A short peptide at the C terminus is responsible for the nuclear localization of RAG2

ARTICLE *in* EUROPEAN JOURNAL OF IMMUNOLOGY · AUGUST 2002

Impact Factor: 4.03 · DOI: 10.1002/1521-4141(200207)32:7<2068::AID-IMMU2068>3.0.CO;2-6 · Source: PubMed

A short peptide at the C terminus is responsible for the nuclear localization of RAG2

Barbara Corneo[1], Alexandre Benmerah[2] and Jean-Pierre de Villartay[1]

[1] Développement Normal et Pathologique du Système Immunitaire, INSERM U429, Hôpital Necker-Enfants Malades, Paris, France
[2] INSERM E9925, Faculté Necker-Enfants Malades, Paris, France

The RAG1 and RAG2 proteins are the lymphoid-specific factors essential for V(D)J recombination, the process that leads to the diversification of antigen receptors on B and T lymphocytes. Nucleolar/nuclear localization of RAG1 is mediated by four basic domains, which are the binding sites for the nuclear transport proteins SRP1 and RCH1, and by a nuclear localization signal (NLS) in the fifth basic domain. The C-terminal region of RAG2 from amino acids (aa) 417 to 484 shows a homology with the PHD domain of other proteins involved in chromatin-mediated gene regulation by protein-protein interactions. Mutations in this domain were shown to be responsible for several diseases and in some case lead to altered subcellular localization of proteins. We found that the C-terminal PHD domain of RAG2 is not responsible for the nuclear localization of the protein. We report here the characterization of a region (aa 491–527) in the C-terminal domain of RAG2, downstream of the putative PHD domain, which directs the nuclear localization of the protein.

Received	15/3/02
Revised	29/4/02
Accepted	16/5/02

Key words: V(D)J Recombination / RAG2 / Nuclear localization signal

1 Introduction

The V(D)J recombination (reviewed in [1]) allows the correct rearrangement of the variable (V), diversity (D) and joining (J) segments constituting the B cell receptor (BCR) and the T cell receptor (TCR). This process begins with DNA cleavage [2] at the recombination signal sequences (RSS) [3], adjacent to the coding segment, by the lymphoid-specific proteins RAG1 [4] and RAG2 [5]. This generates to a double-strand break at the RSS/coding segment border with generation of hairpin-sealed coding ends and blunt, phosphorylated signal ends. Subsequently, the ubiquitously expressed general DNA repair machinery of the cells, including the recently described Artemis factor [6], rejoins the two chromosomal broken ends leading to coding joint (reviewed in [7]). Deletional analysis has defined the minimal regions of RAG1 and RAG2 required for their catalytic function on V(D)J extrachromosomal substrates. Amino acids (aa) 384–1008 of RAG1 [8] and aa 1–387 of RAG2 [9, 10] represent the active core region of the two proteins required for V(D)J recombination. However, both the N-terminal region of RAG1 and the C-terminal region of RAG2, outside of the core domains, are evolutionary conserved and present important functions *in vivo*. The N-terminal RAG1 region enhances the V(D)J activity of the protein [11, 12]. This region contains three stretches of basic aa: BI (aa 141–146), BII (aa 218–236) and BIII (aa 243–249) that are important for the interaction of RAG1 with the SRP1 protein [13]. Two other basic regions, BIV (aa 826–840) and BV (aa 969–973), are localized in the C-terminal domain of RAG1. BIV is the region involved in RAG1 interaction with RCH1 [14]. SRP1 and RCH1 are nuclear transport proteins that allow the nuclear translocation of proteins containing a NLS [15, 16]. BV contains the dominant NLS in RAG1 (putative NLS [4]: R(140–141)KKEKR; dominant NLS [17]: R(969)RFRK). Endogenous RAG1 and RAG2 proteins localize in the periphery of the nucleus in primary mouse thymocytes as well as in NIH-3T3 fibroblast cell lines expressing low levels of RAG1 and RAG2 from a retroviral vector [17]. When RAG1 and RAG2 proteins are expressed at high levels in the same cells, RAG1 appears to localize in the nucleolus while RAG2 remains in the nucleus [17]. Hydrophobic cluster analysis (HCA) [18] suggested that RAG2 is probably constituted of two distinct globular domains separated by a hinge region of 60 aa. The N-terminal globular domain corresponds to the core region, required for V(D)J activity. This region was described as a six-bladed β-propeller [18], involved in protein-protein or protein-DNA interactions. Each blade consists of four

