphocytic leukemia, leads to aberrant transcription of a target gene on 9q34. *Mol. Cell. Biol.* **10**: 4016-4026.
15. von Lindern, M., S. van Baal, J. Wiegant, A. Raap, A. Hagemeijer, and G. Grosveld (1992). *Can*, a putative oncogene associated with myeloid leukemogenesis, can be activated by fusion of its 3' half to different genes: characterization of the *set* gene. *Mol. Cell. Biol.* **12**: 3346-3355.

Chapter 8

Cre-mediated site-specific translocation between nonhomologous mouse chromosomes

J. VAN DEURSEN*, M. FORNEROD*, B. VAN REES†, AND G. GROSVELD*

*Department of Genetics, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105; †Department of Genetics, Erasmus University of Rotterdam, Dr. Molenwaterplein 50, 3000 DR Rotterdam, The Netherlands

Communicated by David M. Prescott, University of Colorado, Boulder, CO, May 3, 1995 (received for review March 1, 1995)

ABSTRACT **Chromosome rearrangements, such as large deletions, inversions, or translocations, mediate migration of large DNA segments within or between chromosomes, which can have major effects on cellular genetic control. A method for chromosome manipulation would be very useful for studying the consequences of large-scale DNA rearrangements in mammalian cells or animals. With the use of the Cre-*loxP* recombination system of bacteriophage P1, we induced a site-specific translocation between the *Dek* gene on chromosome 13 and the *Can* gene on chromosome 2 in mouse embryonic stem cells. The estimated frequency of Cre-mediated translocation between the nonhomologous mouse chromosomes is approximately 1 in 1200–2400 embryonic stem cells expressing Cre recombinase. These results demonstrate the feasibility of site-specific recombination systems for chromosome manipulation in mammalian cells *in vivo*, breaking ground for chromosome engineering.**

Chromosomal rearrangements caused by genetic-recombination processes can alter the specific combination of genes present in any particular genome and, moreover, the spatiotemporal level of expression of these genes. To study the effects of long-distance intra- or interchromosomal-recombination events on gene expression and function, as well as chromosome structure and organization in mammals, the development of an *in vivo* approach for chromosome manipulation is essential.

The Cre-*loxP* recombination system of bacteriophage P1 has been shown to mediate efficient recombination (excision) of DNA sequences between *loxP* sites in close proximity in the genome in both mouse embryonic stem (ES) cells (1) and transgenic mice (2–4). The aim of the present study was to test whether the Cre-*loxP* system can also be applied to induce site-specific recombinations between target sequences located on nonhomologous chromosomes, extending genetic manipulations into the field of chromosome engineering. To this end, we endeavored to mimic human chromosomal translocation (6;9) (5, 6), which is associated with a specific subtype of acute myeloid leukemia, in mouse ES cells. Human t(6;9) results in head-to-tail fusion of the *DEK* gene located at 6p23 to the *CAN* gene at 9q34 (7). The leukemia-specific fusion gene produces a 5.5-kb mRNA encoding a DEK–CAN chimeric protein consisting of the near complete DEK polypeptide fused to the C-terminal two-thirds of the CAN protein (6, 7). In the mouse genome, the *Dek* and *Can* genes are also located on nonhomologous chromosomes. *Dek* is positioned on chromosome 13 (8, 9) and *Can* on chromosome 2 (8), and generation of a murine *Dek–Can* fusion gene via *loxP* sites and Cre recombinase would therefore result from an interchromosomal recombination event.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Construction of Targeting Vectors. To construct targeting vector pDek-loxP-hygro, the *loxP* site from pGEM-30 (1) was excised by digestion with *Sal* I and *Xho* I and inserted into the *Xho* I restriction site of pGEM-hygro (10), generating pGEM-loxP-hygro. The *loxP-hygro* insert was cloned as a *Hin*dIII cassette into the *Hin*dIII restriction site located within the *Dek* intron containing the breakpoints (icb) of an 8.5-kb mouse genomic *Dek* fragment isolated from a 129/Sv/E library in phage EMBL3 (10, 11). To obtain targeting vector pCan-neo-loxP, pneo-loxP was constructed by replacing the 2.0-kb *Bam*HI herpes simplex virus thymidine kinase fragment from pGH-1 (1) with an *Sph* I linker sequence. The *neo-loxP* insert was excised from the resulting vector by partial *Sph* I digestion and ligated into the *Sph* I restriction site within the *Can*-icb of a 12.7-kb genomic *Can* fragment isolated from a 129/Sv/E DNA library.

Generation and Identification of Targeted ES Clones. A total of 8×10^6 ES cells (E14) suspended in 0.4 ml of phosphate-buffered saline (PBS; 17 mM KH_2PO_4/5 mM Na_2HPO_4/150 mM NaCl, pH 7.4) were electroporated with 10 μg of targeting vector DNA (pDek-loxP-hygro linearized with *Sal* I and pCan-neo-loxP excised with *Hin*dIII) by using a Progenetor II gene pulser (1080 μF and 128 V for 1 sec). Subsequently, hygromycin B- and G418-resistant clones were selected and screened by Southern blot analysis as described (12).

Detection of Cre-Mediated Translocations by PCR. Genomic DNA (1 μg) was subjected to 30 cycles of amplification by primer combination 1 (forward *Dek* primer, 5′-CGGCATGAGGTGAGGGAC-3′) and 3 (reverse *Can* primer, 5′-TCCCAGTCCGAGGCTCTG-3′) (see Fig. 1*C*). PCR was carried out in the following mixture with a final volume of 50 μl: 50 mM Tris·HCl, pH 8.5/15 mM $(NH_4)_2SO_4$/3.5 mM $MgCl_2$/250 mM of each dNTP/200 ng of each primer/5 units of *Taq* DNA polymerase (Cetus). Temperature cycling was 94°C for 1 min, 55°C for 2 min, and 72°C for 2 min. From each PCR sample, 1 μl was used as a substrate for 30 further amplifications by nested primers 2 (forward *Dek* primer, 5′-CGGAGCAGAGAGCTCCAG-3′) and 4 (reverse *Can* primer, 5′-CTATCCCCAGCGAGCATG-3′) (see Fig. 1*C*). PCR mixtures and temperature cyclings were as described above. From each PCR sample, 10 μl was resolved on 1.5% agarose gels, DNA was blotted to Hybond-N+ (Amersham) in NaOH, and blots were probed subsequentially with internal *Dek* and *Can* oligonucleotides (5′-GCTACCTGCTCAGTTATCATGC-3′ and 5′-GAGCTCAGGCCAGACCTCC-3′, respectively) and a 109-bp *Xho* I–*Sal* I *loxP* fragment from pGEM-30. The 343-bp PCR fragment was inserted into vector pCR II (Invitrogen) according to the procedure indicated by the manufacturer, and DNA sequence analysis was performed as described (13).

Abbreviations: ES cells, embryonic stem cells; icb, intron containing the breakpoints.

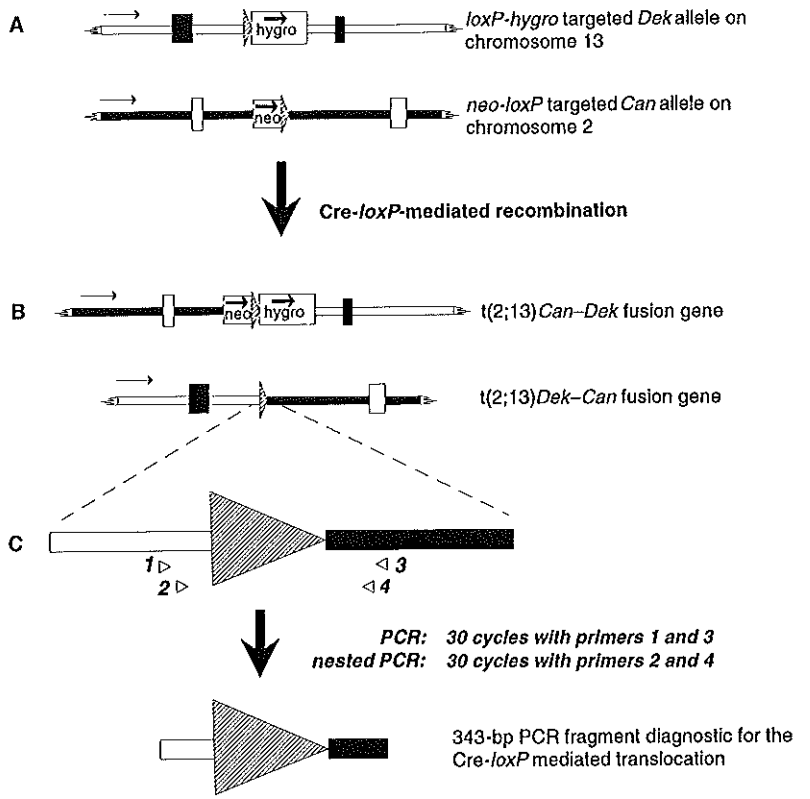


FIG. 1. Strategy for Cre-*loxP* mediated t(2;13) generating a *Dek–Can* fusion gene. (*A*) Targeted *Dek* and *Can* alleles with *loxP* sites at positions within introns analogous to the icbs in human t(6;9) (the mouse analogous introns will be referred to as *Dek*- and *Can*-icb, respectively). (*B*) The products of Cre-*loxP*-mediated translocation. (*C*) The PCR assay designed for identifying *Dek–Can* fusion genes created by Cre-*loxP* recombination. *Dek* and *Can* exons are indicated as (high) solid and open boxes, while *Dek* and *Can* introns are indicated as open and solid bars, respectively. *loxP* sites are represented by hatched triangles. Arrows mark gene orientation. Open arrowheads designate PCR primers 1–4.

RESULTS AND DISCUSSION

Our strategy for Cre-*loxP*-mediated translocation between murine *Dek* and *Can* is outlined schematically in Fig. 1. First, *loxP* sites are targeted into the endogenous *Dek* and *Can* genes of mouse ES cells by two consecutive rounds of homologous recombination at positions analogous to the icbs in the human t(6;9) (6) homologues. Subsequently, the *cre* recombinase gene is expressed transiently in the double-targeted ES clones to enable Cre enzyme-mediated recombination. Cells carrying a Cre-*loxP*-dependent translocation between chromosomes 2 and 13 are identified by PCR analysis using *Dek*- and *Can*-specific primers. Finally, as proof for Cre-mediated recombination, the identity of diagnostic chimeric PCR fragments is verified by DNA sequence analysis.

Fig. 2 shows the two-step gene-targeting scheme employed to generate ES cell clones with *loxP* sequences in *Dek* and *Can* at positions homologous to the breakpoint-specific introns of human t(6;9). In the first step, E14 ES cells were transfected with linearized targeting construct pDek-loxP-hygro (Fig. 2*A*), which has ≈8.5 kb of homology with the endogenous mouse *Dek* gene. Of 550 hygromycin B-resistant clones, 22 were homologous recombinants as identified by Southern blot analysis (Fig. 2*B*), representing a gene targeting frequency of 1 in 25. Two independent primary targeted ES clones, E14-*Dek-loxP*-210 and -493, were submitted to the second round of gene targeting in which linearized vector pCan-neo-loxP, carrying 12.7 kb of homology with the genomic *Can* locus, was introduced by electroporation (Fig. 2*C*). In total, 300 neomycin-resistant clones were analyzed by Southern blotting and about 1 in 3 displayed the desired *loxP* insertion in the endogenous *Can* locus (Fig. 2*D*).

