Biochemical and Biophysical Research Communications **288,** 137–141 (2001)
doi:10.1006/bbrc.2001.5722, available online at http://www.idealibrary.com on IDEAL®

# Identification of Nash1, a Novel Protein Containing a Nuclear Localization Signal, a Sterile $\alpha$ Motif, and an SH3 Domain Preferentially Expressed in Mast Cells

Takafumi Uchida,*,† Atsuhito Nakao,*,[1] Nobuhiro Nakano,* Atsuo Kuramasu,‡ Hirohisa Saito,§ Ko Okumura,* and Hideoki Ogawa*,†

**Atopy (Allergy) Research Center and †Department of Dermatology, Juntendo University School of Medicine, Tokyo 113-8421, Japan; ‡Department of Cellular Pharmacology, Tohoku University School of Medicine, Miyagi 980-8575, Japan; and §Division of Allergy, National Children's Medical Research Center, Tokyo 154-8509, Japan*

Received September 6, 2001

**By using a serial analysis of gene expression (SAGE), we have identified a novel full-length cDNA that is preferentially expressed in human cord blood-derived mast cells. The predicted protein showed unique primary structure with a nuclear localization signal (NLS), a sterile $\alpha$ motif (SAM), and a Src homology 3 domain (SH3) (termed Nash1). Nash1 was mapped to human chromosome 21q11.1 and highly expressed in spleen, liver, peripheral blood, and mast cell lines. In consistent with the presence of NLS, Nash1 was localized in the nucleus. Interestingly, screening gene databases for Nash1-related sequences revealed the existence of a Nash1-related gene termed SLY that was preferentially detected in lymphoid cells. We also found at least two additional candidates for this gene family in the database. These findings suggested that Nash1 and Nash1-related proteins consisted of a novel family of signaling/adaptor proteins, and Nash1 might function as a signaling component of mast cells, possibly in the nucleus.** 

***Key Words:*** **NLS; SAM; SH3; mast cells; SAGE.**

We have previously reported gene expression profile preferentially detected in human cord blood-derived mast cells by using a serial analysis of gene expression (SAGE) (1). SAGE was performed on poly(A)$^+$ RNA purified from human cord blood-derived mast cells and a total of 22914 tags, representing 9181 unique transcripts, were collected and sequenced. By selecting tags that were detected more frequently in the human cord blood-derived mast cells than in other tissues including brain, colon, mammary gland, ovary, kidney, prostate, and skeletal muscle, we enriched genes characteristic of mast cells. Importantly, the study revealed that about 70% of the selected genes were previously uncharacterized, suggesting that some of the uncharacterized genes were important for mast cell functions.

In this study, we characterized one of the novel genes that were found to be preferentially expressed in human cord blood-derived mast cells by our SAGE method. The predicted protein of the novel gene contains a nuclear localization signal (NLS) (2), a sterile $\alpha$ motif (SAM) (3), and a Src homology 3 domain (SH3) (4, 5) (termed Nash1), suggesting that Nash1 functions as a signaling component of mast cells. Interestingly, we found some Nash1-related sequences in the database. Thus, Nash1 and its related proteins may consist of a novel family of possible signaling/adaptor proteins.

## MATERIALS AND METHODS

*cDNA cloning.* Human Nash1 cDNA was identified as an insert sequence of expressed sequence tag (EST) clone (purchased from Genome Systems Inc.; AA741162) in an attempt to search for genes preferentially expressed in human cord blood-derived mast cells using SAGE (1). Briefly, SAGE was performed according to the Serial Gene Analysis of Gene Expression Detailed Protocol version 1.0c and analyzed using SAGE analysis software 3.04b generously provided by K. W. Kinzler (Johns Hopkins University, Baltimore, MD) with minor modification by us as reported (1). The SAGE data obtained from human cord blood-derived mast cells were compared with 15 other SAGE libraries from various tissues or cells. The SAGE libraries we used were NC1, NC2, BB542, NHA (fifth), normal pool (sixth), normal cerebellum, Duke thalamus, mammary epithelium, Br N, HOSE4, ISOSE29-11, 293-CTRL, Chen Normal Pr, Duke HMVEC from NCBI SAGE Library Browser (www.ncbi.nlm.nih.gov/SAGE/sagelb.cgi) and skeletal muscle (6) from Rochester Muscle Database web site (www.urmc.roshester.edu/smd/CRC/Swindex.html). We found that the tag sequence of the uncharacterized gene (human Nash1) (EST; AA741162) was expressed in human cord blood-derived mast cells about fivefold greater than any other tissues (1). Mouse Nash1 cDNA was cloned in the screening of mouse liver cDNA libraries using human cDNA as a probe. Rat Nash1 cDNA was identified by RT-PCR from rat RBL-2H3 cells using a sense primer,