[I 23017]

Abbreviations: NLS: Nuclear localization signal **PHD:** Plant homeodomain **aa:** Amino acid **GFP:** Green fluorescence protein

0014-2980/02/0707-2068$17.50+.50/0

β-strands separated by loops of various length, extending outside the structure and therefore exposed to solvent. The C-terminal region of RAG2 from aa 417 to 484 shows a homology with the zinc-finger domain, constituted by seven cysteines and a histidine C4HC3, of other proteins [18]. This homology region, called PHD (plant homeodomain), TTC (trithorax consensus) or LAP (leukemia-associated protein), was identified in more than 400 eukaryotic proteins involved in regulation of gene expression [19]. Recently, the first solution structure of the PHD domain from the KAP1 corepressor was reported [20], showing that structural integrity of the PHD domain is essential for KAP1 to repress transcription. As the PHD domain appears to function as a protein-protein interaction motif, disruption of the structural integrity by mutations of critical residues fundamentally disrupt this action. Point mutations or deletions in the PHD domain of several proteins are in fact responsible for the pathogenesis of different human diseases, such as myeloid leukemia and autoimmune diseases. The AIRE gene encodes for a nuclear protein constituted by two PHD domains and is mutated in APECED (autoimmune polyendocrinopathy candidiasis ectodermal dystrophy) [21]. Mutations in the coding sequence of AIRE lead to truncated forms of the protein, lacking at least one of the two PHD domains. These mutants show an altered nuclear localization [22]. We investigated if the PHD-homologous domain in the C-terminal moiety of RAG2 was implicated in the nuclear localization of the protein. We demonstrate here that while the PHD domain is not involved, a region constituted of the last 37 aa of the protein (aa 491–527) is responsible for the nuclear localization of RAG2. This region alone directs the nuclear translocation of the normally cytoplasmic protein Eps15-DII [23, 24].

2 Results and discussion

2.1 The RAG1-GFP protein is localized in the nucleolus and the RAG2-GFP protein in the nucleus

To examine their intracellular localization, RAG1 and RAG2 coding sequences were fused to EGFP, an expression plasmid containing the coding sequence of the green fluorescence protein (GFP) gene. When transfected in 293T cells, EGFP-C2 alone localized in both the cell nucleus and the cytoplasm (Fig. 1A). In contrast, the RAG1-GFP fusion protein localized in the cell nucleolus (Fig. 1B) and RAG2-GFP was found in the nucleus but spared the nucleolus (Fig. 1C), as previously reported using specific antisera against murine RAG1 and RAG2 proteins [17, 25–28]. Identical results were obtained in HeLa cells (data not shown).

2.2 The β-propeller domain of RAG2 is localized in the cytoplasm, the PHD domain in the nucleus

The RAG2 protein can be schematically divided into four distinct regions (Fig. 2A): the 355 aa β-propeller separated from the 76 aa PHD domain by a hinge region of 53 aa. Lastly, 43 aa at the C terminus of the protein follow the PHD domain. To better define the sequence requirement for directing RAG2 to the nucleus, we analyzed the subcellular localization of various parts of the RAG2 protein fused to EGFP. The N-terminal β-propeller domain associated to the hinge region was localized in the cytoplasm (Fig. 2B). In contrast, the RAG2 protein devoid of the β-propeller was found in the nucleus but spared the nucleolus (Fig. 2C). This suggested that the

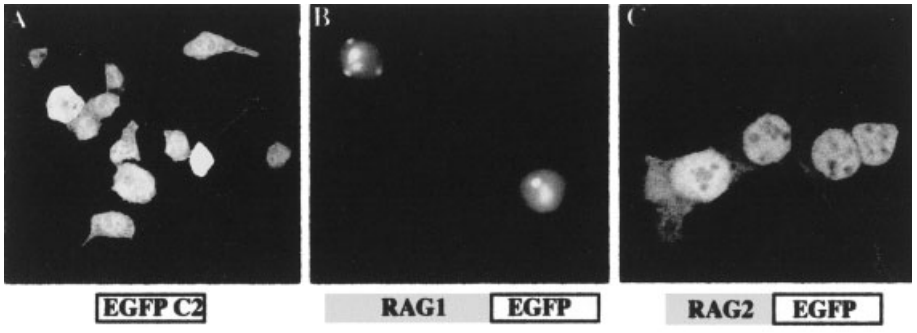

Fig. 1. Subcellular localization of EGFP-tagged RAG1 and RAG2 proteins transfected in 293T cells. RAG1 and RAG2 coding sequences were PCR amplified and subcloned into EGFP-C2. Each construct was transfected in cover-slip adherent 293T cells by FuGene6. 48 h after transfection, samples were washed and examined under an epifluorescence or confocal microscope. (A) EGFP-C2 localized in both cell nucleus and cytoplasm. (B) RAG1-GFP localized the nucleolus. (C) RAG2-GFP localized in nucleus but spared the nucleolus.