To provoke *loxP*-directed chromosome translocations, we expressed the Cre enzyme in eight independent, double-targeted ES cell clones by transfection with supercoiled Cre-encoding plasmid pMC-cre (1). ES cells were harvested 72 h postelectroporation, and genomic DNA was isolated. DNA samples of 1 μg, representing the DNA content of ≈1 × 10^5 ES cells, were screened by nested PCR analysis for Cre-mediated t(2;13) events (see Fig. 1*C*). Five of 24 PCR samples were found to contain an amplified DNA fragment of the expected size (343 bp) that hybridized to *Dek*-, *Can*-, and *loxP*-specific probes (Fig. 3). This PCR fragment was never present in control amplification reactions on genomic DNA from ES clones with single *loxP* insertions in either *Dek* or *Can*. We inserted the 343-bp PCR fragment into vector pCR II and determined the nucleotide sequence. As shown in Fig. 4, the DNA sequence matched that of a fragment anticipated from

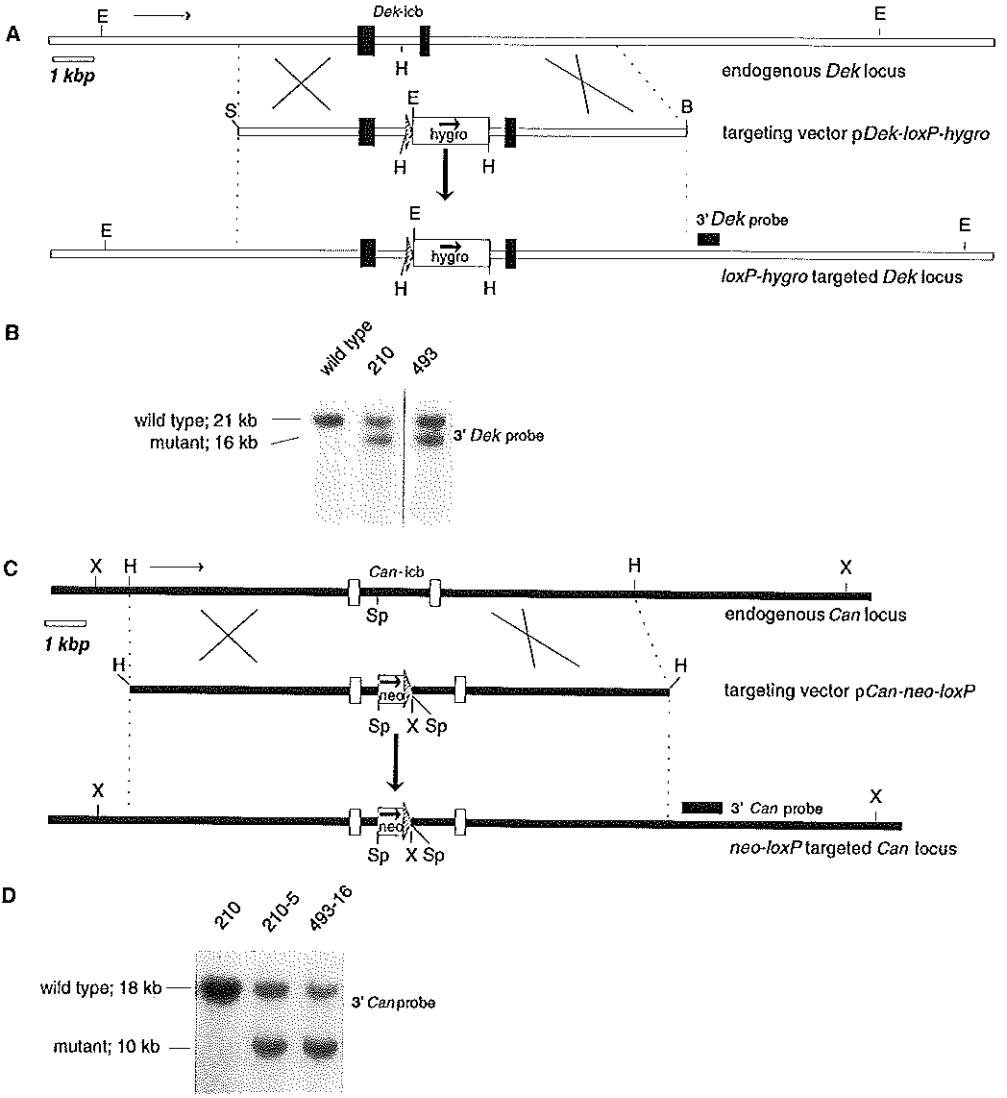


FIG. 2. Insertion of *loxP* sites in *Dek* and *Can* of mouse ES cells via two subsequent rounds of gene targeting. (*A*) Genomic *Dek* segment containing the intron analogous to the human *DEK* icb in t(6;9); the targeting vector pDek-loxP-hygro, in which a *loxP-hygro* cassette was inserted into the *Hin*dIII site of *Dek*-icb; and the mutant *Dek* allele after homologous recombination. (*B*) Southern blot analysis. *Eco*RI-cleaved genomic DNA from wild-type ES cells and homologous recombinants Dek-loxP-hygro-210 and -493 was hybridized with a 0.5-kb Pst I fragment (see *A*, 3′ *Dek* probe), giving rise to a 21-kb and 16-kb fragment for the wild-type and mutant alleles, respectively. (*C*) Partial genomic map of the mouse *Can* gene comprising the intron analogous to the human *CAN* icb in t(6;9); the gene targeting construct, in which a *neo-loxP* cassette was inserted into the *Sph* I site of *Can*-icb; and the targeted *Can* allele. (*D*) Southern analysis of the *Can-neo-loxP* mutation. A Southern blot containing ES cell DNA digested with *Xba* I was probed with a 1.0-kb *Bgl* II–*Eco*RI genomic *Can* fragment (see *C*, 3′ *Can* probe), detecting a wild-type and targeted fragment of 18 and 10 kb, respectively. Lane 210, ES clone *Dek-loxP-hygro*-210 as control; lanes 210-5 and 493-16, DNA samples from two double-targeted ES clones. Exons and introns of *Dek* and *Can*, and *loxP* sequences are indicated as in Fig. 1. Closed bars represent probes used for hybridization. Gene orientation is shown by horizontal arrows. Abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal* I; Sp, *Sph* I; and X, *Xba* I.

a Cre-*loxP*-mediated translocation between chromosomes 2 and 13, containing *Dek* 5′ *loxP*-flanking sequences, a single *loxP* core, 3′ *loxP*-flanking sequences, and *Can* sequences.

The sensitivity of the nested-PCR assay was tested by using plasmid pDek-Can (not shown). From this plasmid, the same chimeric DNA sequences present in the 343-bp t(2;13)-specific PCR fragment, with the exception of the 109-bp *loxP* cassette, can be amplified. This recombination motif was omitted to exclude any false-positive results in the experimental samples caused by physical carryover of test plasmid DNA or its PCR products. Nested-PCR amplifications on mixtures of 1 μg of genomic DNA from wild-type E14 ES cells and various amounts of mixed-in test

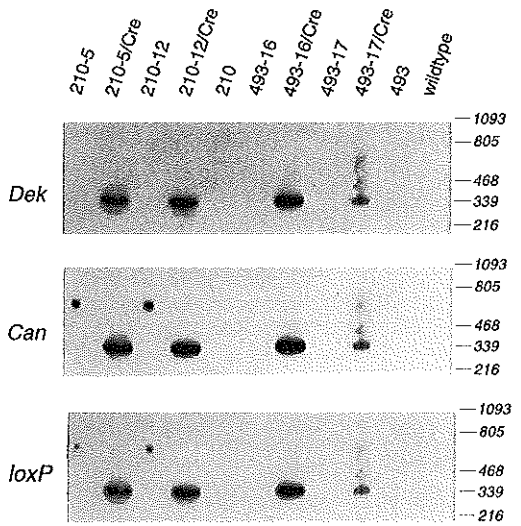


FIG. 3. Analysis of Cre-*loxP*-mediated (2;13) translocations. Southern blots of PCR-amplified genomic DNAs from four double-targeted ES clones (210-5, 210-12, 493-16, and 493-17) and three controls (210, 493, and wild type) illustrating Cre-mediated translocations. A 343-bp PCR fragment diagnostic for Cre-*loxP*-mediated t(2;13), hybridizing to probes specific for *Dek* (*Top*), *Can* (*Middle*), or *loxP* (*Bottom*) is specifically amplified from double-targeted clones transfected with pMC-cre (lanes indicated by Cre). The sizes of a selection of *Pst* I-digested λ DNA fragments are indicated to the right of each panel.

plasmid pDek-Can indicated a detection limit of approximately 10–20 molecules of test-plasmid DNA (data not shown). Assuming that the PCR detection limit is the same for pDek-Can and the Cre-mediated (2;13) translocations and that the transfection efficiency of pMC-cre amounts to ≈5% (1), we estimate the chromosome translocation frequency to be ≈1 in 1200–2400 ES cells expressing the Cre enzyme.

Further enrichment for Cre-mediated recombinants may be possible by application of conditional positive selection, in which drug resistance of ES cells would be dependent upon reconstruction of a positive selection marker (for instance puromycin-*N*-acetyltransferase) by the site-specific recombination event, as has been proposed by Qin *et al.* (14).

In conclusion, we have demonstrated precise chromosome rearrangement between two genes located on nonhomologous chromosomes in mammalian cells by using the Cre-*loxP* recombination system of bacteriophage P1 in combination with gene targeting in mouse ES cells. Site-specific recombinases have previously been shown to mediate interchromosomal recombination in *Saccharomyces cerevisae* (pSR1 recombinase) (15, 16), *Drosophila melanogaster* (FLP recombinase) (17, 18), and tobacco plants (Cre recombinase) (14). The combination of gene targeting, site-specific recombination, and ES cell technology provides powerful methodology for chromosome engineering in both cell culture and whole animals. Human diseases associated with chromosomal abnormalities, such as large deletions, inversions, and translocations (19), might be precisely mimicked in the mouse. However, knowledge about position and transcriptional direction of genes on mouse chromosomes is required for placing correctly oriented *loxP* sites in target genes, since that will determine the structure and fate of the rearranged chromosomes after Cre recombination. More generally, the use of this methodology may allow better insight in genetic regulation of complex gene clusters (such as Hox, immunoglobulin, or T-cell receptor clusters), phenomena like chromosomal imprinting, and X chromosome inactivation (20), or for exact gene mapping and construction of mouse strains (21).

Further studies will focus on the isolation of t(2;13) clones from our PCR-positive pools of ES cells. Subsequently, these t(2;13) clones can be tested for their expression of *Dek–Can* fusion transcripts and analyzed by fluorescence *in situ* hybridization (using chromosome-specific probes) to assess whether the Cre-mediated translocation chromosomes are monocentromeric. Furthermore, chimeric mice can be generated from Cre-mediated t(2;13) ES cells to explore putative effects of *Dek–Can* expression on tumorigenesis *in vivo*. In an alternative approach, we could create mice with *loxP* sites in *Dek* and *Can* from our double-targeted ES clones and provoke Cre-mediated (2;13) translocations specifically in the cells of the hematopoietic lineages. This can be done either by interbreeding mutant mice with transgenic mice expressing a myeloid-specific Cre recombinase construct or by infection of bone marrow from mutant mice with a retroviral vector expressing Cre recombinase, followed by transplantation into sublethally irradiated recipients. Although in both approaches the recombinase would be expressed constitutively, the low frequency of Cre-mediated reversion of the translocation would probably not hamper clonal outgrowth of t(2;13)-carrying hemopoietic precursor cells.

We thank Hua Gu and K. Rajewski for sending the plasmids pGH-1, pMC-cre, and pGEM-30; D. Bootsma and M. Jaeglé for generous help and advice; S. Morris, S. Hiebert, J. Boer, and J. van Ree for their critical reading; and Peggy Burdick for her editing of the manuscript.

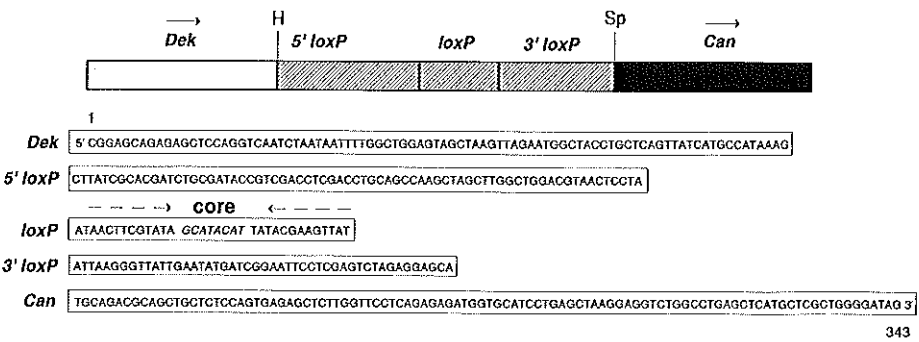


FIG. 4. Diagram and DNA sequence analysis of the 343-bp t(2;13)-specific PCR fragment. Restriction sites at the *Dek-loxP* and *loxP-Can* boundaries [*Hin*dIII (H) and *Sph* I (Sp), respectively], are indicated in the diagram. Arrows mark the orientation of *Dek* and *Can*. Broken arrows above the *loxP* sequence designate the directions of inverted repeat elements flanking the central core region (shown in italics).

This work was supported in part by the National Institutes of Health Cancer Center Support CORE grant, the American Lebanese Syrian Associated Charities, and the Dutch Cancer Foundation Koningin Wilhelmina Fonds.