[1] To whom correspondence should be addressed at Atopy (Allergy) Research Center, Juntendo University School of Medicine, 2-1-1, Hongo, Bunkyo-ku, Tokyo, 113-8421, Japan. Fax: +81-3-3813-5512. E-mail: anakao@med.juntendo.ac.jp.



0006-291X/01 $35.00

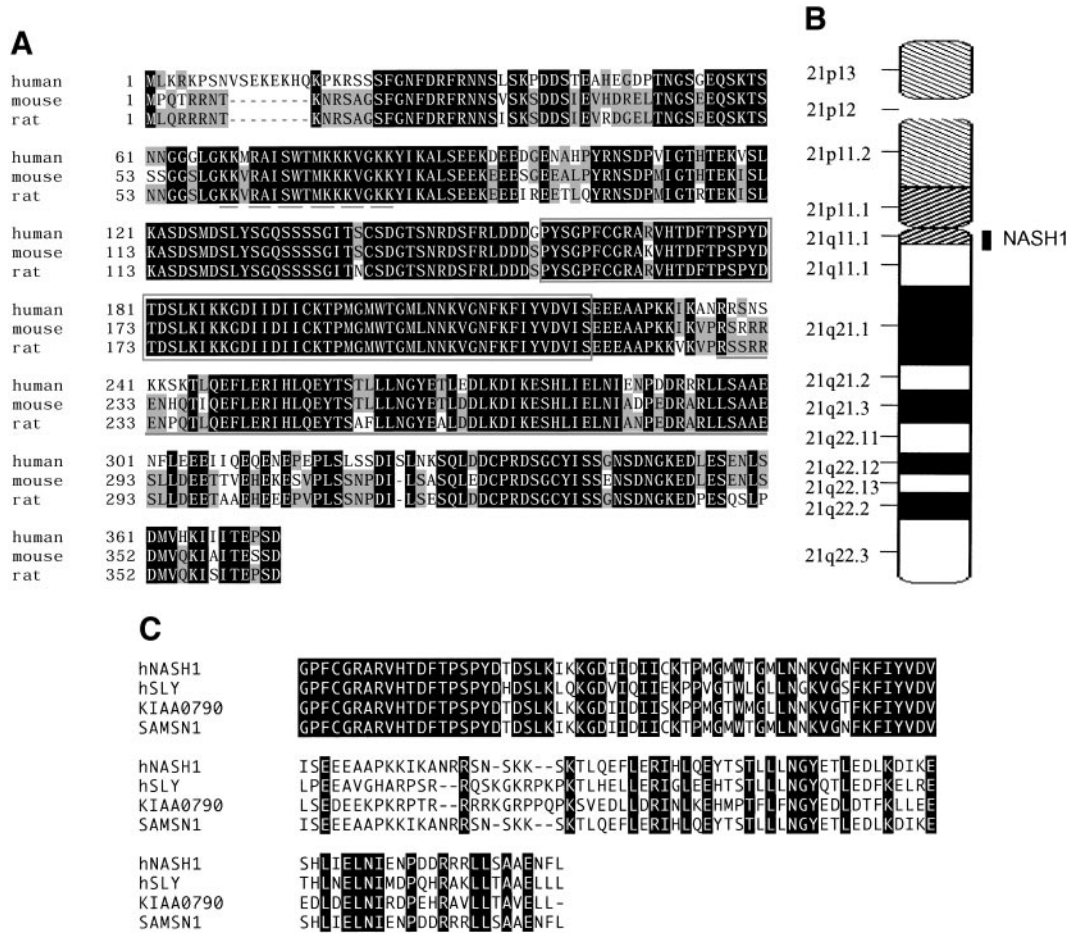


**FIG. 1.** Structural characteristics and alignment of human, mouse, and rat Nash1. (A) Deduced amino acid sequences and homology between human, mouse, and rat Nash1. Nuclear localization signal is indicated by a dotted line, SH3 domain is boxed, and SAM domain is underlined. Sequence matches are highlighted. (B) Diagrammatic representation of the genomic locus encoding Nash1. (C) Alignment of the SH3 and SAM domain of human Nash1 (AA74162) with human homologue of SLY, SAMSN1 (AF218085), and KIAA0790 (XM044017).