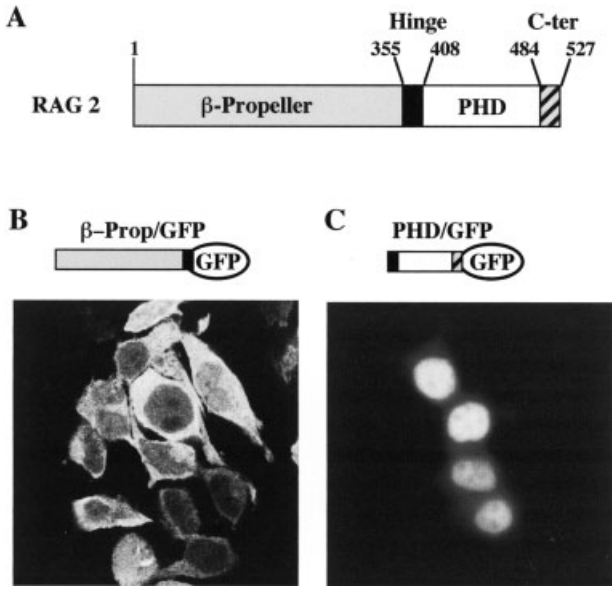

Fig. 2. Intracellular localization of various regions of the RAG2 protein fused to EGFP and transfected in 293T cells. (A) Schematic representation of the various domain of RAG2. (B) The N-terminal β-propeller localized to the nucleus, (C) the C terminus PHD domain localized to the nucleus.

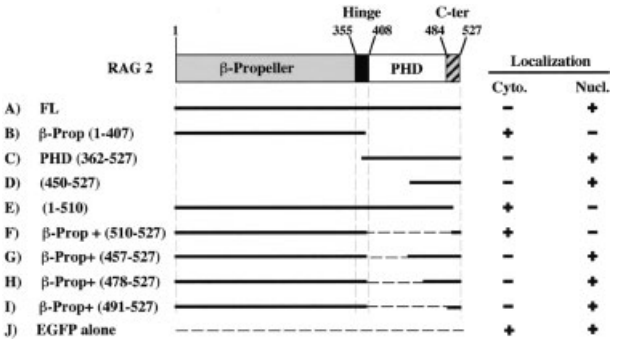

Fig. 3. Subcellular localization of proteins generated by serial deletion of RAG2 fused to EGFP. (A) Full-length RAG2, (B) β–propeller (aa 1–407), (C) PHD domain (aa 362–527). (D) truncated PHD domain (aa 450–527), (E) aa 1–510, (F) β-Propeller + aa 510–527, (G) β-Propeller + aa 457–527, (H) β-Propeller + aa 478–527. (I) β-Propeller + aa 491–527, (J) EGFP alone.

172 aa C-terminal region of RAG2, which contains the PHD domain, was responsible for the proper nuclear localization of the full-length RAG2.

2.3 The PHD domain of RAG2 is not involved in the nuclear localization of RAG2

Given the role of the PHD domain in the nuclear localization of other proteins, a deletional analysis of the C-terminal region of RAG2 was performed to precisely define the protein sequence that mediates the nuclear localization (Fig. 3). When a truncated PHD domain, lacking four out of the seven conserved cysteins that constitutes the PHD motif, was fused to GFP, the resulting protein still localized to the nucleus (Fig. 3D), strongly suggesting that the PHD domain itself was not responsible for the nuclear localization of RAG2. In contrast, a RAG2 protein lacking the very last 17 aa, but containing an intact β-propeller, hinge and PHD domain, had lost its nuclear localization and accumulated in the cytoplasm (Fig. 3E). This indicated that the last 17 aa were required for proper nuclear targeting of RAG2. However, these 17 aa were not sufficient on their own to drive nuclear localization, as a chimeric protein constituted of the β-propeller fused to the last 17 aa was still retained in the cytoplasm (Fig. 3F). This was not the result of an active retention of the β-propeller domain in the cytoplasm, as a classical monopartite NLS (SV40 NLS) [29] fused to β-propeller-EGFP was able to target the fusion protein into the nucleus (data not shown). Lastly, we evaluated different portions of the C terminus of RAG2, starting from aa 457, for their ability to redirect the cytoplasmic β-propeller-GFP protein to the nucleus (Fig. 3G–I). The fusion proteins produced by all three constructs were localized in the nucleus. Altogether these data indicate that the 37-aa minimal region of RAG2 necessary to direct its localization to the nucleus comprises aa 491–527. This region can be regarded as the NLS of RAG2. Although classical NLS motifs could not be found within this region using specific computer programs as PSORT [30], the high proportion of basic residues (33.4%) could, by itself, account for this effect.

