

Evolutionary specialization of the nuclear targeting apparatus

(nuclear transport/ α and β karyopherins/Armadillo and HEAT motifs)

HARMIT S. MALIK, THOMAS H. EICKBUSH, AND DAVID S. GOLDFARB*

Department of Biology, University of Rochester, Rochester, NY 14627

Communicated by Masayasu Nomura, University of California, Irvine, CA, October 2, 1997 (received for review August 1, 1997)

ABSTRACT The α - and β -karyopherins (Kaps), also called importins, mediate the nuclear transport of proteins. All α -Kaps contain a central domain composed of eight approximately 40 amino acid, tandemly arranged, armadillo-like (Arm) repeats. The number and order of these repeats have not changed since the common origin of fungi, plants, and mammals. Phylogenetic analysis suggests that the various α -Kaps fall into two groups, $\alpha 1$ and $\alpha 2$. Whereas animals encode both types, the yeast genome encodes only an $\alpha 1$ -Kap. The β -Kaps are characterized by 14–15 tandemly arranged HEAT motifs. We show that the Arm repeats of α -Kaps and the HEAT motifs of β -Kaps are similar, suggesting that the α -Kaps and β -Kaps (and for that matter, all Arm and HEAT repeat-containing proteins) are members of the same protein superfamily. Phylogenetic analysis indicates that there are at least three major groups of β -Kaps, consistent with their proposed cargo specificities. We present a model in which an α -independent β -Kap progenitor gave rise to the α -dependent β -Kaps and the α -Kaps.

The nuclear pore complex is remarkable for its capacity to catalyze the bi-directional translocation of diverse macromolecules (1–3). In contrast, the α and β karyopherins (here designated α/β Kaps) families of nuclear targeting factors [also called α/β importin, pore-targeting apparatus, nuclear localization signal (NLS) receptor-P97 complex, or Kap60/95] appear to have become specialized to mediate the transport of selected kinds of NLS cargo (4). The prototypic β -Kap95 factors of yeast and humans function as heterodimers with α -Kap60, an NLS receptor, to mediate the docking of NLS cargo at the cytoplasmic face of the nuclear pore complex. α -Kap60 mediates the import of a wide range of karyophilic proteins that contain simian virus 40 T-antigen-like or nucleoplasmin-like NLS sequences (1–3).

In contrast to α -Kap dependent β -Kap95, yeast β -Kap104 (5), and its human homologue transportin (here called human Kap104) (6–8) are α -Kap independent transport factors that exhibit both the NLS-binding and nucleoporin binding activities needed to import an abundant group of M9 signal-containing heterogeneous nuclear ribonucleoproteins. Recent results suggest that yeast β -Kap121 and β -Kap123 (9) and the human β -Kap121 homologue (10) are also α -Kap independent transport factors that mediate the import of ribosomal proteins. β -Kap95, -Kap104, -Kap121, and -Kap123 also have been termed $\beta 1$ –4, respectively (8, 10). The *in vivo* substrate ranges and binding properties of the different NLS receptors for their cognate NLSs remain inadequately defined. For example, it is still unclear if the different α -Kap and β -Kap NLS receptors have discrete or overlapping NLS-cargo specificities.

Although little is known about the secondary structure of the β -Kaps, an earlier report has identified these proteins as

containing HEAT motifs (11) (the HEAT acronym is based on proteins in which these repeats were originally identified, ref. 11). These motifs are 38–45 amino acids in length and have been proposed to characterize proteins involved in transport processes. HEAT motifs appear in both eukaryotic and prokaryotic proteins (11).

The complexity of the extant α/β Kap nuclear targeting apparatus raises interesting evolutionary questions about the origin and function of these genetically diverse gene families. This apparatus must have evolved in the earliest nucleated eukaryotes given their need to compartmentalize proteins to either the nucleus or cytoplasm. Here, we have investigated the evolutionary relationships among these factors by using molecular phylogenetic techniques. This analysis suggests that the various α - and β -Kaps have evolved from a common ancestor. Three major specialization events have occurred during the evolution of these eukaryotic gene families. First, the α -Kaps arose together with the α -dependent β -Kap from an α -independent β -Kap progenitor. Second, the α -Kaps divided into two groups, the $\alpha 1$ -Kaps and $\alpha 2$ -Kaps. Finally, gene duplication and functional specialization has occurred within the $\alpha 2$ -Kap family.