1. Gu, H., Zou, Y.-R. & Rajewski, K. (1993) *Cell* **73,** 1155–1164.
2. Orban, P. C., Chui, D. & Marth, J. D. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6861–6865.
3. Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. & Rajewski, K. (1994) *Science* **265,** 103–106.
4. Araki, K., Araki, M., Miyazaki, J.-I. & Vassali, P. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 160–164.
5. Rowley, J. D. & Potter, D. (1976) *Blood* **47,** 705–721.
6. Von Lindern, M. L., Poutska, A., Lerach, H. & Grosveld, G. (1992) *Mol. Cell. Biol.* **12,** 1687–1697.
7. Von Lindern, M. L., Poutska, A., Lerach, H. & Grosveld, G. (1990) *Mol. Cell. Biol.* **10,** 4016–4026.
8. Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Epping, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., Stein, L. D., Nadeau, J. H. & Lander, E. S. (1993) *Science* **262,** 57–66.
9. Yon, J., Jones, T., Garson, K., Sheer, D. & Fried, M. (1993) *Hum. Mol. Genet.* **2,** 237–240.
10. Van Deursen, J. & Wieringa, B. (1992) *Nucleic Acids Res.* **20,** 3815–3820.
11. te Riele, H., Robanus Maandag, E. & Berns, A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5128–5132.
12. Van Deursen, J., Lovell-Badge, R., Oerlemans, F., Schepens, J. & Wieringa, B. (1991) *Nucleic Acids Res.* **19,** 2637–2643.
13. Von Lindern, M. L., Fornerod, M., van Baal, S., Jaegté, M., de Wit, T., Buijs, A. & Grosveld, G. (1992) *Mol. Cell. Biol.* **12,** 1687–1697.
14. Qin, M., Bayley, C., Stockton, T. & Ow, D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1706–1710.
15. Matzaki, H., Araki, H. & Oshima, Y. (1990) *J. Bacteriol.* **172,** 610–618.
16. Sauer, B. (1992) *J. Mol. Biol.* **223,** 911–928.
17. Golic, K. G. (1991) *Science* **252,** 958–961.
18. Chou, T. B. & Perrimon, N. (1992) *Genetics* **131,** 643–653.
19. Rabbits, T. H. (1994) *Nature (London)* **372,** 143–149.
20. Kilby, N. J., Snaith, M. R. & Murray, J. A. H. (1993) *Trends Genet.* **9,** 413–421.
21. Sauer, B. (1993) *Methods Enzymol.* **225,** 890–900.

Chapter 9: Summary and Discussion

progress can be based on incorrect but productive ideas
L. Wolpert (1995)

Summary

Translocation (6;9) is a recurrent chromosomal aberration that defines a specific subtype of acute myeloid leukemia. The translocation fuses a gene on chromosome 6, named *DEK*, to a gene on chromosome 9, named *CAN*, in a head-to-tail manner. In leukemic cells with t(6;9), a fusion messenger is expressed that encodes a chimeric DEK-CAN protein. In a similar manner, the *CAN* gene is fused to *SET* in leukemic cells with a very rare chromosomal rearrangement, leading to a chimeric SET-CAN protein. How do these two fusion proteins function in these leukemias? And, what could be the role of their normal cellular counterparts?

As a first step to answer these questions we studied the subcellular localization of DEK and CAN. It appeared that DEK was a nuclear protein, while CAN was part of the nuclear pore complex (NPC). This was a surprising and puzzling result, because NPC proteins had not earlier been implicated in oncogenesis, and the reason as to why such a protein would be oncogenic was not immediately evident. However, the finding did suggest that CAN might play a role in nucleocytoplasmic transport, which was recently shown to be the case by van Deursen *et al.* (1996). The putatively oncogenic fusion proteins DEK-CAN and SET-CAN were found only in the nucleoplasm, suggesting that relocation of the C-terminal part of CAN could be important for the putative mode of action of the two proteins, but how remained unclear.

To get more intelligible leads towards the function of DEK-CAN and SET-CAN, we set out to identify CAN-interacting proteins; this approach could also give clues about the normal function of CAN in nucleocytoplasmic transport. Via immunopurification techniques, two previously unknown proteins that interact with CAN were cloned and characterized. Nup88 is an NPC component and its association with the NPC is dependent on the presence of CAN. The protein interacts with CAN's central region, probably via coiled-coil interactions. This may well be the reason why Nup88 does not associate with DEK-CAN or SET-CAN, since the two fusion proteins lack part of the central region of CAN. The second CAN-interacting protein associates with the C-terminal FG-repeat region of CAN and is the human homologue of yeast CRM1, a protein known to be required for correct chromosome structure. This human CRM1 (hCRM1) is localized to the nuclear and cytoplasmic face of the NPC as well as the nucleoplasm and nucleolus. In *CAN-/-* mouse embryos lacking CAN, hCRM1 remained in the nuclear envelope, suggesting that this protein binds to other repeat-containing nucleoporins as well. Furthermore, we obtained evidence that hCRM1 is regularly released

from the NPC into the nucleoplasm. Interestingly, hCRM1 shares a domain of significant homology to importinβ, a cytoplasmic transport factor that interacts with nucleoporin repeat regions. Together, these data suggest that hCRM1 is a soluble nuclear transport factor that interacts with the NPC.

In addition to importinβ and its yeast homologue Kap95p, nine other proteins share the homology domain, that we named the CRIME domain (CRM1, IMportinβ, Etcetera). The function of the majority of these proteins remains to be determined, however, one of these proteins is yeast CSE1, a protein that genetically interacts with the yeast homologue of importinα. Therefore we infer that this CRIME domain could define a new group of NPC-interacting transport factors.

In contrast to Nup88, hCRM1 does associate with the DEK-CAN and SET-CAN fusion protein. Because hCRM1 is not related to any proteins known to be involved in oncogenic transformation, its role in leukemogenesis, if any, remains to be determined (see below). Genes encoding Nup88 and hCRM1 were localized on human chromosomes 17p13 (proximal) and 2p16 respectively, not coinciding with currently know cytogenetically defined oncogenic loci.

Progress in the study of translocation (6;9) would be greatly stimulated by an *in vivo* test system, recapitulating the disease. Therefore, a technique was developed for *in vivo* genome manipulation using the viral recombinase Cre. We recreated a translocation between heterologous chromosomes giving rise to a *DEK-CAN* gene in mouse embryonic stem cells. This opens the possibility to study the effects of this, and other chromosome translocations in transgenic mice.

In this chapter, I will discuss some implications of these findings, and suggest a number of future directions.

How do nucleoporin-derived oncoproteins work?

Thus far, three NPC components have been found as part of cancer or leukemia related fusion proteins, Tpr, CAN and Nup98. As discussed in Chapter 2, Tpr-containing fusion proteins probably use part of the Tpr protein for constitutive dimerization, thereby constitutively activating the kinase fusion partners. CAN and Nup98 are nucleoporins of another type, characterized by a long FG-repeat domain. Interestingly, in both CAN and Nup98 derived fusion proteins, this repeat region is incorporated (Figure 1), suggesting that the functional contribution of Nup98 and CAN could be similar. HoxA9, the fusion partner of Nup98 in t(7;11), is a DNA binding transcription factor that belongs to the Hox protein family, which has been implicated in cancer and leukemia (see Chapter 2). Therefore it is useful to review some ideas generated by discovery of the Nup98-HoxA9 fusion protein and compare them to what is known about DEK-CAN and SET-CAN.

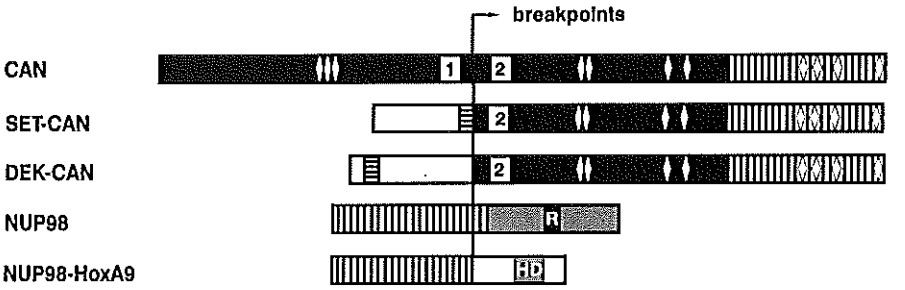


Figure 1 Similarity between Nup98 and CAN-derived fusion proteins. Proteins are represented as bars and are aligned on their break/fusion points. Predicted structural motifs are indicated as follows: FG-repeats as vertical bars, FXF-repeats as diamonds, α-helical protein interaction domains as white boxes: (1) coiled coil #1 and adjacent leucine zipper; (2) coiled-coil #2. Horizontal stripes indicate acidic regions. R, RNA binding domain; HD, homeodomain.

Maybe the most straightforward model explaining the action of the Nup98-HoxA9 fusion gene parallels that of a number of oncogenic fusion proteins (e.g. E2A-PBX1) where a DNA binding domain is fused to a heterologous transactivating domain, resulting in a chimeric transactivator, deregulating expression of key target genes. This model predicts an transcriptional regulating potential of the Nup98 moiety, something that can be tested using a reporter gene driven by a HoxA9 responsive promoter. Interestingly, also DEK and SET are implicated in DNA binding; DEK associates with metaphase chromosomes and binds DNA *in vitro* (Chapter 4), whereas SET has been shown to function as an adenovirus replication factor *in vitro* (Nagata *et al.*, 1995). Therefore, also DEK-CAN and SET-CAN may function as transcription factors, a notion that is consistent with their nuclear localization (Chapter 4).

Another model, the filter model, assumes a localization of Nup98-HoxA9 at the NPC, and proposes that essential HoxA9 cofactors could be seized by the NPC-bound fusion protein, thereby preventing them from entering the nucleoplasm. Absence of such cofactors might deregulate HoxA9-controlled transcriptional events, and cause oncogenic transformation (Borrow *et al.*, 1996). Yet another model, the gene gating model (Borrow *et al.*, 1996) also depends on a NPC association of Nup98-HoxA9, and suggests that export of transcripts from HoxA9 activated genes to the cytoplasm could be enhanced by virtue of NPC localized HoxA9 activity. Weaknesses in these models are that 1) HoxA9 is not expressed in myeloid cells (Nakamura *et al.*, 1996; Rubin *et al.*, 1987), 2) cofactors known to bind Hox proteins do so only together with DNA (Chan *et al.*, 1994; Chang *et al.*, 1995; van Dijk and Murre, 1994), and 3) information is lacking on the subcellular localization of Nup98-HoxA9. Expression studies in HeLa cells show that CAN-derived fusion proteins are located in the nucleus and not at the NPC (Chapter 4).

If nucleoporin FG-repeat regions play an important role in the fusion proteins, it seems likely that this contribution is mediated through interaction with other proteins. So far, two cellular proteins have been shown to interact with these repeat regions, importinβ (Chi *et al.*, 1995; Görlich *et al.*, 1995; Moroianu *et al.*, 1995; Radu *et al.*, 1995) and hCRM1 (Chapter 5 and 6). There are currently no indications that importinβ has a strong affinity for CAN, and in line with this, DEK-CAN and importinβ do not co-localize (Judith Boer, unpublished results). If hCRM1 indeed is a nuclear transport factor that interacts with the NPC (as argued below), nuclear localization of DEK-CAN and SET-CAN may result in an increased presence of hCRM1 in the nucleoplasm, disturbing the transport equilibrium. Both positive or dominant negative effects of such a disturbance can be envisioned, but sequestering of all hCRM1 in the nucleoplasm as can be achieved by overexpression of C-terminal CAN (Chapter 6) is unlikely considering the low expression levels of DEK-CAN in leukemic cells (Chapter 3). Alternatively, still assuming a transport function for hCRM1, CAN-linked transport molecules such as hCRM1 might carry factors important for DEK-mediated transactivation or repression. Currently, the involvement of hCRM1 in DEK-CAN related leukemogenesis can not be directly addressed due to lack of an *in vivo* or *in vitro* experimental system. However, analogous interacting proteins (or hCRM1 itself, see below) may bind to Nup98-HoxA9, a fusion protein that is possibly more amenable to experimental recapitulation of the fusion-gene related disease.