2.4 The NLS in the C-terminal region of RAG2 can mediate nuclear translocation of a heterologous protein

To evaluate if the last 37 aa of RAG2 could act as a NLS independently of the β-propeller region we fused them to a heterologous protein, Eps15. Eps15 is a constitutive component of cathrin-coated pits [23, 24]. We have previously generated and characterized various Eps15-derived GFP constructs [31]. The one corresponding to the central domain of Eps15 (DII) was used as a cytosolic reporter molecule because of its main cytoplasmic staining (Fig. 4A). The last 37 aa of RAG2 were sufficient to target the Eps15-DII-GFP protein to the nucleus (Fig. 4B). This clearly demonstrates that this region of 37 aa in the C-terminal domain of RAG2 behaves as an NLS which can mediate the nuclear translocation of an

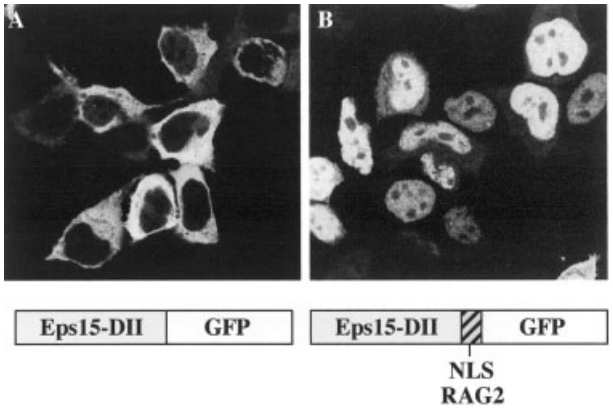

Fig. 4. Nuclear translocation of the heterologous cytoplasmic protein Eps15-DII by the RAG2-NLS. The last 37 aa of RAG2 (aa 491–527) were fused to the cytoplasmic protein Eps15-DII-GFP. (A) Eps15-DII-GFP localized in cytoplasm. (B) Eps15-DII+RAG2(NLS)-GFP localized to nucleus.

otherwise cytoplasmic protein. It also suggests that this NLS is autonomous and does not need other sequences in RAG2 to drive its nuclear localization.

2.5 Nature of the RAG2 NLS

NLS have been classified in several groups according to their aa composition and usually contain highly basic residues (reviewed in [32]). In some proteins such as SV40 Large T, a single peptide, which comprises a series of basic residues (PKKKRKV) is responsible for nuclear import. In some other proteins, a bipartite NLS consists of two stretches of basic residues separated by 10–12 aa. The last 17 residues of RAG2 contain an interesting motif (PAKKSFLRRLFD) that is part of the NLS as judged by the requirement of this peptide for nuclear import. The structure of this motif highly resembles the c-myc NLS (PAAKRDVKLKLD) where the proline and aspartic acid residues flanking the basic cluster play important role in nuclear import. In RAG2 this motif, although necessary, is clearly not sufficient for nuclear targeting. The 24 residues upstream of this motif, which are part of the 37-aa minimal region for nuclear import of RAG2, contain 8 R/K residues (in particular the RKK stretch) that probably play an important role in nuclear targeting. *In vitro* mutagenesis of each of these residues should help clarify their precise role. Altogether the RAG2 NLS more or less concords with conventional NLS sequences which suggests that RAG2 is probably imported in the nucleus through the action of members of the importin family.