RESULTS AND DISCUSSION

Phylogeny of the α -Kaps. Phylogenetic analysis was first used to establish the relationship between different α -Kaps. α -Kaps are ≈ 60 -kDa polypeptides previously characterized as being composed of eight tandem 38–45 residue Armadillo (Arm) repeats (identified in *Drosophila* armadillo protein) flanked by ≈ 100 -aa N- and C-terminal domains (12, 13). Alignment of all available α -Kap-like sequences revealed extensive similarity within the Arm repeats, N-, and C-terminus domains (Fig. 1).

Because the α -Kaps contain internal tandem repeats, the alignment in Fig. 1 cannot be used in constructing a phylogeny unless it can be unambiguously determined that the order of individual repeats within the domain has been conserved. Otherwise, nonhomologous repeats would be aligned, leading to incorrectly inferred phylogenies. A phylogenetic analysis of each of the eight separate Arm repeats from three highly divergent α -Kaps is shown in Fig. 2 and indicates that the order of the eight Arm repeats has been conserved from yeast to humans. For example, Arm repeat no. 1 in the yeast α -Kap is more similar to repeat no. 1 in the two human α -Kaps than to any other Arm repeat in the yeast and human proteins. As shown in Fig. 2, a similar conclusion can be made for each of the eight Arm repeats. We conclude that the order of the Arm repeats has been conserved in all known α -Kaps. This result predicts that the number and order of the Arm repeats of the extant α -Kaps is similar to that of the progenitor α -Kap. The precise conservation of Arm repeat order further suggests that the eight repeats are not functionally interchangeable. This

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.

© 1997 by The National Academy of Sciences 0027-8424/97/9413738-5\$2.00/0
PNAS is available online at <http://www.pnas.org>.

Abbreviations: Kap, karyopherin; Arm, armadillo; NLS, nuclear localization signal.

*To whom reprint requests should be addressed. e-mail: Goldfarb@Nucleus.Biology.Rochester.edu.

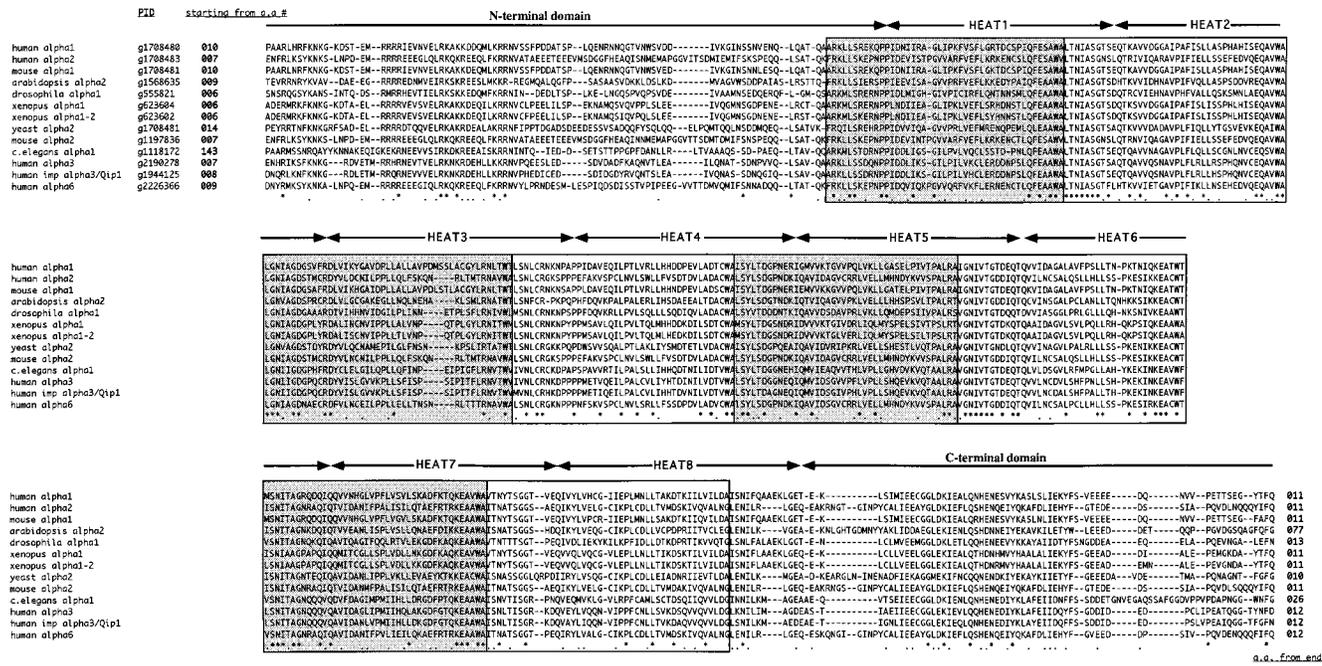


FIG. 1. α -Kap alignment. Complete ORFs from the various α -Kaps were aligned by using the CLUSTALV package of programs (27) and high gap penalties. * and • below the alignment indicate identities and similarities, respectively among all α -Kaps. The boxes indicate the most recent definition of the Arm repeats (12, 13), and the horizontal arrows indicate the HEAT motifs (protein identification) numbers of each sequence are indicated at the beginning of the alignment. Note that the *Caenorhabditis elegans* α -Kap, a composite of several expressed sequence tags, has a considerably longer N terminus, which may be an artifact.