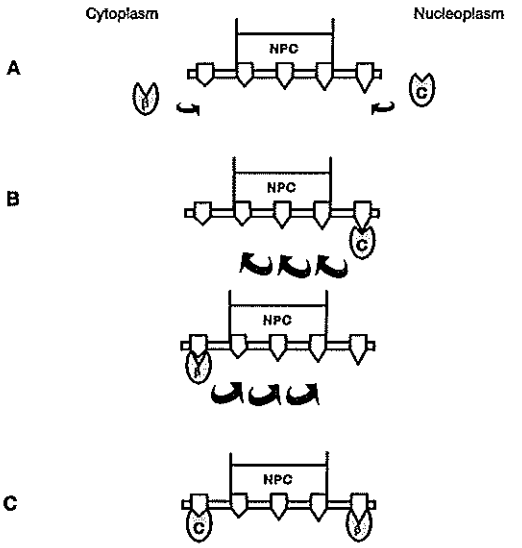


Figure 2 A model for the interaction of hCRM1 (C) with the nuclear pore complex (NPC) based on homology with importinβ (β). A, hCRM1 docks at the nuclear face of the NPC via a low-affinity interaction with a nucleoporin repeat region (pointed triangle). B and C, hCRM1 translocates through the channel via a series of docking and undocking reactions towards higher affinity binding sites. Affinity is represented by triangle fit.

Nup88

Although Nup88 does not bind DEK-CAN or SET-CAN, it is not excluded that this protein contributes to a DEK-CAN or SET-CAN-dependent transformed state. It might be necessary

that all Nup88 is complexed to CAN for proper cell function. The fusion of one *CAN* allele to *DEK* or *SET* could result in free Nup88, which may enhance the deregulated state of the cell. On the other hand Nup88 could be a negative modulator of the function of C-terminal CAN. Loss of Nup88 binding which occurs in DEK-CAN and SET-CAN may, in this scenario, contribute to the oncogenic capacity. More detailed studies of the function of Nup88 could addess these possibilities.

What is the function of hCRM1?

As discussed in Chapter 5, three lines of evidence support the idea that hCRM1 is a nuclear transport factor that dynamically interacts with the NPC. First, nuclear overexpression of the FG-repeat region of CAN, containing its hCRM1-interaction domain, resulted in depletion of hCRM1 from the NPC, suggesting that hCRM1 can move from the NPC to the nucleoplasm; second, the nuclear envelopes of cells from CAN-depleted embryos still contain hCRM1, suggesting that hCRM1 interacts with other FG-repeat containing nucleoporins; and third, hCRM1 shares a region of significant homology with the cytoplasmic transport factor importinβ. As importinβ is immunolocalized to both cytoplasmic and nuclear faces of the NPC as well as to the cytoplasm (Görlich *et al.*, 1995), its localization seems to mirror that of hCRM1. It has been hypothesized that importinβ moves through the NPC because of increased affinity for nucleoporin repeats towards the nucleus (Nehrbass and Blobel, 1996; Rexach and Blobel, 1995). Therefore it is conceivable that for hCRM1 the opposite is the case, having an increased affinity for nucleoporin repeats towards the cytoplasm (Figure 2). In order to test this hypothesis, relative affinities of hCRM1 for different nucleoporin repeats could be compared in a two-hybrid system. Alternatively, the FG-repeat of a nucleoporin located at the nucleoplasmic site of the NPC could be replaced by that of CAN, or especially its hCRM1-binding part, by gene targeting. The prediction of such experiments would be that hCRM1 accumulates at the nuclear site of the NPC, while the distribution of importinβ is unaffected.

A related but equally important question is what the transport substrate of hCRM1 would be. At the moment it is hard to say more than that a nuclear importinβ-like factor would seem to fit in a current importinα-mediated RNA export model (compare Figure 2 with Figure 3 of Chapter 2).

The CRIME family

The N-terminal homology domain present in human and yeast CRM1 and importinβ connects a group of previously unrelated proteins (Chapter 6). Apart from this domain, common characteristics appear to be their similarity in size and evolutionary conservation (Table 1).

Many of the proteins are still largely uncharacterized. However, *S. cer.* CSE1 is a protein required for correct chromosome segregation (Xiao *et al.*, 1993). Interestingly, a temperature sensitive *cse1* allele can be complemented by high expression of *srp*, the yeast homologue of the vertebrate importinα (Belanger *et al.*, 1994), making CSE1 the third otherwise unrelated protein that contains the CRIME domain and is implicated in nucleocytoplasmic transport. If the genetic interaction of *srp* with *crm1* interaction represents a physical association, a common attribute of CRIME proteins could be Srp/importinα interaction. Thus, it would be interesting to test whether hCRM1 could bind importinα or importinα-like proteins.

Another CRIME protein that should be mentioned is Nmd5. This protein was identified through a two hybrid screen with Upf1, a yeast protein involved in nonsense-mediated mRNA decay (A. Jacobson, University of Massachusetts, unpublished results). Nonsense-mediated mRNA decay refers to the phenomenon of rapid turnover of nonsense-containing mRNAs (reviewed in Maquat, 1995; Peltz *et al*, 1994). Rapid decay of such molecules in the yeast *S. cer.* depends on Upf1, Nmd2, and Upf3 proteins. Mutations of each of these proteins lead to the selective stabilization of mRNAs containing early nonsense mutations without affecting the decay rates of most other mRNAs. Mutation of Nmd5 does not seem to alter the decay rate of nonsense-containing mRNAs (A. Jacobson, unpublished results), although this does not preclude its involvement in this decay pathway. Recent studies showed that the Upf1 protein is distributed throughout the cell in approximately 200-400 clusters, predominantly in the cytoplasm, but definitively detectable in the nucleus (Peltz *et al.*, 1994). In addition, a specific two-hybrid interaction of Upf1 has been found with *SNP1*, a gene encoding a U1 snRNP 70 kDa homologue (He and Jacobson, 1995), a nuclear protein. This suggests that at least some components of Upf1-containing complexes cycle in and out of the nucleus. Therefore its homology with importinβ and (h)CRM1 suggests that Nmd5 could be an NPC interacting component of the decay complex.

Where do we go from here?

In order to understand the functional contribution of DEK-CAN and SET-CAN to oncogenesis, an experimental system that reproduces the oncogenic effect seems to be indispensable. In the absence of such a system, the significance of factors such as Nup88 or hCRM1 for the leukemic process are difficult to interpret, especially since neither protein looks like a proto-oncoprotein. It seems to me that further biochemical analyses of DEK-CAN and SET-CAN have a lower priority than studies towards *in vivo* or *in vitro* models for their action. An attractive possibility for the latter appears to be an *in vivo* mouse model generated by Cre-mediated chromosome translocation. Our results described in chapter 8 show the feasibility of such approach, which is further supported by recent studies of Corral *et al.* (1996). These authors mimicked the

Table I CRIME domain proteins

Protein	length (aa)	Characteristics[a]	Human EST[b]
hCRM1	1071	Human CAN-binding protein	
CRM1	1084	*Saccharomyces cerevisiae* protein	
CRM1+	1077	*Schizosaccharomyces pombe* protein	
CSE1	960	Required for chromosome segregation in yeast	
CAS	971	Human homologue of CSE1	
Nmd5p	1048	Putatively Upf1p interacting protein	R44286
HRC1004	1004	Hypothetical protein	W19612
SPAC22H10.03c	993	Hypothetical protein	--
YER110c	1113	Hypothetical protein	--
Pse1p	1089	Protein secretion enhancer	H29017
LPH2p	1032	Hypothetical protein	--
D9509.15p	944	Hypothetical protein	R49703, R44286
Importinβ	876	*Xenopus* NLS receptor large subunit[c]	
Kap95p	861	yeast homologue of importinβ	

[a]*Saccharomyces cerevisiae* proteins unless otherwise indicated; for references see Chapter 6, Figure 2.
[b]Accession number of human expressed sequence tags with significant homology ($p<5x10^{-5}$).
[c]Also named p97 or karyopherinβ

oncogenic translocation (9;11) in the mouse using Cre-mediated recombination, resulting in a leukemic phenotype comparable to the human disease. In the absence of a test system, the biochemical and celbiological analysis of the DEK-CAN and SET-CAN proteins can still be useful in the analysis of similar fusion proteins, such as Nup98-HoxA9, or others that are still unknown.

The findings presented in this thesis could also improve our understanding of the mechanism of nucleocytoplasmic transport. Especially hCRM1, that may be a novel transport factor might fulfil this promise. In view of the high homology between the human and yeast CRM1 proteins, it would probably be better to proceed with the analysis of this protein in yeast, an eukaryotic organism that is more amenable to genetic approaches than human or mouse. *S. pombe* and *S. cer.* CRM1 proteins have been immunolocalized very prominently in the periphery of the nucleus (Adachi and Yanagida, 1989) making it likely that also yeast CRM1 associates with the NPC via one or more nucleoporins. To test this, physical interactions between *S. cer.* CRM1 and several nucleoporins can be assessed by two-hybrid analysis (Stutz et al., 1995) or co-immunoprecipitation (Chapter 5). In addition, these experiments may confirm the hypothesis put forward by Kraemer *et al.* (1995) that Nup159 is the yeast homologue of CAN.

The *crm1* gene has three cold sensitive alleles in *S. pombe*: *crm1-809*, *crm1-119* and *crm1-N1* (Adachi and Yanagida, 1989; Nishi *et al.*, 1994). Transport defects due to CRM1

depletion thus can be easily studied in these mutants by shifting them to the restrictive temperature. RNA export defects can be assessed by *in situ* hybridization with oligo dT or rRNA-specific probes (Amberg et al., 1992), whereas protein import defects can be tested using an SV40NLS-invertase reporter protein, that normally is transported to the nucleus in an NLS-dependent fashion (Nelson and Silver, 1989). In addition, complementation studies can be performed with other CRIME family members, each of which could also be tested for nucleoporin interaction.

The antibiotic Leptomycin B (LMB) seems to target CRM1 function specifically in *S. pombe*, since (i) this compound causes the same phenotypical characteristics as mutations in *crm1*, and (ii) amplification of the wild-type *crm1* gene leads to LMB resistance (Nishi *et al.*, 1994). It will be interesting to test the effect of this antibiotic on nucleocytoplasmic transport. Once a biological model for t(6;9) has been established, it would also be interesting to test this antibiotic on DEK-CAN function.

Together, these experiments should determine whether CRM1 functions in nucleocytoplasmic transport, and whether the CRIME homology domain defines a new class of nucleoporin-interacting transport factors.

Acknowledgement I would like to thank Dr. Allan Jacobson and Judith Boer for communicating unpublished results.

References

Adachi, Y. and 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 at its perphery. *J Cell Biol*, **108**, 1195-1207.

Amberg, D. C., Goldstein, A. L., and Cole, C. N. (1992) Isolation and characterization of *RAT1*: An essential gene of *Sacceromyces cerevisiae* required for the efficient nucleoplasmic trafficking of mRNA. *Genes Dev* **6**, 1173-1189.

Belanger, K.D., Kenna, M.A., Wei, S. and Davis, L.I. (1994) Genetic and physical interactions between Srp1p and nuclear pore complex proteins Nup1p and Nup2p. *J Cell Biol*, **126**, 619-630.

Borrow, J., Shearman, A.M., Stanton, V.P., Jr., Becher, R., Collins, T., Williams, A.J., Dube, I., Katz, F., Kwong, Y.L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J. and Housman, D.E. (1996) The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nature Genet.*, **12**, 159-167.

Chan, S.K., Jaffe, L., Capovilla, M., Botas, J. and Mann, R.S. (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, **78**, 603-615.

Chang, C.P., Shen, W.F., Rosenfeld, S., Lawrence, H.J., Largman, C. and Cleary, M.L. (1995) Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev*, **9**, 663-674.

Chi, N.C., Adam, E.J.H. and Adam, S.A. (1995) Sequence and characterization of cytoplasmic nuclear protein import factor p97. *J Cell Biol*, **130**, 265-274.

Corral, J., Lavenir, I., Impey, H., Warren, A.J., Forster, A., Larson, T.A., Bell, S., McKenzie, A.N., King, G. and Rabbitts, T.H. (1996) An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell*, **85**, 853-861.

Doye, V. and Hurt, E.C. (1995) Genetic approaches to nuclear pore structure and function. *Trends Genet*, **11**, 235-241.

Görlich, D., Vogel, F., Mills, A.D., Hartmann, E. and Laskey, R.A. (1995) Distinct functions for the two importin subunits in nuclear protein import. *Nature*, **377**, 246-248.

He, F. and Jacobson, A. (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. *Genes Dev*, **9**, 437-454.