3 Concluding remarks

The C-terminal domain of RAG2 including the PHD domain has been considered dispensable for the V(D)J recombination activity *in vitro*, but proved to have important functions *in vivo*. Indeed, this region is required for the proper rearrangement of endogenous immunoglobulins genes as opposed to extrachromosomal substrates [33]. Moreover, a conserved degradation signal involved in the cell cycle regulation of RAG2 is constituted by Thr490 and by multiple cationic residues within the interval from the aa 499–508 [34]. We described here another important function carried by the sequence of RAG2 outside of the core catalytic domain: the region spanning aa 491–527 in the C-terminal domain contains an NLS. The expression of truncated forms of RAG2 through a retroviral transfer, where the transgene is expressed at relatively low level, should permit to address the role of this NLS *in vivo* on the rearrangement of endogenous Ig or TCR genes.

4 Materials and methods

4.1 Plasmid constructions

Standard PCR and digestion techniques were used to generate the different RAG1 and RAG2 constructs by cloning into EGFP-N2 or EGFP-C2 (Clontech). Details on Eps15-DII-GFP construction were previously described [31]. All plasmids were sequenced to confirm that each construct sequence was in frame with the GFP sequence and that no mutations were introduced during the PCR amplification.

4.2 Cell culture and tranfections

293T and HeLa cells were grown in DMEM supplemented with 10% fetal bovine serum (Gibco). After cell adhesion on gelatin-treated cover-slips in six-well plates, 5 μg of each plasmid constructs were transfected into cells by FuGene6 (Roche). At 48 h following transfection, cells were gently washed in PBS and samples were fixed 15 min in 3.7% PFA. After three washes in PBS, cover-slips were mounted in Mowiol (Hoechst).

4.3 Immunofluorescence

Samples were examined under an epifluorescence microscope (Axioplan 2, Carl Zeiss, Jena, Germany) attached to a cooled CCD Spot2 camera (Diagnostic Instruments, Burroughs, MI) or to a confocal microscope (Carl Zeiss).

Acknowledgements: We are grateful to Jerôme Feldman for his technical advice. This work was supported by institu-

tional grants from Institut National de la Santé et de la Recherche Médicale and Ministère de l'Education Nationale de la Recherche et de la Technologie, and grants from Association de Recherche sur le Cancer (ARC) et Commissariat à l'énergie atomique (CEA-LRC 7V). B.C. is supported by scholarship from Ligue contre le Cancer and Fondation pour la Recherche Médicale (FRM).

References

1 **Fugmann, S. D., Lee, A. I., Shockett, P. E., Villey, I. J. and Schatz, D. G.,** The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* 2000. **18:** 495–527.

2 **McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M. and Oettinger, M. A.,** Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* 1995. **83:** 387–395.

3 **Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S.,** Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* 1979. **280:** 288–294.

4 **Schatz, D. G., Oettinger, M. A. and Baltimore, D.,** The V(D)J recombination activating gene, RAG-1. *Cell* 1989. **59:** 1035–1048.

5 **Oettinger, M. A., Schatz, D. G., Gorka, C. and Baltimore, D.,** RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 1990. **248:** 1517–1523.

6 **Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A. and de Villartay, J. P.,** ARTEMIS, a novel DNA double-strand break repair / V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 2001. **105:** 177–186.

7 **Grawunder, U., West, R. B. and Lieber, M. R.,** Antigen receptor gene rearrangement. *Curr. Opin. Immunol.* 1998. **10**: 172–180.

8 **Kirch, S. A., Sudarsanam, P. and Oettinger, M. A.,** Regions of RAG1 protein critical for V(D)J recombination. *Eur. J. Immunol.* 1996. **26:** 886–891.

9 **Sadofsky, M. J., Hesse, J. E. and Gellert, M.,** Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucleic Acids Res.* 1994. **22:** 1805–1809.

10 **Cuomo, C. A. and Oettinger, M. A.,** Analysis of regions of RAG-2 important for V(D)J recombination. *Nucleic Acids Res.* 1994. **22:** 1810–1814.

11 **McMahan, C. J., Difilippantonio, M. J., Rao, N., Spanopoulou, E. and Schatz, D. G.,** A basic motif in the N-terminal region of RAG1 enhances V(D)J recombination activity. *Mol. Cell. Biol.* 1997. **17:** 4544–4552.

12 **Roman, C. A., Cherry, S. R. and Baltimore, D.,** Complementation of V(D)J recombination deficiency in RAG-1(-/-) B cells reveals a requirement for novel elements in the N-terminus of RAG-1. *Immunity* 1997. **7:** 13–24.