hypothesis may be tested by constructing recombinant α -Kaps with shuffled repeat units.

The Arm repeat domains of all available α -Kap sequences were used to construct a phylogeny by the Neighbor-Joining method (Fig. 3). Based on the phylogeny shown in Fig. 3, we propose that the members of the group containing the yeast *SRP1* encoded α -Kap be referred to as “ $\alpha 1$ -Kaps” (or $\alpha 1$ -importins), and that the members of the other groups be called “ $\alpha 2$ -Kaps” (or $\alpha 2$ -importins). These two groups are likely to

reflect a functional divergence of α -Kaps, because each group is broadly distributed in different phyla. Note, however, that within the $\alpha 2$ clade shown in Fig. 3 there are three distinct sub-clades, each as different from the one another as they are from the $\alpha 1$ group. We have designated these three clades as subdivisions of a single $\alpha 2$ group. This designation was done in recognition of the fact that yeast contains only an $\alpha 1$ -Kap (see below). Thus, the $\alpha 2A$, $\alpha 2B$, and $\alpha 2C$ Kaps all are likely to have arisen from an archetypal $\alpha 1$ -Kap. Alternatively, we could have just as easily designated the three $\alpha 2$ clades as $\alpha 2$, $\alpha 3$, and

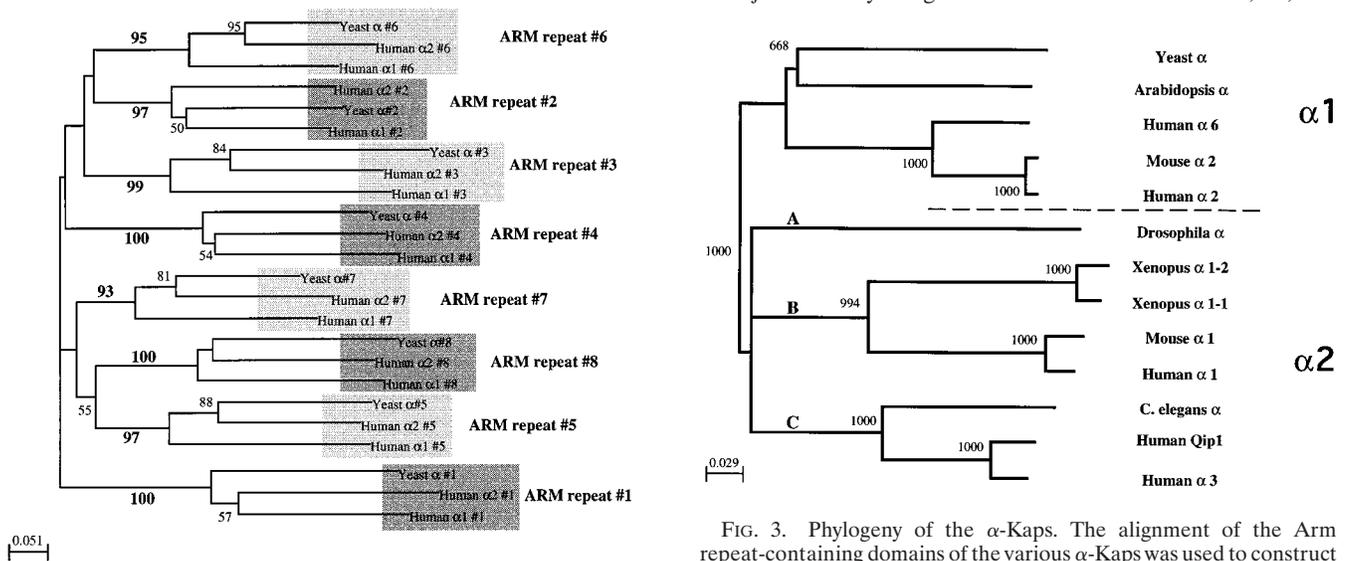


FIG. 2. Conservation of the order of Arm repeats within the α -Kaps. The eight individual Arm repeats from yeast $\alpha 1$ -Kap, human $\alpha 1$ -Kap, and human $\alpha 2$ -B-Kap were aligned, and a phylogenetic relationship was derived by using the Neighbor-Joining method (28). The tree shown here is a 50% consensus tree with bootstraps represented as a percentage of 1,000 replications.