Kraemer, D.M., Strambio-de-Castillia, C., Blobel, G. and Rout, M.P. (1995) The essential yeast nucleoporin NUP159 is located on the cytoplasmic side of the nuclear pore complex and serves in karyopherin-mediated binding of transport substrate. *J Biol Chem*, **270**, 19017-19021.

Maquat, L.E. (1995) When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *Rna*, **1**, 453-465.

Moroianu, J., Hijikata, M., Blobel, G. and Radu, A. (1995) Mammalian karyopherin à1á and à2á heterodimers: à1 or à2 subunit binds nuclear localization signal and á subunit interacts with peptide repeat-containing nucleoporins. *Proc Natl Acad Sci USA*, **92**, 6532-6536.

Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A. and Matsumoto, K. (1995) Replication factor encoded by a putative oncogene, *set*, associated with myeloid leukemogenesis. *Proc Natl Acad Sci USA*, **92**, 4279-4283.

Nakamura, T., Largaespada, D.A., Shaughnessy, J.D., Jr., Jenkins, N.A. and Copeland, N.G. (1996) Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nature Genet.*, **12**, 149-153.

Nehrbass, U. and Blobel, G. (1996) Role of the nuclear transport factor p10 in nuclear import. *Science*, **272**, 120-122.

Nelson, M., and Silver, P. A. (1989) Context affects nuclear protein localization in *Saccheromyces cerevisiae*. *Mol Cell Biol* **9**, 384-389.

Nishi, K., Yoshida, M., Fujiwara, D., Nishikawa, M., Horinouchi, S., and 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.

Peltz, S.W., He, F., Welch, E. and Jacobson, A. (1994) Nonsense-mediated mRNA decay in yeast. *Prog Nucleic Acid Res Mol Biol*, **47**, 271-98.

Radu, A., Moore, M.S. and 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.

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

Rubin, M.R., King, W., Toth, L.E., Sawczuk, I.S., Levine, M.S., P, D.E. and Nguyen-Huu, M.C. (1987) Murine Hox-1.7 homeo-box gene: cloning, chromosomal location, and expression [published erratum appears in Mol Cell Biol 1988 Dec;8(12):5593]. *Mol Cell Biol*, **7**, 3836-41.

van Deursen, J., Boer, J., Kasper, L. and Grosveld, G. (1996) G_2 arrest and impaired nucleocytoplasmic transport in mouse embryos lacking the proto-oncogene *CAN/Nup214*. *EMBO J*, in press.

van Dijk, M.A. and Murre, C. (1994) extradenticle raises the Dna binding specificity of homeotic selector gene products. *Cell*, **78**, 617-624.

Wolpert, L. (1995) Evolution of the cell theory. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **349**, 227-233.

Xiao, Z., McGrew, J.T., Schroeder, A.J. and Fitzgerald-Hayes, M. (1993) *CSE1* and *CSE2*, two new genes required for accurate mitotic chromosome segration in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **13**, 4691-4702.

Het CAN Eiwit: Een Schakel in Nucleocytoplasmatisch Transport met Oncogene Eigenschappen

door

Maarten Fornerod

SAMENVATTING

Translocatie (6;9) is een chromosomale afwijking die kenmerkend is voor een bepaald type acute myeloide leukemie, die vooral voorkomt bij kinderen en jonge volwassenen. Als gevolg van deze translocatie worden twee genen, die normaal gesproken op chromosoom 6 en 9 liggen, kop-staart gefuseerd. Deze genen zijn *DEK* op chromosoom 6 en *CAN* op chromosoom 9. In leukemische cellen met translocatie (6;9) wordt een fusie-mRNA geproduceerd, dat codeert voor een chimeer DEK-CAN eiwit. In leukemische cellen met een andere, zeer zeldzame, chromosomale afwijking is het *CAN* gen gefuseerd met *SET*, en wordt waarschijnlijk een SET-CAN eiwit geproduceerd.

De vraag was waarom deze fusie-eiwitten betrokken zijn bij leukemie, en wat de normale rol van de fusie partners in de cel is. Antwoord op deze vragen zou, op lange termijn, de weg kunnen openen naar een remedie voor de betrokken leukemiën, en zou, op iets kortere termijn, inzicht kunnen geven in celgroei en differentiatie, processen die zijn verstoord bij kanker. Bovendien waren deze eiwitten het bestuderen waard, omdat ieder natuurlijk eiwit een deel van het antwoord in zich draagt over de werking en ontstaansgeschiedenis van de cel. In dit proefschrift heeft de nadruk gelegen op de studie van het CAN eiwit, omdat deze door twee verschillende leukemie-gerelateerde chromosomale afwijkingen wordt getroffen, en daarom een meer algemene rol in leukemie zou kunnen spelen.

Als eerste stap hebben we antilichamen opgewekt tegen DEK en CAN, en de subcellulaire lokalisatie van DEK, CAN en DEK-CAN bepaald. DEK bleek gelokaliseerd te zijn in de kern, terwijl CAN deel uitmaakt van het nucleaire porie complex (NPC). NPC's zijn zeer grote eiwitstructuren (ongeveer 125 megadalton) die ingebed zijn in de nucleaire envelop en de poort vormen tussen het cytoplasma en nucleoplasma. Alle macromoleculen die zich tussen deze twee cellulaire compartimenten bewegen, zoals eiwitten, RNA, en ribosomale onderdelen, moeten het NPC passeren, echter, het mechanisme van dit transport is nog grotendeels onbekend. De ontdekking dat CAN deel uitmaakt van het NPC was een verrassend resultaat, ten eerste omdat NPC componenten nog niet eerder als oncogene eiwitten geïdentificeerd waren, en ten tweede omdat het in het geheel niet duidelijk was hoe zo'n soort eiwit leukemie zou kunnen veroorzaken. Het suggereerde echter wel dat CAN een rol zou

kunnen spelen in nucleocytoplasmatisch transport, een idee dat recent bewaarheid is. Het DEK-CAN fusie eiwit bleek, evenals het SET-CAN eiwit, gelokaliseerd te zijn in de kern. Dit resultaat gaf aan dat de kern lokalisatie van een gedeelte van het CAN eiwit belangrijk zou kunnen zijn voor de oncogene rol van de twee fusie eiwitten, maar hoe was onduidelijk.

Om wat meer begrijpelijke aanwijzingen te krijgen over de functie van DEK-CAN en SET-CAN, hebben wij eiwitten geïsolereerd die in de cel een interactie aangaan met CAN. Deze aanpak zou ook aanwijzingen op kunnen leveren over de normale functie van CAN in nucleocytoplasmatisch transport. Door middel van immunozuivering hebben we twee CAN-bindende eiwitten geïdentificeerd en de overeenkomstige genen gecloneerd. Het eerste eiwit, Nup88 (voor Nucleoporin [=NPC eiwit] van 88 kilodalton) bleek, net zoals CAN, een NPC component te zijn. Het eiwit bevat een zogeheten coiled-coil domein, en bindt waarschijnlijk daarmee aan centraal gelegen eiwit sequenties van CAN. Dit zou heel goed de reden kunnen zijn dat Nup88 niet bindt aan DEK-CAN of SET-CAN, omdat deze twee een gedeelte van het centrale gebied van CAN missen. Verder hebben we gevonden dat de binding van Nup88 aan het NPC afhankelijk is van CAN. Het tweede CAN-bindende eiwit associeert met het C-terminale, gerepeteerde gedeelte van CAN. Het eiwit bleek een zeer grote gelijkenis te vertonen met het CRM1 eiwit van gist (50% identiek), een eiwit dat een (wellicht indirecte) rol speelt bij de instandhouding van een juiste chromosomale structuur. Dit eiwit, dat we hCRM1 genoemd hebben (humaan CRM1), hebben we gelokaliseerd aan beide kanten van het NPC, maar ook in de nucleus en nucleolus. In cellen zonder CAN eiwit was hCRM1, in tegenstelling tot Nup88, nog steeds aanwezig in de nucleaire envelop. Dit wijst erop dat hCRM1 ook aan andere NPC componenten kan binden, wellicht aan eiwitten die een zelfde soort gerepeteerd gedeelte bezitten als CAN. Ook hebben we aanwijzingen dat hCRM1 regelmatig loslaat van het NPC en zich in het nucleoplasma beweegt. Interessant genoeg vertoont hCRM1 een gelijkenis met importinβ, een cytoplasmatische transport factor, die contact aangaat met het NPC via gerepeteerd gebieden in nucleoporins. Bij elkaar genomen wijzen deze vindingen erop dat hCRM1 een nucleaire transport factor is die via CAN, en wellicht ook via andere nucleoporins, contact aangaat met het NPC.

Met behulp van de computer hebben wij een aantal gisteiwitten gevonden in de databank, die homologie vertonen met zowel hCRM1 als importinβ. Deze homologie ligt in alle eiwitten in een N-terminaal gebied, dat we het CRIME gebied hebben genoemd (van CRm1, IMportinβ, Etcetera). De meerderheid van deze gist eiwitten is nog nauwelijks gekarakteriseerd, echter, een van hen is CSE1, een eiwit dat een genetische interactie aangaat met een andere transport factor, genaamd importinα. Het is dus denkbaar dat het CRIME gebied een nieuwe groep van transport factoren definieert, die contact aangaan met het NPC.

In tegenstelling tot Nup88 bindt hCRM1 wel aan DEK-CAN en SET-CAN. hCRM1 behoort echter niet tot een groep eiwitten, die een gekende functie hebben in oncogenese. De vragen of dit eiwit een rol speelt in de t(6;9) leukemogenese, en zo ja, welke, blijven dus

voorlopig onbeantwoord. De genen die coderen voor Nup88 en hCRM1 zijn respectievelijk gelokaliseerd in chromosoomband 17p13 (proximaal) en 2p16, geen gebieden waar op grond van de huidige gegevens een oncogen of anti-oncogen verwacht wordt.

De voortgang van het onderzoek naar de gevolgen van translocatie (6;9) zou zeer gestimuleerd worden met een *in vivo* testsysteem. Met het oog hierop hebben we een methodiek ontwikkeld, die met behulp van een viraal recombinase chromosoom translocaties mogelijk maakt in muize embryonale stamcellen. Dit opent de mogelijkheid om de (6;9) translocatie na te bootsen in transgene muizen.

De experimentele studies beschreven in dit proefschrift hebben de vraag hoe het *CAN* gen betrokken is bij het ontstaan van leukemie niet beantwoord. Wel geven zij nieuwe richtingen aan voor verder onderzoek en dragen bij tot het inzicht dat nucleocytoplasmatisch transport een rol zou kunnen spelen bij het ontstaan van leukemie. Bovendien heeft het onderzoek elementen aangedragen, die de basale kennis over het transport tussen de kern en het cytoplasma zullen vergroten.

Curriculum Vitae

Maarten Fornerod (1963) began his academic education at the University of Leyden, The Netherlands in 1983 with a general curriculum in Biology. This was followed by a specialization in Biochemistry that included two undergraduate research projects of nine months each. The first involved construction of vectors for plant transformation, and was supervised by Dr. Lambert A.M. Hensgens and Prof. Dr. Rob A. Schilperoort at the Department of Plant Molecular Biology, University of Leyden; the second examined aspects of mouse int-1 protein biosynthesis, and was supervised by Drs. Marc J. Van de Vijver and Roel Nusse at the department of Molecular Biology of the Netherlands Cancer Institute in Amsterdam. After graduation (July 1989), Mr. Fornerod was stationed for 18 months and 20 days in the lab of Dr. Gerard C. Grosveld, Erasmus University Rotterdam, The Netherlands as a conscientious objector to military service, where he participated in ongoing research into the molecular basis of translocation (6;9)(p23;q34)-associated acute myeloid leukemia. In 1991 he remained in the same laboratory as graduate student to study the proteins involved in this type of leukemia, supervised by Dr. Gerard C. Grosveld (co-promotor) and Prof. Dr. Dirk Bootsma (promotor). In 1993 the graduate student moved with Dr. Grosveld's lab to St. Jude Children's Research Hospital in Memphis, U.S.A., and participated in the start-up of its Department of Genetics. Here, his research focussed on the cellular function of one of the proteins involved in t(6;9), CAN/Nup214. In November 1996 Maarten Fornerod was awarded a two year EMBO postdocteral fellowship on the project "The Role of *Saccharomyces cerevisiae* CRM1 and CSE1 Proteins in Nucleocytoplasmic Transport", and relocated to the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, to join the lab of Dr. Iain Mattaj, where he currently works.

Maarten Fornerod can be reached by E-mail at <maarten.fornerod@EMBL-Heidelberg.de> and/or <seismo@dds.nl>.

