ACCELERATED COMMUNICATION

The C terminus of the nuclear protein NuMA: Phylogenetic distribution and structure

PATRICIA C. ABAD,[1,3] I. SAIRA MIAN,[2,3] CEDRIC PLACHOT,[1]
ANIYSHA NELPURACKAL,[1] CAROL BATOR-KELLY,[1] AND SOPHIE A. LELIÈVRE[1]

[1]Department of Basic Medical Sciences, Purdue University, West Lafayette, Indiana 47907-2026, USA
[2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

(RECEIVED June 1, 2004; FINAL REVISION June 1, 2004; ACCEPTED July 11, 2004)

Abstract

The C terminus of the nuclear protein NuMA, NuMA-CT, has a well-known function in mitosis via its proximal segment, but it seems also involved in the control of differentiation. To further investigate the structure and function of NuMA, we exploited established computational techniques and tools to collate and characterize proteins with regions similar to the distal portion of NuMA-CT (NuMA-CTDP). The phylogenetic distribution of NuMA-CTDP was examined by PSI-BLAST- and TBLASTN-based analysis of genome and protein sequence databases. Proteins and open reading frames with a NuMA-CTDP-like region were found in a diverse set of vertebrate species including mammals, birds, amphibia, and early teleost fish. The potential structure of NuMA-CTDP was investigated by searching a database of protein sequences of known three-dimensional structure with a hidden Markov model (HMM) estimated using representative (human, frog, chicken, and pufferfish) sequences. The two highest scoring sequences that aligned to the HMM were the extracellular domains of β3-integrin and Her2, suggesting that NuMA-CTDP may have a primarily β fold structure. These data indicate that NuMA-CTDP may represent an important functional sequence conserved in vertebrates, where it may act as a receptor to coordinate cellular events.

Keywords: nuclear mitotic apparatus protein; β3-integrin; chordate; mammary epithelial cells; differentiation

NuMA is widely expressed in the nuclei of mammalian cells (Kallajoki et al. 1992; Tang et al. 1993; Lelièvre et al. 1998). A prominent feature of this 2101-amino-acid protein is an unusually long coiled-coil domain spanning residues 216–1700, a region similar to the coiled-coils found in structural proteins like myosin heavy chains, cytokeratins, and nuclear lamins (Yang et al. 1992; Harborth et al. 1995). The exact function of the NuMA coiled-coil remains unknown. In contrast, the globular NuMA-NT (residues 1–215) and NuMA-CT (residues 1701–2101) flanking the coiled-coil are associated with the function of NuMA in mitosis (Compton and Cleveland 1993). On the basis of sequence analysis, a calponin homology domain has been reported within NuMA-NT and proposed as a likely interaction site for actin-related protein 1 during mitosis (Novatchkova and Eisenhaber 2002). The proximal portion of NuMA-CT contains binding sites for several proteins involved in the control of mitosis, including tubulin, LGN, and protein 4.1R (Fig. 1; Mattagajasingh et al. 1999; Du et al. 2001; Haren and Merdes 2002).

Interestingly, NuMA-CT seems also associated with functions other than mitosis. A fusion protein between NuMA lacking the distal portion of its C terminus (trun-

Reprint requests to: Sophie A. Lelièvre, Department of Basic Medical Sciences, Purdue University, 625 Harrison Street, LYNN, West Lafayette, IN 47907-2026, USA; e-mail: lelievre@purdue.edu; fax: (765) 494-0781.

[3]These authors contributed equally to this work.

Abbreviations: ESTs, expressed sequence tags; HMM, hidden Markov model; NCBI, National Center for Biotechnology Information; NuMA, nuclear mitotic apparatus protein; NuMA-CT, NuMA C terminus; NuMA-CTDP, distal portion of NuMA CT; NuMA-NT, NuMA N terminus; RAR, retinoic acid receptor; RCSB, Research Collaboratory for Structural Biology; SAM, Sequence Alignment and Modeling.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.04906804.