FIG. 3. Phylogeny of the α -Kaps. The alignment of the Arm repeat-containing domains of the various α -Kaps was used to construct a phylogeny by the Neighbor-Joining method (28). Shown here is the consensus tree of 1,000 bootstrap replications. The tree was midpoint-rooted between the $\alpha 1$ and $\alpha 2$ -Kaps. The phylogenetic relationship of the proteins based on the Maximum Parsimony algorithm, by using the PAUP package of programs (30), also was determined and gave the identical topology. Nomenclatures based both on the submitted names as well as a proposed scheme are indicated.

α 4. Previous number designations of α -Kaps have been made somewhat arbitrarily, for example, according to the order of their submission to the database (see Fig. 1). The previous nomenclature incorrectly implies that the human and mouse α 2-Kaps are more closely related to the yeast α 1-Kap than are the human and mouse α 1-Kaps. The approach we have taken to assigning α -Kap nomenclature is more consistent with the derived α -Kap phylogeny shown in Fig. 3.

The α -Kap phylogeny is notable for the existence of only a single α 1-Kap and no α 2-Kap gene in *Saccharomyces cerevisiae* (*SRP1*) as well as multiple α 1- and α 2-Kaps in humans. Because of the availability of the entire yeast genome sequence, it is unlikely that another α -Kap exists in yeast. This fact indicates that the yeast α 1-Kap must be capable of performing all of the transport duties required of an α -Kap in this single-extensive organism. The α 2-Kap family has undergone more extensive specialization than the α -Kaps. Three highly divergent clades of the α 2-Kaps, designated α 2A, α 2B, and α 2C, may serve the distinct developmental, tissue-type specific, and NLS cargo-specific functions of more complex organisms. This hypothesis is supported by the observations that the expression of *Drosophila* α 2A-Kap (Pendulin; refs. 14 and 15) and mouse α 2B-Kap (16) are developmentally regulated and that α 1 and α 2B-Kaps in humans differ in their *in vitro* NLS binding properties (17). Even within the three clades of the α 2-Kaps there has been significant sub-specialization. Two similar α 2B-Kaps identified in *Xenopus* (18), and the

human α 3/Qip1 (α 2C-Kap clade) are examples of this specialization. Future sequencing of α -Kap genes from divergent species should help further resolve the various α -Kap groups and provide insight into when these groups appeared in the eukaryotic lineage. For example, are α 2-Kaps restricted to metazoans? Do they occur in vascular plants? [To date, only α 1-Kap expressed sequence tags have been reported in plants (not shown).] Do mammalian equivalents of the *Drosophila* α 2A-Kap (Pendulin) exist?

Repeat Structure and Phylogeny of the β -Kaps. β -Kaps in yeast fall into four major classes—the α -dependent β -Kap95p, the heterogeneous nuclear ribonucleoprotein-transporting β -Kap104p, and the ribosomal protein transporting factors β -Kap121p and β -Kap123p. As mentioned earlier, β -Kaps are characterized by internal HEAT motifs. At the amino acid sequence level, HEAT motifs are significantly more degenerate than Arm repeats, but each tandem repeat appears to represent a conserved secondary structure motif consisting of antiparallel α -helices with short connecting loops (11, 19). In the β -Kaps, specific HEAT repeats have been proposed to bear Ran-binding determinants (see ref. 20). We have attempted to construct a comprehensive map of the HEAT and non-HEAT-containing domains of these β -Kaps. As shown in Fig. 4, the β -Kaps are composed predominately of 14–15 tandem HEAT repeats, flanked by 100–150 amino acid arms (the KAP95s are missing the C-terminal arms). The HEAT repeats are in tandem except for the interruption of 1–3 non-HEAT se-