5′-AATGCTAAAGAGGAAGCCATCCAATG-3′, and an anti-sense primer, 5′-GGGTGTGGACATGAGAGTGTGTGTTCAGTCACT-3′, that were chosen based on regions of sequences conserved between human and mouse Nash1 cDNAs.

*Northern blotting.* RNA samples from Ramos, KU-812, Jurkat, U937, THP-1, HMC-1, NIH3T3, and HEK293 cell lines were prepared using Isogen solution (Nippon Gene) according to the manufacturer's recommendation. Northern blot analysis was performed as described before (1). A human multiple tissue membrane for Northern blot was purchased from Clontech. The cDNA restriction fragment encoding complete coding region of human Nash1 was used as a probe.

*Subcellular localization of Nash1.* The full-length of human Nash1 cDNA was subcloned into the pEGFP-C1 expression vector (Clontech), which placed the green fluorescent protein (GFP) at the N-terminus of the translated protein. COS7 cells were transiently transfected with the GFP-Nash1 expression vector using lipofectin method (Fugene6, Boehringer Mannheim) according to the manufacturer's recommendation. After 24 h of the transfection, cells were washed with PBS twice and then fixed in 4% paraformaldehyde for 30 min. Cells were then washed with PBS and treated with 100 $\mu$g/ml of RNase A for 30 min at 37°C. Cells were then mounted in mounted media containing propidium ionide (1.5 $\mu$g/ml). The cells were examined with a confocal imaging system (LM510; Carl Zeiss).

*The nucleotide accession number.* The nucleotide sequence data reported in this paper appeared in the DDBJ/EMBL/GenBank with Accession No. of AB067679 (rat Nash1).

## RESULTS AND DISCUSSION

We identified an EST clone, AA741162, as a gene preferentially expressed in human cord blood-derived mast cells using SAGE (1). The tag sequence of the EST clone AA741162 was expressed in human cord blood-derived mast cells about fivefold greater than any other tissues (1). The open reading frame of AA741162 predicted a protein of 373 amino acids in length with a candidate molecular mass of 42 kD. Comparison of the deduced amino acid sequences of this protein with protein sequences in the Swiss Protein Data Bank and SMART (7) predicted that the protein contained an NH2-terminal putative bisplit nuclear localization signal K(K/R)K motifs (residue 68–84) (2), a SH3 domain (residue 159–223) (4, 5), and a COOH-

terminal SAM domain (residue 236–300) (3), but lacked any known catalytic domain (Fig. 1A). Due to its unique primary structure, we termed the protein as Nash1 (a NLS, SAM, and SH3 containing protein1) and decided to further characterize the gene. We also cloned mouse Nash1 full-length cDNA and rat Nash1 full-length cDNA using human cDNA as a probe for screening mouse liver cDNA library or by RT-PCR using RBH-2H3 cell mRNA with two primers of which sequences were conserved between mouse and human cDNAs, respectively. The homology in amino acids between human and mouse Nash1 or mouse and rat Nash1 was 80.5 and 90.4%, respectively (Fig. 1A).

Human Nash1 cDNA sequence was compared with the Human Genome Project Data Base and the chromosomal localization of Nash1 gene was mapped to the chromosome 21q11.1 (Fig. 1B). Nash1 gene was found to comprise eight exons.

Searching the database (BLAST) for Nash1-related sequences revealed the existence of Nash1-related protein termed SLY that was recently reported by Beer *et al.* (8). SLY was cloned based on an expression cloning screen of a murine T cell lymphoma cDNA library and the human homologue was mapped to chromosome Xq25-26.3. Human homologue of SLY has 41.8% identity with human Nash1 in amino acid sequence. SLY also contains the unique structure with a COOH-terminal SAM domain and a SH3 domain in the middle portion (Fig. 1C).