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

14 **Cuomo, C. A., Kirch, S. A., Gyuris, J., Brent, R. and Oettinger, M. A.,** Rch1, a protein that specifically interacts with the RAG-1 recombination-activating protein. *Proc. Natl. Acad. Sci. USA* 1994. **91:** 6156–6160.

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

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

17 **Spanopoulou, E., Cortes, P., Shih, C., Huang, C. M., Silver, D. P., Svec, P. and Baltimore, D.,** Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. *Immunity* 1995. **3:** 715–726.

18 **Callebaut, I. and Mornon, J. P.,** The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis. *Cell. Mol. Life Sci.* 1998. **54:** 880–891.

19 **Jacobson, S. and Pillus, L.,** Modifying chromatin and concepts of cancer. *Curr. Opin. Genet. Dev.* 1999. **9:** 175–184.

20 **Capili, A. D., Schultz, D. C., Rauscher, I. F. and Borden, K. L.,** Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *EMBO J.* 2001. **20:** 165–177.

21 **Bjorses, P., Pelto-Huikko, M., Kaukonen, J., Aaltonen, J., Peltonen, L. and Ulmanen, I.,** Localization of the APECED protein in distinct nuclear structures. *Hum. Mol. Genet.* 1999. **8:** 259–266.

22 **Rinderle, C., Christensen, H. M., Schweiger, S., Lehrach, H. and Yaspo, M. L.,** AIRE encodes a nuclear protein co-localizing with cytoskeletal filaments: altered suB cellular distribution of mutants lacking the PHD zinc fingers. *Hum. Mol. Genet.* 1999. **8:** 277–290.

23 **Fazioli, F., Minichiello, L., Matoskova, B., Wong, W. T. and Di Fiore, P. P.,** Eps15, a novel tyrosine kinase substrate, exhibits transforming activity. *Mol. Cell. Biol.* 1993. **13:** 5814–5828.

24 **Tebar, F., Sorkina, T., Sorkin, A., Ericsson, M. and Kirchhausen, T.,** Eps15 is a component of clathrin-coated pits and vesicles and is located at the rim of coated pits. *J. Biol. Chem.* 1996. **271:** 28727–28730.

25 **Leu, T. M. and Schatz, D. G.,** RAG-1 and RAG-2 are components of a high-molecular-weight complex, and association of rag-2 with this complex is RAG-1 dependent. *Mol. Cell. Biol.* 1995. **15:** 5657–5670.

26 **Ferguson, S. E., Accavitti, M. A., Wang, D. D., Chen, C. L. and Thompson, C. B.,** Regulation of RAG-2 protein expression in avian thymocytes. *Mol. Cell. Biol.* 1994. **14:** 7298–7305.

27 **Silver, D. P., Spanopoulou, E., Mulligan, R. C. and Baltimore, D.,** Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. *Proc. Natl. Acad. Sci. USA* 1993. **90:** 6100–6104.

28 **McMahan, C. J., Sadofsky, M. J. and Schatz, D. G.,** Definition of a large region of RAG1 that is important for coimmunoprecipitation of RAG2. *J. Immunol.* 1997. **158:** 2202–2210.

29 **Mattaj, I. W. and Englmeier, L.,** Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* 1998. **67**: 265–306.

30 **Nakai, K. and Horton, P.,** PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* 1999. **24:** 34–36.

31 **Benmerah, A., Poupon, V., Cerf-Bensussan, N. and Dautry-Varsat, A.,** Mapping of Eps15 domains involved in its targeting to clathrin-coated pits. *J. Biol. Chem.* 2000. **275:** 3288–3295.

32 **Jans, D. A., Xiao, C. and Lam, M. H. C.,** Nuclear targeting signal recognition: a key control point in nuclear transport. *Bioessays* 2000. **22:** 532–544.

33 **Kirch, S. A., Rathbun, G. A. and Oettinger, M. A.,** Dual role of RAG2 in V(D)J recombination: catalysis and regulation of ordered Ig gene assembly. *EMBO J.* 1998. **17**: 4881–4886.

34 **Li, Z., Dordai, D. I., Lee, J. and Desiderio, S.,** A conserved degradation signal regulates RAG-2 accumulation during cell division and links V(D)J recombination to the cell cycle. *Immunity* 1996. **5:** 575–589.

Correspondence: Jean-Pierre de Villartay, INSERM U429, Hôpital Necker-Enfants Malades, 149, rue de Sèvres, F-75015 Paris, France
Fax: +33-01-4273-0640
e-mail: devillar@infobiogen.fr