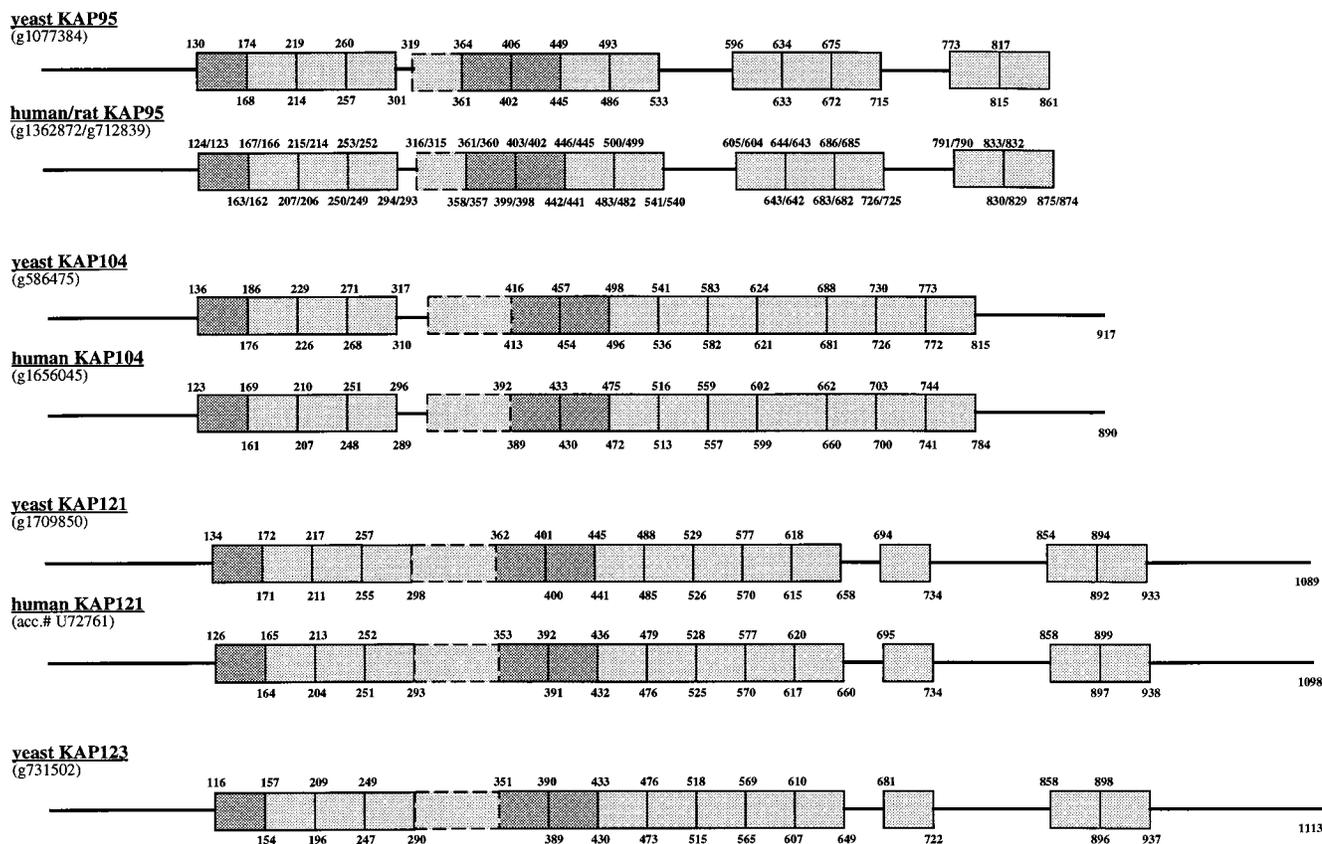


FIG. 4. Distribution of HEAT motifs in the β -Kaps. HEAT motifs are indicated as boxes, and non-HEAT-containing sequences are represented by thick lines. The entire human and yeast ORFs were aligned by CLUSTALV (27) in pairs. A window of about 40 amino acids, matching the HEAT consensus (11), then was used to match the ORF pairs with a high gap penalty, to score for individual HEAT motifs. The alignment windows were evaluated individually on the basis of matches to the residues identified previously as being important for the secondary structure. Mismatches of one important residue per α -helix were allowed as long as the secondary structure prediction was unaffected. Secondary structure predictions were used based on the SSPRED method (29) to confirm the primary sequence alignments. The darkly shaded boxes represent HEAT motifs that are unambiguously homologous in the various β -Kaps (data not shown). The dashed boxes represent HEAT motifs that are not close matches to the universal HEAT consensus (Fig. 3) but are predicted to assume a similar secondary structure (11, 19, 20). Each HEAT repeat is bounded by numbers above and below the box that indicate the beginning and ending amino acid for that particular repeat. Accession numbers/protein identification numbers are indicated below the protein names.

quence or “partial” HEAT sequence-containing regions. Most of the length variation of the different β -Kaps is derived from non-HEAT motif expansion domains in the larger β -Kaps.