Interestingly, additional Nash1-related sequences, SAMSN1 (AF218085) and KIAA0790 (XM044017) were found in the database, both of which had a COOH-terminal SAM domain and a SH3 domain in the middle portion (Fig. 1C), although the 5′ region of KIAA0790 were not identified yet. Amino acid sequence of SAMSN1 was identical to the amino acid residue 19 to 343 of human Nash1. SAMSN1 is thus likely to be an alternative splicing variant form of Nash1.

The distribution of Nash1 in human tissues was determined by Northern blot analysis of mRNA (Fig. 2A). Nash1 transcript of approximately 4.3 kb was highly detected in spleen, liver, and peripheral blood lymphocytes. We also examined expression of Nash1 in several human cell lines of hematopoietic cell origin and found that Nash1 was highly expressed in KU812 and HMC1 human mast cell/basophil cell lines (Fig. 2B), which was consistent with the fact that Nash1 was identified as a gene preferentially expressed in human cord-blood derived mast cells by SAGE. A Nash1-related gene, SLY, was reported to be highly expressed in lymphoid tissues such as spleen, thymus, lymph node, and bone marrow and also in lung, but not in liver (8). Different expression pattern observed between Nash1 and SLY may suggest distinct role of the two related proteins.

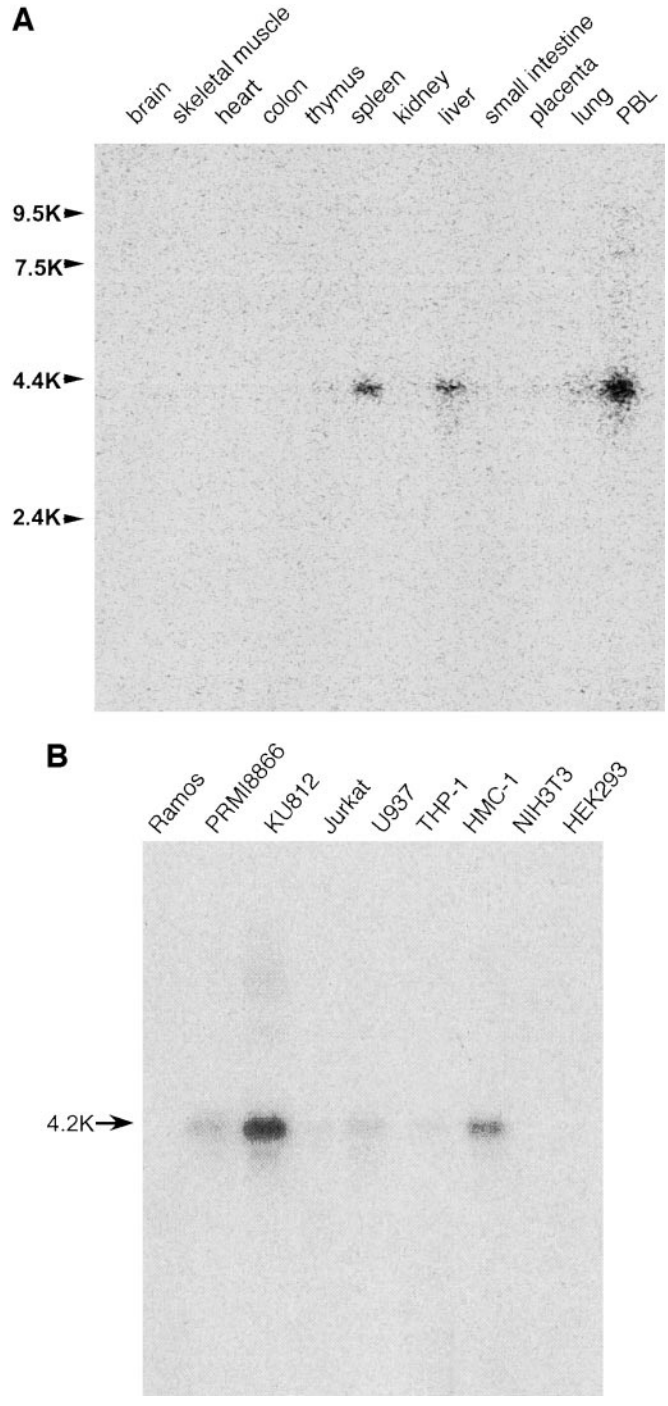


**FIG. 2.** Northern blot analysis of human Nash1 mRNA from human tissues and cell lines. (A) Human multiple tissue Northern blot membrane or (B) Northern blot containing total RNA of various cell lines (lane 1, Ramos; lane 2, RPMI8866; lane 3, KU812; lane 4, Jurkat; lane 5, U937; lane 6, THP-1; lane 7, HMC1; lane 8, NIH3T3; lane 9, HEK293) were hybridized with a full-length human Nash1 cDNA probe. Reprobing of the blot with a $\beta$-actin demonstrated equivalent RNA loading (data not shown).