List of publications (in reverse chronological order)

1) **Maarten Fornerod**, Sjozef van Baal, VirginiaValentine, David N. Shapiro, and Gerard Grosveld. Chromosomal localization of genes encoding CAN/Nup214-interacting proteins: hCRM1 localizes to 2p16, whereas Nup88 localizes to 17p13 and is physically linked to SF2p32. **Genomics**, submitted.

2) Erik Bonten, Aarnoud van der Spoel, **Maarten Fornerod**, Gerard Grosveld, and Alessandra d'Azzo. Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis. **Genes & Development**, in press.

3) **Maarten Fornerod**, Jan van Deursen, Sjozèf van Baal, Donna Davis, K.Gopal Murti and Gerard Grosveld. The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. **The EMBO Journal**, in press.

4) **Maarten Fornerod**, Judith Boer, Sjozèf van Baal, Hans Morreau and Gerard Grosveld (1996). Interaction of cellular proteins with the leukemia specific fusion protein DEK-CAN and SET-CAN and their normal counterpart, the nucleoporin CAN. **Oncogene** 13: 1801-1808.

5) **Maarten Fornerod** (1996). Prevention of radioactive contamination of cell culture incubators during metabolic labelling experiments. **BioTechniques** 20: 876-877.

6) **Maarten Fornerod**, Judith Boer, Sjozèf van Baal, Martine Jaeglé, Marieke von Lindern, K. Gopal Murti, Donna Davis, Jaqueline Bonten, Arjan Buijs and Gerard Grosveld (1995). Relocation of the carboxyterminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. **Oncogene** 10: 1739-1748.

7) Jan van Deursen, **Maarten Fornerod**, Bas van Rees and Gerard Grosveld (1995). Cre-mediated site specific translocation between nonhomologous mouse chromosomes. **Proceedings of the National Academy of Sciences USA** 92: 7376-7380.

8) Yanping Yang, Elio F. Vanin, Michael A. Whitt, **Maarten Fornerod**, Ronald Zwart, Richard D. Schneiderman, Gerard Grosveld and Arthur W. Nienhuis (1995). Inducible, High level Production of infectious Murine Leukemia Retroviral vector particles pseudotyped with Vesicular Stomatitis Virus G envelope protein in a stable cell line. **Human Gene Therapy** 6: 1203-1213.

9) Marcel H.M. Koken, Hanny H.M. Odijk, Marcel van Duin, **Maarten Fornerod** and Jan H.J. Hoeijmakers (1993). Augmentation of protein production by a combination of the T7 RNA polymerase system and ubiquitin fusion. **Biochemical and Biophysical Research Communications** 195: 643-653.

10) L.A.M. Hensgens, **Maarten Fornerod**, S. Rueb, A.A. Winkler, S. Vanderveen and R.A. Schilperoort (1993). Translation Controls the Expression Level of a Chimeric Reporter Gene. **Plant Molecular Biology** 20: 921-938.

11) Marieke von Lindern, **Maarten Fornerod**, Sjozèf van Baal, Martine Jaeglé, Ton de Wit, Arjan Buijs and Gerard Grosveld (1992). The translocation (6;9) associated with a specific subtype of Acute Myeloid Leukemia, results in the fusion of two genes, *dek* and *can*, and the expression of a chimeric, leukemia-specific *dek-can* mRNA. **Molecular and Cellular Biology** 12: 1687-1697.

12) Marieke von Lindern, **Maarten Fornerod**, Nike Soekarman, Sjozèf van Baal, Martine Jaeglé, Anne Hagemeijer, Dirk Bootsma and Gerard Grosveld (1992). Translocation t(6;9) in acute non-lymphocytic Leukaemia results in the formation of a *dek-can* fusion gene. **Baillieres Clinical Haematology** 5: 857-879.

Epilogue

A well-respected scientist already told me: once you get to the acknowledgement part you will be in heaven. That certainly holds true for me, now, less than twelve hours before this manuscript is due at the print shop. It may not hold true for you, especially when it turns out I haven't even mentioned your name in spite of everything you've done for me. Or misspelt it. I duly apologize, and I can only bring to my defence the late hour and the advanced form of keyboard-weariness. Probably it's for these reasons that acknowledgement sections of PhD theses are such a rich source of embarrassing mistakes (hey, there's a nice PhD project). Perhaps I can console you with the thought that, conceivably, some day I've forgotten every single person I've mentioned here, but not you, being haunted by your benevolence.

First than, I am grateful to my PhD supervisor and co-promotor, Gerard Grosveld, for allowing me to work in his lab and providing me with many good ideas. Professor Bootsma I thank for his trust and, especially these last months, for being a faithful link across the North Atlantic. Sjozèf van Baal, I have never been able to call you my technician, because you've always been your own. Nevertheless you have made major contributions to the work, throughout my entire PhD period. Importantly, you introduced or created many invaluable computer tools, without which several discoveries would not have been possible.

Rotterdam and Memphis, it's difficult to compare. I have many good memories to both places, mostly of late night hours. The old guard, Marieke von Lindern, the source, I'm typing this on your computer, thanks for having WP6.1 for Windows, Dies Meijer, I really tried to convert the lab to Dylan, but I'm afraid I only could convert myself, Martine Jaeglé, I saw you swaying through the corridors today and I was afraid you would take off, Hans Morreau, in other times you would have been a prophet; in yet other times you would be executed. Thanks my students, Monica Imi (I know you will be calling me on my birthday forever, and I never call you on yours) and Bas van Rees for enduring my emerging teaching skills. The VH-CML lab, it's gone forever, but it was a great lab. Thanks to the people who made it great: Nike Soekarman, Anneke Graus, Robert Kraaij, Ton de Wit, Ton Vertegaal, Marella de Bruin, Shen-Yan, Ingrid Ruigrok, Paula Mommers, Monique van der Knaap, Ronald Zwart, Antoinette Lambooy, Michael Dimitri Breems. And the ones that crossed the ocean.

Memphis, it really turned out to be the best of times, and, inevitably, sometimes the worst of times. Arjan "Dr. Tube" Buijs, I always will remember our persisting discussions on the question whether there had ever been rain before the Great Flood, thank you for being a great and upright colleague, and for and making those phenomenal noises. Jan "Dr. Deur" van Deursen, I've never seen someone quite like you, having so many right ideas at the right time. Keep in touch my friend! My other room 3054 comrades Lawryn Kasper, Jin-Xing

Jiang, Jacqueline Bonten, Tony McKinney, thanks for your friendship and keeping the lab alive, Judith Boer, my dear bread-partner, thank you for your knowledge, your true sense, and your paranymphing (the other paranymph is Martin de Vries, but I'll thank him some other time). Many people in the Department of Genetics have made the science a real joy (or at least, at times, bearable). Thanks, especially, Peggy Burdick, Aarnoud van der Spoel, Marie-Pili Martin, Robbert Rottier, Manuela Kapsetaki. Furthermore I want to thank Gopal Murti for stimulating discussions, Donna Davis, together we fought and won the porewar, Deanna Williams, my aerobics teacher, for keeping me in shape (read: sane), and Kent Williams for keeping up with me as a room mate / tenant despite my ailing household skills (my bike made it in one piece). Last but not least I would like to thank my parents for their support and great patience.

November 21, 1996, 1:45 AM. That's it.