Because of the high sequence degeneracy among HEAT motifs, it is difficult to match homologous HEAT repeats among the β -Kaps. However, we were able to unambiguously identify three sets of homologous HEAT repeats among the four β -Kap groups—the first, sixth, and seventh (Fig. 4, dark boxes). An approximately 79-residue sequence (Fig. 5) containing the sixth and seventh HEAT repeats, which may contain part of a Ran-binding determinant (20), was used to construct a phylogeny of the four β -Kap groups (Fig. 6). An alignment of the entire protein sequences based on their similar domain structures (Fig. 4) yielded a similar tree (not shown). Not included in this analysis (because they were missing the sixth and seventh HEAT repeats) are a number of expressed sequence tags from vertebrates and invertebrates that fall in the Kap95, Kap104, or Kap121 lineages. No expressed sequence tags homologous to yeast Kap123p were identified. Thus, *KAP123*, which is nonessential for growth, may be unique to yeast or is expressed at low levels in metazoans. The tree in Fig. 6 suggests that the Kap95p, Kap104p, and Kap121/123p groups are very old and about equally divergent, consistent with their proposed functions in defining eukaryotic nuclear transport pathways (1–10). Although it is not possible to root this tree, the most parsimonious explanation to account for two of three major branches being α -Kap independent is that the Kap95p ancestor arose from an α -Kap independent progenitor.

Alignment of Arm and HEAT Repeats. If, as we postulate, α -dependence is a derived character, then it is likely that the α -Kaps would have appeared in tandem with the α -dependent β -Kaps. This simultaneous appearance could have occurred by a genetic event that separated the NLS-binding domain from the nuclear pore complex docking domain of an α -independent β -Kap-like ancestor into separate genes. Evidence that the ancestral α -KAP gene arose from a β -KAP progenitor is that the α -Kap Arm and β -Kap HEAT repeats can be aligned by shifting by 10 residues the currently accepted definition of the tandem Arm repeat junctions (refs. 12 and 13; Fig. 7). Indeed, the definition of the central domain of α -Kaps as containing HEAT motifs is better because the relatively poor first Arm repeat becomes an excellent match to a HEAT repeat (Figs. 1 and 7). Second, the proposed secondary structure of the HEAT repeat-containing proteins suggests they are in a repeating helical packing structure (19). If, as we suggest, this structure is present in α -Kaps and other Arm repeat-containing proteins (19, 21), the currently defined (12, 13) Arm repeat junction interrupts one of the two helical secondary structures found in every repeat (Figs. 1 and 7). The original definition (22) of the Arm repeat junction is in accordance with the HEAT junction, and correctly positions the border of the sequence outside of the two helical regions. It should be noted that the phylogenetic analysis of the α -Kaps that used HEAT junctions yields the same tree topologies as

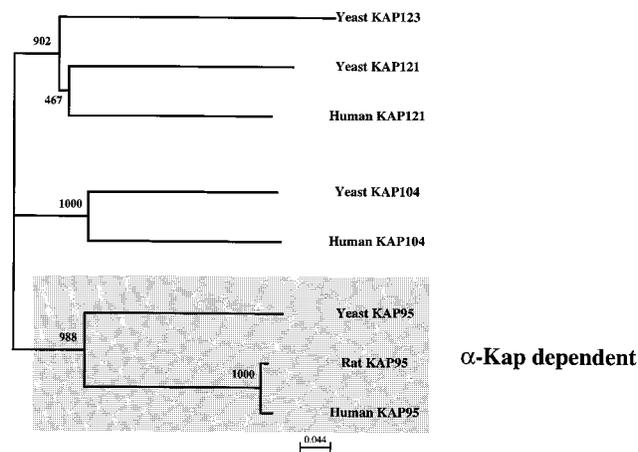


FIG. 6. Phylogeny of the β -Kaps. Neighbor-Joining trees (28) based on the alignment in Fig. 5, with bootstrap values of 1,000 replications. The phylogenetic relationship of the proteins based on Maximum Parsimony (30) differs in that the human Kap121 falls outside a branch containing both the yeast Kap121 and yeast Kap123.

shown in Figs. 2 and 3, which are based on the Arm junctions (data not shown).

The finding that Arm and HEAT repeats define the same two-helical secondary structure suggests that Arm and HEAT repeat-containing proteins belong to the same protein superfamily and may share analogous functions. Both domains appear to function as scaffolds to bind various protein ligands. For example, the Arm repeats of β -catenin are required for binding to α -catenin, cadherin, and adenomatous polyposis coli (23). The HEAT motifs of the A subunit of PP2A associate with tumor antigens and with the B and C subunits of PP2A (19). The same is likely to be true for the Arm and HEAT motifs of α -Kaps and β -Kaps, respectively, which form a heterodimer in the targeting of NLS-bearing substrates to the nuclear pore complex.