To gain insight for possible function of Nash1, subcellular organization of Nash1 was analyzed by immunofluorescence study. As shown in Fig. 3A, the GFP-

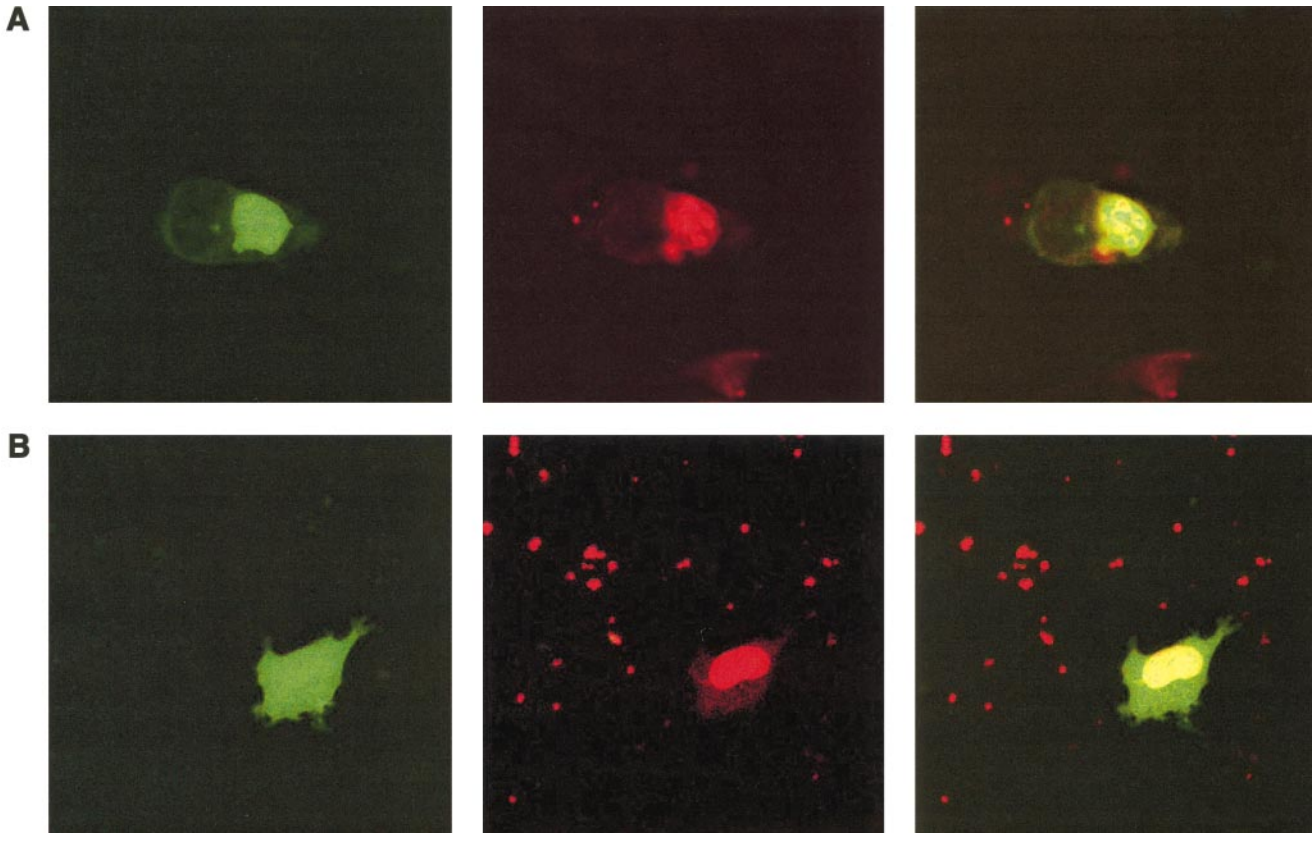


**FIG. 3.** Subcellular localization of Nash1. GFP was placed in frame with the NH2-terminus of human Nash1 cDNA and the fusion plasmid, GFP-Nash1 (A), or a control GFP plasmid (B) was expressed in COS7 cells. The GFP fluorescence was visualized using confocal microscopy (left panels) and the nuclei was counterstained by propidium iodide (middle panels). Right panels were the merging of the two staining. Colocalization was indicated by yellow.