Appendix

Appendix 1 Nucleotide sequence of Nup88 cDNA (accession no. Y08612). Aminoacid translation is given under first nucleotide of codon, the stopcodon is represented by an asteriks. The first methionione is in italics, and a possible polyadenylation signal is underlined. Sequences not present in the alternatively spliced Nup88 mRNA (Appendix 2) are in bold typeface.

```
   1  GATAAACCCACAAGACACAAAACATACCTTTCGAGCAGTTGGGCCAAGATGGCGGCCGCC   60
                                                  M  A  A  A        4

  61  GAGGGACCGGTGGGCGACGGCGAGCTGTGGCAGACCTGGCTTCCTAACCACGTCGTGTTC  120
   5  E  G  P  V  G  D  G  E  L  W  Q  T  W  L  P  N  H  V  V  F     24

 121  TTGCGGCTCCGGGAGGGACTGAAAAACCAGAGTCCAACCGAAGCTGAGAAACCAGCTTCT  180
  25  L  R  L  R  E  G  L  K  N  Q  S  P  T  E  A  E  K  P  A  S     44

 181  TCGTCGTTGCCTTCGTCGCCGCCGCCGCAGTTGCTGACGAGAAACGTGGTCTTTGGCCTC  240
  45  S  S  L  P  S  S  P  P  P  Q  L  L  T  R  N  V  V  F  G  L     64

 241  GGCGGAGAGCTTTTCCTGTGGGACGGAGAAGACAGCTCCTTCTTAGTCGTTCGCCTTCGG  300
  65  G  G  E  L  F  L  W  D  G  E  D  S  S  F  L  V  V  R  L  R     84

 301  GGCCCCAGCGGCGGCGGCGAAGAGCCCGCCCTGTCCCAGTACCAGAGATTGCTTTGCATA  360
  85  G  P  S  G  G  G  G  E  E  P  A  L  S  Q  Y  Q  R  L  L  C  I    104

 361  AATCCACCCCTGTTTGAAATCTATCAAGTCTTGTTAAGCCCAACACAACATCATGTAGCA  420
 105  N  P  P  L  F  E  I  Y  Q  V  L  L  S  P  T  Q  H  H  V  A    124

 421  CTTATAGGrATAAAAGGACTTATGGTATTAGAATTACCTAAAAGATGGGGGAAGAATTCT  480
 125  L  I  G  I  K  G  L  M  V  L  E  L  P  K  R  W  G  K  N  S    144

 481  GAATTTGAAGGTGGAAAATCAACAGTGAATTGTAGTACCACTCCAGTTGCGGAGAGATTT  540
 145  E  F  E  G  G  K  S  T  V  N  C  S  T  T  P  V  A  E  R  F    164

 541  TTCACCAGTTCCACCTCTCTGACTCTAAAGCATGCTGCATGGTATCCAAGTGAAATCCTG  600
 165  F  T  S  S  T  S  L  T  L  K  H  A  A  W  Y  P  S  E  I  L    184

 601  GATCCCCACGTAGTGCTGTTAACATCAGACAACGTAATCAGAATTTACTCTCTACGTGAG  660
 185  D  P  H  V  V  L  L  T  S  D  N  V  I  R  I  Y  S  L  R  E    204

 661  CCGCAGACACCCACTAACGTGATAATACTTTCAGAAGCCGAAGAGGAAAGTCTAGTACTC  720
 205  P  Q  T  P  T  N  V  I  I  L  S  E  A  E  E  E  S  L  V  L    224

 721  AATAAAGGAAGGGCGTATACCGCATCTCTAGGAGAGACAGCAGTTGCATTTGACTTTGGG  780
 225  N  K  G  R  A  Y  T  A  S  L  G  E  T  A  V  A  F  D  F  G    244

 781  CCATTGGACGCAGTCCCAAAGACTCTATTTGGACAAAACGGCAAAGATGAAGTAGTGGCA  840
 245  P  L  D  A  V  P  K  T  L  F  G  Q  N  G  K  D  E  V  V  A    264

 841  TACCCACTGTACATCTTATATGAAAATGGAGAGACTTTCCTGACATACATCAGTCTGTTA  900
 265  Y  P  L  Y  I  L  Y  E  N  G  E  T  F  L  T  Y  I  S  L  L    284

 901  CACAGCCCTGGAAATATTTGGAAAGCTGTTGGGTCCATTGCCCATGCATCTGCGGCTGAA  960
 285  H  S  P  G  N  I  W  K  A  V  G  S  I  A  H  A  S  A  A  E    304

 961  GATAACTATGGTTATGATGCGTGTGCTGTACTCTGCTTACCCTGTGTCCCCAATATCTTA 1020
 305  D  N  Y  G  Y  D  A  C  A  V  L  C  L  P  C  V  P  N  I  L    324

1021  GTGATCGCTACTGAATCAGGAATGCTGTATCACTGTGTCGTGCTAGAAGGGGAAGAAGAA 1080
 325  V  I  A  T  E  S  G  M  L  Y  H  C  V  V  L  E  G  E  E  E    344
```

```
1081  GATGACCACACGTCAGAAAAGTCCTGGGATTCCAGGATTGACCTCATTCCTTCTCTGTAT  1140
 345  D  D  H  T  S  E  K  S  W  D  S  R  I  D  L  I  P  S  L  Y    364

1141  GTGTTTGAATGTGTTGAGTTGGAGCTTGCTTTGAAACTGGCATCTGGAGAGGATGACCCT  1200
 365  V  F  E  C  V  E  L  E  L  A  L  K  L  A  S  G  E  D  D  P    384

1201  TTTGATTCTGACTTTTCTTGTCCAGTCAAACTTCATAGAGATCCCAAGTGTCCTTCAAGA  1260
 385  F  D  S  D  F  S  C  P  V  K  L  H  R  D  P  K  C  P  S  R    404

1261  TATCACTGTACTCATGAAGCTGGTGTACATAGTGTTGGGCTAACTTGGATTCATAAACTT  1320
 405  Y  H  C  T  H  E  A  G  V  H  S  V  G  L  T  W  I  H  K  L    424

1321  CACAAATTTCTTGGATCAGATGAAGAAGATAAGGATAGTTTACAGGAACTCTCTACAGAA  1380
 425  H  K  F  L  G  S  D  E  E  D  K  D  S  L  Q  E  L  S  T  E    444

1381  CAGAAATGCTTTGTTGAACACATCCTTTGTACGAGGCCATTGCCCTGCAGGCAGCCAGCT  1440
 445  Q  K  C  F  V  E  H  I  L  C  T  R  P  L  P  C  R  Q  P  A    464

1441  CCAATTCGAGGATTTTGGATTGTACCTGACATTCTGGGACCCACGATGATCTGCATCACC  1500
 465  P  I  R  G  F  W  I  V  P  D  I  L  G  P  T  M  I  C  I  T    484

1501  AGTACCTATGAATGCCTCATATGGCCGTTATTAAGTACAGTCCATCCAGCGTCTCCTCCC  1560
 485  S  T  Y  E  C  L  I  W  P  L  L  S  T  V  H  P  A  S  P  P    504

1561  CTGCTTTGTACTCGAGAAGATGTTGAAGTGGCAGAGTCTTCCCTCCGTGTTCTGGCTGAA  1620
 505  L  L  C  T  R  E  D  V  E  V  A  E  S  S  L  R  V  L  A  E    524

1621  ACCCCAGATTCCTTTGAAAAGCATATTAGAAGCATTTTGCAACGTAGTGTTGCCAATCCA  1680
 525  T  P  D  S  F  E  K  H  I  R  S  I  L  Q  R  S  V  A  N  P    544

1681  GCATTTTTGAAAGCTTCTGAAAAGGACATAGCCCCTCCTCCTGAAGAATGCCTTCAGCTC  1740
 545  A  F  L  K  A  S  E  K  D  I  A  P  P  P  E  E  C  L  Q  L    564

1741  CTCAGCAGAGCCACCCAGGTGTTCAGAGAGCAGTACATTCTCAAACAGGACTTGGCAAAG  1800
 565  L  S  R  A  T  Q  V  F  R  E  Q  Y  I  L  K  Q  D  L  A  K    584

1801  GAGGAGATTCAGCGGAGGGTCAAATTATTATGTGACCAAAAAAAGAAACAACTAGAAGAT  1860
 585  E  E  I  Q  R  R  V  K  L  L  C  D  Q  K  K  K  Q  L  E  D    604

1861  CTCAGTTATTGTCGAGAAGAGAGGAAAAGTCTGCGGGAAATGGCTGAGCGTTTAGCTGAC  1920
 605  L  S  Y  C  R  E  E  R  K  S  L  R  E  M  A  E  R  L  A  D    624

1921  AAATATGAGGAAGCTAAAGAAAAACAAGAGGATATCATGAACAGGATGAAAAAACTACTT  1980
 625  K  Y  E  E  A  K  E  K  Q  E  D  I  M  N  R  M  K  K  L  L    644

1981  CACAGTTTTCACTCTGAGCTCCCAGTTCTCTCTGATAGTGAGCGAGACATGAAGAAAGAA  2040
 645  H  S  F  H  S  E  L  P  V  L  S  D  S  E  R  D  M  K  K  E    664

2041  TTACAGCTGATACCTGATCAACTTCGACATTTGGGCAATGCCATCAAACAGGTTACTATG  2100
 665  L  Q  L  I  P  D  Q  L  R  H  L  G  N  A  I  K  Q  V  T  M    684

2101  AAAAAGGATTATCAACAGCAAAAGATGGAGAAGGTGTTGAGTCTTCCAAAACCCACCATT  2160
 685  K  K  D  Y  Q  Q  Q  K  M  E  K  V  L  S  L  P  K  P  T  I    704

2161  ATTCTCAGTGCCTACCAGCGAAAGTGCATTCAGTCCATCCTGAAAGAGGAGGGTGAACAT  2220
 705  I  L  S  A  Y  Q  R  K  C  I  Q  S  I  L  K  E  E  G  E  H    724

2221  ATAAGGGAAATGGTGAAGCAAATCAATGATATCCGCAATCATGTAAACTTCTGACACCAC  2280
 725  I  R  E  M  V  K  Q  I  N  D  I  R  N  H  V  N  F  *         741

2281  CAGGAGCTGACTCACACCTGAACTGAACACCATTGAAGGCTTAAACCCATATTGTAAAAC  2340

2341  AGGTAGAATTATCTAATTTATAAAAAGGTGTTTTGATGAAAAAAAAAAAA            2390
```

Appendix 2 Nucleotide sequence of an alternative Nup88 3' untranslated region (accession no. Y08613). Sequences shared by the major Nup88 mRNA are in bold, its stopcodon is in italics. A possible polyadenylation signal is in underlined. Sequence tagged site (STS) WI-6584 is represented by nucleotides 959-1190.

```
   1  ***TGA***CACCACCAGGGAGCTGACTCACACCTGAACTGAACACCATTGAAGGCTTAAACCCAT   60
  61  **ATTGTAAAACAG**CATCAACAGCAGACCGGCAATTGCGCTTTTGATCTAAAGGAAAGATGA   120
 121  AAGCCTGGATCCTGCAATACACTGAGATTATAGAAAACTGCAAAGCAAACCAAATCCAGG      180
 181  TTGCTCTAACTTAAATTCCGCAATGCTAGTCAGGCAGATTGCTCCAACCTGAGAGCTGTG      240
 241  AGTTCAGGTGGGGAGGTTCACATTCACTGTGATTAACTAGCTAGCTCCTTTCTCTGTCAC      300
 301  AATTACCTGGAATACCCAAGGCAGACTATGACAAGGGTCAGGTGACACACGTTAATATCC      360
 361  ATTGCTGGCAACATGTGCCTGCACTGACTCCAGGAACCGTCTGTTGAATGAACGCTAAAT      420
 421  CCCAGAGCTTCCCAGTGTGCCACAGCGTGAGACGAGGAAAAACGTGAGAAAGGTAACTCA      480
 481  CCTTCTTTCTGCAAGAGATTCCATTCTGTTTCACTAACTAAACGATAACCACTAGGAGAC      540
 541  ACATAGGCACTCTCCATGCCCTGGGTAACGCTGGAGAGGAGGGACGCTGTCTCTTCTTGT      600
 601  TCTGGTGTCATCGCCTTGATTGCTCTTTCCTGATCTTTGGTTAACATAAACCCACTTGGC      660
 661  ATCTGGAGTGACCCAAGGGATGCAGTATCAAAGTTCTCAGACACCTGAATCTGCTCCTAC      720
 721  CAGTGGTCCAAAATCTGACTCGTCCAGGTTTCCAGCAGATTTTGCTTTGTAGTTATAGCC      780
 781  TAAAGCTTTGGATTGCAATGAGCCCGAGGTTCCCCAAGCTGTCTGTAGACTGTGCTCTCC      840
 841  CTGAGTCCCATCTTTAAACATGTCATTGTCCGATTTACTGAAAGGATCTCCAGATGGCAA      900
 901  CAGCAAGCCTGCATCTAAGGAATGAACTGAGCCCACGGGTGCTGTCTAAATGCTCCTGAA      960
 961  ATGTGAAAATACATATGGCCAAGTTAGACAATTCTCTGCGTTTTTAAAGACAAAAACTAT     1020
1021  AATAAATGTATTCCGCCACTCATTACACATGATAATCATTAATGTGCTACCTTCACAGAC     1080
1081  TCATTGAACAAACACTAAGTGTCATTTCTACATGCTGGACACAAAACCAGGCCCTCAGGA     1140
1141  TACAACTGTGAGCAAGATATAATCCCTGACCTCGAGATACATACGGCTTCAAAAAAAAAA     1200
```