Model for the Evolution of the Kaps. Based on these results we propose the following model for the evolution of β and α/β Kap-mediated nuclear transport. Both β and α/β systems likely evolved from an ancient α -Kap-independent β -Kap progenitor. The β -KAP95 gene family arose at about the time of the separation between the β -KAP104 and β -KAP121/123 families. Recently, additional genes with partial homology to the β -Kaps have been reported (24, 25), suggesting further specialization within this family. We propose that the ancestral α -KAP gene, evolved in tandem with the β -KAP95 progenitor. Because yeast has a single α 1-Kap and the α 2-Kaps appear to be more adapted to specialized functions, we propose the first α -Kap was probably of the α 1 type. A second major phase of evolution within the Kaps, the appearance and sub-specialization of the α 2-Kaps, occurred after the separation of the yeast and human lineages to serve the specialized targeting and nuclear functions of early metazoan cell types, associated with development (26). For example, the loss of the *Drosophila*

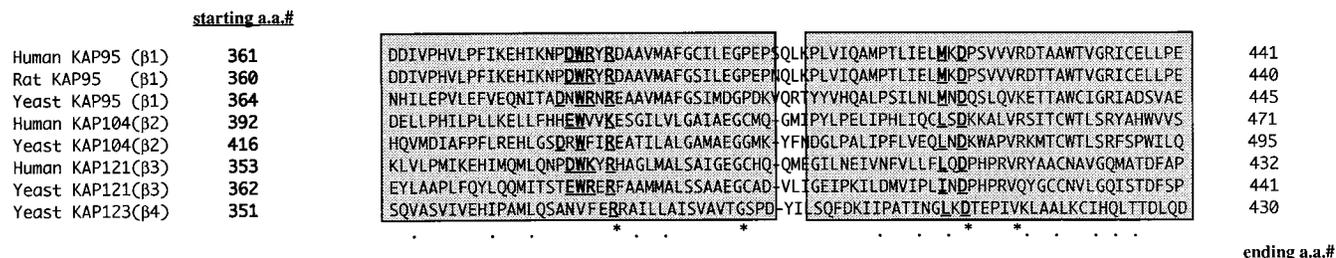


FIG. 5. Alignment of conserved sixth and seventh HEAT motifs of β -Kap sequences. * and • at the bottom of the alignment represent identities and similarities among the β -Kap sequences, respectively. Putative Ran-binding determinants (20) are highlighted in bold.

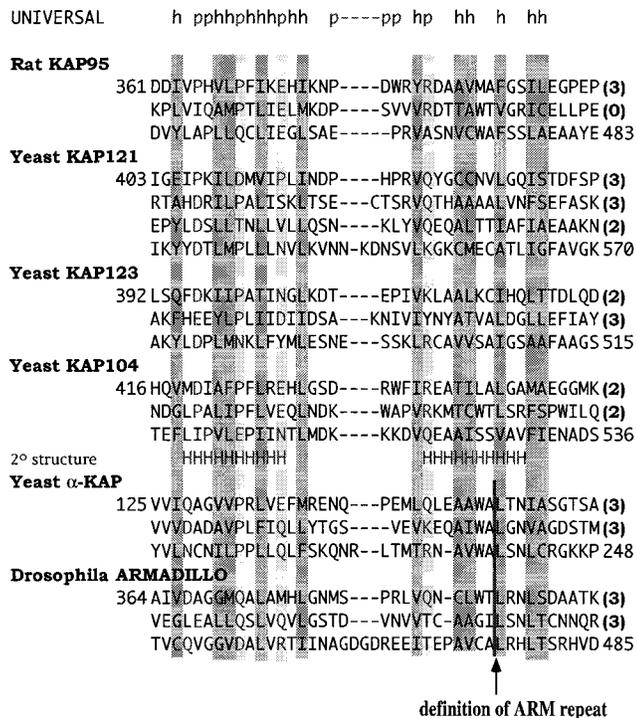


FIG. 7. Alignment of Arm and HEAT motifs. HEAT motifs from the various β -Kaps and Arm repeat-containing segments from Yeast α 2-Kap and *Drosophila* Armadillo (β -catenin) were aligned to the universal HEAT consensus (11). p and h indicate polar and hydrophobic residues, and HHH indicates α -helical secondary structure prediction. The arrow indicates the Arm repeat junction identified in Fig. 1 (12). Numbers at the beginning and end of each amino acid stretch represent the number of amino acids to the end of the ORF. Numbers in parentheses represent the number of amino acids between the consecutive HEAT repeats.