Nash1 fusion construct was localized in the nucleus after transfection of the construct into COS7 cells, which was in agreement with the presence of NLS sequence in Nash1. Interestingly, the nuclear staining pattern was 'granular'. In contrast, fluorescence was seen throughout the cytoplasm and nucleus in COS7 cells transfected with a control GFP construct (Fig. 3B). We observed similar results in HEK293 or Hela cells (data not shown), suggesting that nuclear localization of Nash1 was a general phenomenon. Mutational studies in the NLS sequence or identification of NLS binding proteins will be required for further characterization of the NLS sequence of Nash1 (9).

SH3 domains are present in a large number of intracellular proteins such as non-receptor type tyrosine kinases, signaling adaptor molecules, transcriptional factors and suggested to plays a critical role in the formation of signaling complexes through the binding to prolin-rich sequences of proteins with the consensus PXXP (4, 5). SAM domain is also a novel protein module of around 70 amino acids that is found in a variety of signaling molecules (3). The presence of these modules in Nash1 thus suggested that Nash1 functioned as a signaling/adaptor component. Taken together with preferential expression of Nash1 in mast cells, future study will be directed to clarify the role of Nash1 in the signal transduction of mast cells.

It is reported that the SAM domain can potentially function as a protein interaction module through the ability to homo- and heterooligomerize with other SAM domains (3, 10). Therefore, the existence of Nash1-related proteins, SLY, SAMSN1, and KIAA0790, suggest that Nash1 might function as a homodimer or heterooligomer with other Nash1-related proteins through their SAM domains. This possibility is under investigation.

In summary, we identified Nash1 as a novel gene that was preferentially expressed in mast cells and was mapped to the chromosome 21q11.1. Nash1 has unique primary structure with an NH2-terminal NLS, a SH3 domain in the middle portion, and a COOH-terminal SAM domain. In consistent with the presence of NLS, Nash1 was localized in the nucleus. Interestingly, we found additional Nash1-related genes with similar primary structure in the database. These findings suggested that Nash1 and Nash1-related proteins consisted of a novel family of signaling/adaptor proteins

and Nash1 might function as a signaling component of mast cells, possibly in the nucleus.

## ACKNOWLEDGMENTS

We thank K. Maeda, H. Ushio, C. Nishiyama, T. Takai, S. Masaki, Y. Akizawa, T. Shimokawa, M. Terada, K. Hira, H. Suto, S. Nakamura, K. Miyahara, and T. Tokura for discussion and technical assistance and E. Kawasaki and M. Matsumoto for secretarial assistance. This work was supported in part by grants from the Ministry of Education, Science, Sports, and Culture, Japan.

## REFERENCES

1. Kuramatsu, A., Kubota, Y., Matsumoto, K., Nakajima, T., Sum, X.-M., Watanabe, T., Saito, H., and Ohtsu, H. (2001) *FEBS Lett.* **2902,** 1–5.
2. Boulikas, T. (1993) *Crit. Rev. Eukaryote Gene Exp.* **3,** 193–227.2.
3. Pasquale, E. B. (1997) *Curr. Opin. Cell Biol.* **9,** 608–615.
4. Pawson, T. (1995) *Nature* **342,** 624.
5. Musacchio, T., Gibson, T., Lehto, V. P., and Saraste, M. (1992) *FEBS Lett.* **307,** 55–61.
6. Welle, S., Bhatt, K., and Thorton, C. A. (1999) *Genome Res.* **9,** 506–513.
7. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5857–5864.
8. Beer, B., Simins, A. B., Schuster, A., and Holzmann, B. (2001) *Biochim. Biophys. Acta* **1520,** 89–93.
9. Mattaj, I. W., and Englmeier, L. (1998) *Annu. Rev. Biochem.* **67,** 265–306.
10. Stapleton, D., Balan, I., Pawson, T., and Sicheri, F. (1999) *Nat. Struct. Biol.* **6,** 44–49.