Appendix 3 Nucleotide sequence of hCRM1 cDNA (accession no. Y08614). Aminoacid translation is given under first nucleotide of codon, the stopcodon is represented by an asteriks. First and second methioniones are indicated with italics, and a possible polyadenylation signal is underlined.

```
   1  AGGAAGGAAGGAGCAGTTGGTTCAATCTCTGGTAATCTATGCCAGCAATTATGACAATGT   60
                                              M  P  A  I  M  T  M  L    8

  61  TAGCAGACCATGCAGCTCGTCAGCTGCTTGATTTCAGCCAAAAACTGGATATCAACTTAT  120
   9    A  D  H  A  A  R  Q  L  L  D  F  S  Q  K  L  D  I  N  L  L   28

 121  TAGATAATGTGGTGAATTGCTTATACCATGGAGAAGGAGCCCAGCAAAGAATGGCTCAAG  180
  29    D  N  V  V  N  C  L  Y  H  G  E  G  A  Q  Q  R  M  A  Q  E   48

 181  AAGTACTGACACATTTAAAGGAGCATCCTGATGCTTGGACAAGAGTCGACACAATTTTGG  240
  49    V  L  T  H  L  K  E  H  P  D  A  W  T  R  V  D  T  I  L  E   68

 241  AATTTTCTCAGAATATGAATACGAAATACTATGGACTACAAATTTTGGAAAATGTGATAA  300
  69    F  S  Q  N  M  N  T  K  Y  Y  G  L  Q  I  L  E  N  V  I  K   88

 301  AAACAAGGTGGAAGATTCTTCCAAGGAACCAGTGCGAAGGAATAAAAAAATACGTTGTTG  360
  89    T  R  W  K  I  L  P  R  N  Q  C  E  G  I  K  K  Y  V  V  G  108

 361  GCCTCATTATCAAGACGTCATCTGACCCAACTTGTGTAGAGAAAGAAAAGGTGTATATCG  420
 109    L  I  I  K  T  S  S  D  P  T  C  V  E  K  E  K  V  Y  I  G  128

 421  GAAAATTAAATATGATCCTTGTTCAGATACTGAAACAAGAATGGCCCAAACATTGGCCAA  480
 129    K  L  N  M  I  L  V  Q  I  L  K  Q  E  W  P  K  H  W  P  T  148

 481  CTTTTATCAGTGATATTGTTGGAGCAAGTAGGACCAGCGAAAGTCTCTGTCAAAATAATA  540
 149    F  I  S  D  I  V  G  A  S  R  T  S  E  S  L  C  Q  N  N  M  168

 541  TGGTGATTCTTAAACTCTTGAGTGAAGAAGTATTTGATTTCTCTAGTGGACAGATAACCC  600
 169    V  I  L  K  L  L  S  E  E  V  F  D  F  S  S  G  Q  I  T  Q  188

 601  AAGTCAAATCTAAGCATTTAAAAGACAGCATGTGCAATGAATTCTCACAGATATTTCAAC  660
 189    V  K  S  K  H  L  K  D  S  M  C  N  E  F  S  Q  I  F  Q  L  208

 661  TGTGTCAGTTTGTAATGGAAAATTCTCAAAATGCTCCACTTGTACATGCAACCTTGGAAA  720
 209    C  Q  F  V  M  E  N  S  Q  N  A  P  L  V  H  A  T  L  E  T  228

 721  CATTGCTCAGATTTCTGAACTGGATTCCCCTGGGATATATTTTTGAGACCAAATTAATCA  780
 229    L  L  R  F  L  N  W  I  P  L  G  Y  I  F  E  T  K  L  I  S  248

 781  GCACATTGATTTATAAGTTCCTGAATGTTCCAATGTTTCGAAATGTCTCTCTGAAGTGCC  840
 249    T  L  I  Y  K  F  L  N  V  P  M  F  R  N  V  S  L  K  C  L  268

 841  TCACTGAGATTGCTGGTGTGAGTGTAAGCCAATATGAAGAACAATTTGTAACACTATTTA  900
 269    T  E  I  A  G  V  S  V  S  Q  Y  E  E  Q  F  V  T  L  F  T  288

 901  CTCTGACAATGATGCAACTAAAGCAGATGCTTCCTTTAAATACCAATATTCGACTTGCGT  960
 289    L  T  M  M  Q  L  K  Q  M  L  P  L  N  T  N  I  R  L  A  Y  308

 961  ACTCAAATGGAAAAGATGATGAACAGAACTTCATTCAAAATCTCAGTTTGTTTCTCTGCA 1020
 309    S  N  G  K  D  D  E  Q  N  F  I  Q  N  L  S  L  F  L  C  T  328

1021  CCTTTCTTAAGGAACATGATCAACTTATAGAAAAAAGATTAAATCTCAGGGAAACTCTTA 1080
 329    F  L  K  E  H  D  Q  L  I  E  K  R  L  N  L  R  E  T  L  M  348

1081  TGGAGGCCCTTCATTATATGTTGTTGGTATCTGAAGTAGAAGAAACTGAAATCTTTAAAA 1140
 349    E  A  L  H  Y  M  L  L  V  S  E  V  E  E  T  E  I  F  K  I  368

1141  TTTGTCTTGAATACTGGAATCATTTGGCTGCTGAACTCTATAGAGAGAGTCCATTCTCTA 1200
 369    C  L  E  Y  W  N  H  L  A  A  E  L  Y  R  E  S  P  F  S  T  388
```

```
1201  CATCTGCCTCTCCGTTGCTTTCTGGAAGTCAACATTTTGATGTTCCTCCCAGGAGACAGC  1260
 389    S  A  S  P  L  L  S  G  S  Q  H  F  D  V  P  P  R  R  Q  L   408

1261  TATATTTGCCCATGTTATTCAAGGTCCGTTTATTAATGGTTAGTCGAATGGCTAAACCAG  1320
 409    Y  L  P  M  L  F  K  V  R  L  L  M  V  S  R  M  A  K  P  E   428

1321  AGGAAGTATTGGTTGTAGAGAATGATCAAGGAGAAGTTGTGAGAGAATTCATGAAGGATA  1380
 429    E  V  L  V  V  E  N  D  Q  G  E  V  V  R  E  F  M  K  D  T   448

1381  CAGATTCCATAAATTTGTATAAGAATATGAGGGAAACATTGGTTTATCTTACTCATCTGG  1440
 449    D  S  I  N  L  Y  K  N  M  R  E  T  L  V  Y  L  T  H  L  D   468

1441  ATTATGTAGATACAGAAAGAATAATGACAGAGAAGCTTCACAATCAAGTGAATGGTACAG  1500
 469    Y  V  D  T  E  R  I  M  T  E  K  L  H  N  Q  V  N  G  T  E   488

1501  AGTGGTCATGGAAAAATTTGAATACATTGTGTTGGGCAATAGGCTCCATTAGTGGAGCAA  1560
 489    W  S  W  K  N  L  N  T  L  C  W  A  I  G  S  I  S  G  A  M   508

1561  TGCATGAAGAGGACGAAAAACGATTTCTTGTTACTGTTATAAAGGATCTATTAGGATTAT  1620
 509    H  E  E  D  E  K  R  F  L  V  T  V  I  K  D  L  L  G  L  C   528

1621  GTGAACAGAAAAGAGGCAAAGATAATAAAGCTATTATTGCATCAAATATCATGTACATAG  1680
 529    E  Q  K  R  G  K  D  N  K  A  I  I  A  S  N  I  M  Y  I  V   548

1681  TAGGTCAATACCCACGTTTTTTGAGAGCTCACTGGAAATTTCTGAAGACTGTAGTTAACA  1740
 549    G  Q  Y  P  R  F  L  R  A  H  W  K  F  L  K  T  V  V  N  K   568

1741  AGCTGTTCGAATTCATGCATGAGACCCATGATGGAGTCCAGGATATGGCTTGTGATACTT  1800
 569    L  F  E  F  M  H  E  T  H  D  G  V  Q  D  M  A  C  D  T  F   588

1801  TCATTAAAATAGCCCAAAAATGCCGCAGGCATTTCGTTCAGGTTCAGGTTGGAGAAGTGA  1860
 589    I  K  I  A  Q  K  C  R  R  H  F  V  Q  V  Q  V  G  E  V  M   608

1861  TGCCATTTATTGATGAAATTTTGAACAACATTAACACTATTATTTGTGATCTTCAGCCTC  1920
 609    P  F  I  D  E  I  L  N  N  I  N  T  I  I  C  D  L  Q  P  Q   628

1921  AACAGGTTCATACGTTTTATGAAGCTGTGGGGTACATGATTGGTGCACAAACAGATCAAA  1980
 629    Q  V  H  T  F  Y  E  A  V  G  Y  M  I  G  A  Q  T  D  Q  T   648

1981  CAGTACAAGAGCACTTGATAGAAAAGTACATGTTACTCCCTAATCAAGTGTGGGATAGTA  2040
 649    V  Q  E  H  L  I  E  K  Y  M  L  L  P  N  Q  V  W  D  S  I   668

2041  TAATCCAGCAGGCAACCAAAAATGTGGATATACTGAAAGATCCTGAAACAGTCAAGCAGC  2100
 669    I  Q  Q  A  T  K  N  V  D  I  L  K  D  P  E  T  V  K  Q  L   688

2101  TTGGTAGCATTTTGAAAACAAATGTGAGAGCCTGCAAAGCTGTTGGACACCCCTTTGTAA  2160
 689    G  S  I  L  K  T  N  V  R  A  C  K  A  V  G  H  P  F  V  I   708

2161  TTCAGCTTGGAAGAATTTATTTAGATATGCTTAATGTATACAAGTGCCTCAGTGAAAATA  2220
 709    Q  L  G  R  I  Y  L  D  M  L  N  V  Y  K  C  L  S  E  N  I   728

2221  TTTCTGCAGCTATCCAAGCTAATGGTGAAATGGTTACAAAGCAACCATTGATTAGAAGTA  2280
 729    S  A  A  I  Q  A  N  G  E  M  V  T  K  Q  P  L  I  R  S  M   748

2281  TGCGAACTGTAAAAAGGGAAACTTTAAAGTTAATATCTGGTTGGGTGAGCCGATCCAATG  2340
 749    R  T  V  K  R  E  T  L  K  L  I  S  G  W  V  S  R  S  N  D   768

2341  ATCCACAGATGGTCGCTGAAAATTTTGTTCCCCCTCTGTTGGATGCAGTTCTCATTGATT  2400
 769    P  Q  M  V  A  E  N  F  V  P  P  L  L  D  A  V  L  I  D  Y   788

2401  ATCAGAGAAATGTCTCAGCTGCTAGAGAACCAGAAGTGCTTAGTACTATGGCCATAATTG  2460
 789    Q  R  N  V  S  A  A  R  E  P  E  V  L  S  T  M  A  I  I  V   808

2461  TCAACAAGTTAGGGGGACATATAACAGCTGAAATACCTCAAATATTTGATGCTGTTTTTG  2520
 809    N  K  L  G  G  H  I  T  A  E  I  P  Q  I  F  D  A  V  F  E   828

2521  AATGCACATTGAATATGATAAATAAGGACTTTGAAGAATATCCTGAACATAGAACGAACT  2580
 829    C  T  L  N  M  I  N  K  D  F  E  E  Y  P  E  H  R  T  N  F   848

2581  TTTTCTTACTACTTCAGGCTGTCAATTCTCATTGTTTCCCAGCATTCCTTGCTATTCCAC  2640
 849    F  L  L  L  Q  A  V  N  S  H  C  F  P  A  F  L  A  I  P  P   868
```

```
2641  CTACACAGTTTAAACTTGTTTTGGATTCCATCATTTGGGCTTTCAAACATACTATGAGGA  2700
 869    T  Q  F  K  L  V  L  D  S  I  I  W  A  F  K  H  T  M  R  N   888

2701  ATGTCGCAGATACGGGCTTACAGATACTTTTTACACTCTTACAAAATGTTGCACAAGAAG  2760
 889    V  A  D  T  G  L  Q  I  L  F  T  L  L  Q  N  V  A  Q  E  E   908

2761  AAGCTGCAGCTCAGAGTTTTTATCAAACTTATTTTTGTGATATTCTCCAGCATATCTTTT  2820
 909    A  A  A  Q  S  F  Y  Q  T  Y  F  C  D  I  L  Q  H  I  F  S   928

2821  CTGTTGTGACAGACACTTCACATACTGCTGGTTTAACAATGCATGCATCAATTCTTGCAT  2880
 929    V  V  T  D  T  S  H  T  A  G  L  T  M  H  A  S  I  L  A  Y   948

2881  ATATGTTTAATTTGGTTGAAGAAGGAAAAATAAGTACATCATTAAATCCTGGAAATCCAG  2940
 949    M  F  N  L  V  E  E  G  K  I  S  T  S  L  N  P  G  N  P  V   968

2941  TTAACAACCAAATCTTTCTTCAGGAATATGTGGCTAATCTCCTTAAGTCGGCCTTCCCTC  3000
 969    N  N  Q  I  F  L  Q  E  Y  V  A  N  L  L  K  S  A  F  P  H   988

3001  ACCTACAAGATGCTCAAGTAAAGCTCTTTGTGACAGGGCTTTTCAGCTTAAATCAAGATA  3060
 989    L  Q  D  A  Q  V  K  L  F  V  T  G  L  F  S  L  N  Q  D  I  1008

3061  TTCCTGCTTTCAAGGAACATTTAAGAGATTTCCTAGTTCAAATAAAGGAATTTGCAGGTG  3120
1009    P  A  F  K  E  H  L  R  D  F  L  V  Q  I  K  E  F  A  G  E  1028

3121  AAGACACTTCTGATTTGTTTTTGGAAGAGAGAGAAATAGCCCTACGGCAGGCTGATGAAG  3180
1029    D  T  S  D  L  F  L  E  E  R  E  I  A  L  R  Q  A  D  E  E  1048

3181  AGAAACATAAACGTCAAATGTCTGTCCCTGGCATCTTTAATCCACATGAGATTCCAGAAG  3240
1049    K  H  K  R  Q  M  S  V  P  G  I  F  N  P  H  E  I  P  E  E  1068

3241  AAATGTGTGATTAAAATCCAAATTCATGCTGTTTTTTTTCTCTGCAACTCCGTTAGCAGA  3300
1069    M  C  D  *                                                   1071

3301  GGAAAACAGCATGTGGGTATTTGTCGACCAAAATGATGCCAATTTGTAAATTAAAATGTC  3360

3361  ACCTAGTGGCCCTTTTTCTTATGTGTTTTTTTGTATAAGAAATTTTCTGTGAAATATCCT  3420

3421  TCCATTGTTTAAGCTTTTGTTTTGGTCATCTTTATTTAGTTTGCATGAAGTTGAAAATTA  3480

3481  AGGCATTTTTAAAAATTTTACTTCATGCCCATTTTTGTGGCTGGGCTGGGGGGAGGAGGC  3540

3541  AAATTCAATTTGAACATATACTTGTAATTCTAATGCAAAATTATACAATTTTTCCTGTAA  3600

3601  ACAATACCAATTTTTAATTAGGGAGCATTTTCCTTCTAGTCTATTTCAGCCTAGAAGAAA  3660

3661  AGATAATGAGTAAAACAAATTGCGTTGTTTAAAGGATTATAGTGCTGCATTGTCTGAAGT  3720

3721  TAGCACCTCTTGGACTGAATCGTTTGTCTAGACTACATGTATTACAAAGTCTCTTTGGCA  3780

3781  AGATTGCAGCAAGATCATGTGCATATCATCCCATTGTAAAGCGACTTCAAAAATATGGGA  3840

3841  ACACAGTTAGTTATTTTTACACAGTTCTTTTTGTTTTTGTGTGTGTGTGCTGTCGCTTGT  3900

3901  CGACAACAGCTTTTTGTTTTCCTCAATGAGGAGTGTTGCTCATTTGTGAGCCTTCATTAA  3960

3961  CTCGAAGTGAAATGGTTAAAAATATTTATCCTGTTAGAATAGGCTGCATCTTTTTAACAA  4020

4021  CTCATTAAAAAACAAAACAACTCTGGCTTTTGAGATGACTTATACTAATTTACATTGTTT  4080

4081  ACCAAGCTGTAGTGCTTTAAGAACACTACTTAAAAAGCAAAATAAACTTGGTTTACATTT  4140

4141  AAAAAAAA  4148
```