α 2-Kap Pendulin is an embryonic lethal mutation that causes defects in the control of cell proliferation and malignancy of hematopoietic organs (14, 15).

Note Added in Proof. Recently, the structure of the Arm-repeat domain of murine β -catenin was solved (31). The junctions between the 12 secondary structure motifs were found to be inconsistent with the Arm repeat definition. These authors proposed an alternative repeat junction that is in agreement with the HEAT motif definitions that we suggest applies to the α -Kaps and β -catenins (Fig. 7).

This work was supported by National Institutes of Health Public Health Service Grants GM40362 (D.S.G.) and GM42790 (T.H.E.).

- Gorlich, D. S. & Mattaj, I. W. (1996) *Science* **271**, 1513–1518.
- Nigg, E. A. (1997) *Nature (London)* **386**, 779–787.
- Corbett, A. H. & Silver, P. A. (1997) *Microbiol. Mol. Biol. Rev.* **61**, 193–211.
- Goldfarb, D. S. (1997) *Curr. Biol.* **7**, 14–17.
- Aitchison, J. D., Blobel, G. & Rout, M. P. (1996) *Science* **274**, 624–627.
- Pollard, V. W., Michael, W. M., Nakielny, S., Siomi, M. C., Wang, F. & Dreyfuss, G. (1996) *Cell* **86**, 985–994.
- Nakielny, S., Siomi, M. C., Siomi, H., Michael, W. M., Pollard, V. & Dreyfuss, G. (1996) *Exp. Cell Res.* **229**, 261–266.
- Bonifaci, N., Moroianu, J., Radu, A. & Blobel, G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5055–5060.
- Rout, M. P., Blobel, G. & Aitchison, J. D. (1997) *Cell* **89**, 715–725.
- Yaseen, N. R. & Blobel, G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 4451–4456.
- Andrade, M. A. & Bork, P. (1995) *Nat. Genet.* **11**, 115–116.
- Peifer, M., Berg, S. & Reynolds, A. B. (1994) *Cell* **76**, 789–791.
- Kussel, P. & Frasch, M. (1995) *J. Cell Biol.* **129**, 1491–1507.
- Kussel, P. & Frasch, M. (1995) *Mol. Gen. Genet.* **248**, 351–363.
- Torok, I., Strand, D., Schmitt, R., Tick, G., Torok, T., Kiss, I. & Mechler, B. M. (1995) *J. Cell Biol.* **129**, 1473–1489.
- Prieve, M. G., Guttridge, K. L., Munguia, J. E. & Waterman, M. L. (1996) *J. Biol. Chem.* **271**, 7654–7658.
- Nadler, S. G., Tritschler, D., Haffar, O. K., Blake, J., Bruce, A. G. & Cleveland, J. S. (1997) *J. Biol. Chem.* **272**, 4310–4315.
- Gorlich, D., Prehn, S., Laskey, R. A., Hartmann, E. (1994) *Cell* **79**, 767–778.
- Ruediger, R., Hentz, M., Fait, J., Mumby, M. & Walter, G. (1994) *J. Virol.* **68**, 123–129.
- Moroianu, J., Blobel, G. & Radu, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6572–6576.
- Hirschl, D., Bayer, P., Muller, O. (1996) *FEBS Lett.* **383**, 31–36.
- Yano, R., Oakes, M., Yamagishi, M., Dodd, J. A. & Nomura, M. (1992) *Mol. Cell. Biol.* **12**, 5640–5651.
- Peifer, M. (1995) *Trends Cell Biol.* **5**, 224–229.
- Fornerod, M., van Deursen, J., van Baal, S., Reynolds, A., Davis, D., Murli, K. G., Fransen, J. & Grosveld, G. (1997) *EMBO J.* **16**, 807–816.
- Gorlich, D., Dabrowski, M., Bischoff, F. R., Kutay, U., Bork, P., Hartmann, E., Prehn, S. & Izaurralde, E. (1997) *J. Cell Biol.* **138**, 65–80.
- Gorlich, D., Kraft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R. A., Mattaj, I. W. & Izaurralde, E. (1996) *Cell* **87**, 21–32.
- Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) *Comput. Appl. Biosci.* **8**, 189–191.
- Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
- Mehta, P. K., Heringa, J. & Argos, P. (1995) *Protein Sci.* **4**, 2517–2525.
- Swofford, D. L. (1991) *Illinois Natural History Survey*, Champaign, IL.
- Huber, A. H., Nelson, W. J. & Weis, W. I. (1997) *Cell* **90**, 871–882.