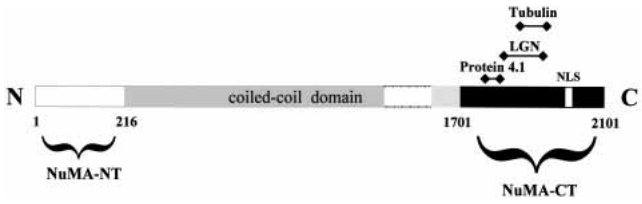

Figure 1. Protein binding regions within NuMA-CT. Numbers designate the position of amino acids. The nuclear localization signal (NLS) is indicated.

cated at amino-acid 1883) and the RAR (Wells et al. 1997) has been shown to prevent neutrophil differentiation and proposed to be involved in acute promyelocytic leukemia by altering NuMA function (Sukhai et al. 2004). In addition, introduction of antibodies directed against NuMA-CT or expression of amino acids 1965–2101 in human mammary epithelial HMT-3522 S1 cells disrupted the distribution of NuMA and altered chromatin structure and differentiation into growth-arrested mammary epithelial tissue structures (acini) obtained on culture in Matrigel (Lelièvre et al. 1998; P. Abad and S. Lelièvre, unpubl.). Thus, NuMA-CT appears to play a role in differentiation. To further decipher the function of NuMA, we have investigated the phylogenetic distribution and structure of NuMA-CTDP.

Results and Discussion

Proteins with a NuMA-CTDP-like region are present in a phylogenetically diverse set of organisms

To date, NuMA-like proteins have been identified in *Homo sapiens* (human), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Bos taurus* (cow), and *Xenopus laevis* (frog). To identify proteins possessing NuMA-CTDP-like regions, we used human NuMA residues 1915–2095, designated hNuMA-CTDP, as the protein query for TBLASTN searches against databases of genomic sequences or ESTs. Both the region missing in the NuMA-RAR fusion protein (Wells et al. 1997) and a region believed to be involved in the control of mammary differentiation (P. Abad and S. Lelièvre, unpubl.) are included within hNuMA-CTDP. We found sequences with statistically significant similarity to hNuMA-CTDP in *Sus scrofa* (pig), *Gallus gallus* (chicken), *Danio rerio* (zebrafish), and *Takifugu rubripes* (pufferfish; Fig. 2). The putative pig and chicken family members were translated ESTs. For pufferfish, part of *T. rubripes* genomic scaffold 307 contained short peptide matches in the −1 frame. This scaffold sequence and the Web version of GENSCAN (Burge and Karlin 1997) were used to identify an 865-amino-acid open reading frame containing the hNuMA-CTDP-like region at its C terminus. The zebrafish hNuMA-CTDP relative was identified in a similar manner using the February 2004 Ensembl assembly of the genome (http://www.ensembl.org/Danio_rerio). Given the techniques and tools available currently, we found no convincing evidence for hNuMA-CTDP family members in three other metazoa with fully sequenced genomes (*Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabadopsis thaliana*). However, we did find a short match to hNuMA-CTDP in a *Ciona intestinalis* genomic scaffold. *C. intestinalis* is considered to be one of the earliest chordates because although the larval stage has a notochord, it is lost in the adult stage. These observations suggest a relatively broad phylogenetic distribution for members of the NuMA-CTDP family, with convincing evidence for their presence in vertebrates (mammals: human, mouse, rat, pig; amphibians: frog; birds: chicken; and teleost fish: zebrafish and pufferfish) and a possibility that NuMA itself might be restricted to the chordate lineage.

NuMA-CTDP family members may possess a fold similar to the extracellular domain of human β3-integrin

To gain insight into the structure and thus function of NuMA-CTDP, we sought to predict a possible three-dimensional structure for the NuMA-CTDP family. There were no matches when hNuMA-CTDP was used as the query for an NCBI Conserved Domain Database Search. Therefore, we estimated an HMM from a set of diverse NuMA-CTDP sequences, hNuMA-CTDP plus its frog, chicken, and teleost relatives (HsNuMA1, XlNuMA, GgEST, and FrORF, as shown in Fig. 1) using the SAM System (http://www.soe.ucsc.edu/research/compbio/sam.html). The resulting NuMA-CTDP HMM was used to search a database of protein sequences of known three-dimensional structure provided by the RCSB (http://www.rcsb.org). The highest scoring protein of the ~55,000 sequences in the February 2004 release from the RCSB was human β3-integrin (1m1x). The sequence of this region of β3-integrin is highly conserved in human β1- and β6-integrins. The next highest scoring protein was another membrane protein, oncoprotein receptor tyrosine kinase Her2 (1n8y). Although the scores of 1m1x and 1n8y against the NuMA-CTDP HMM were not statistically significant (E values above 1.0), the results are of biological interest because the regions of similarity correspond to largely β extracellular domains that regulate protein interactions and are involved in signaling (Borges et al. 2000; Hynes 2002). These results suggest that the three-dimensional structure of NuMA-CTDP might be primarily β, and that this region could be a site of intermolecular interaction(s).

If NuMA-CTDP is a site of interaction with other proteins, then by analogy with β3-integrin, potential binding partners may include proteins located at the plasma membrane. More specifically, this observation suggests that in addition to its known location in the nucleus, NuMA could

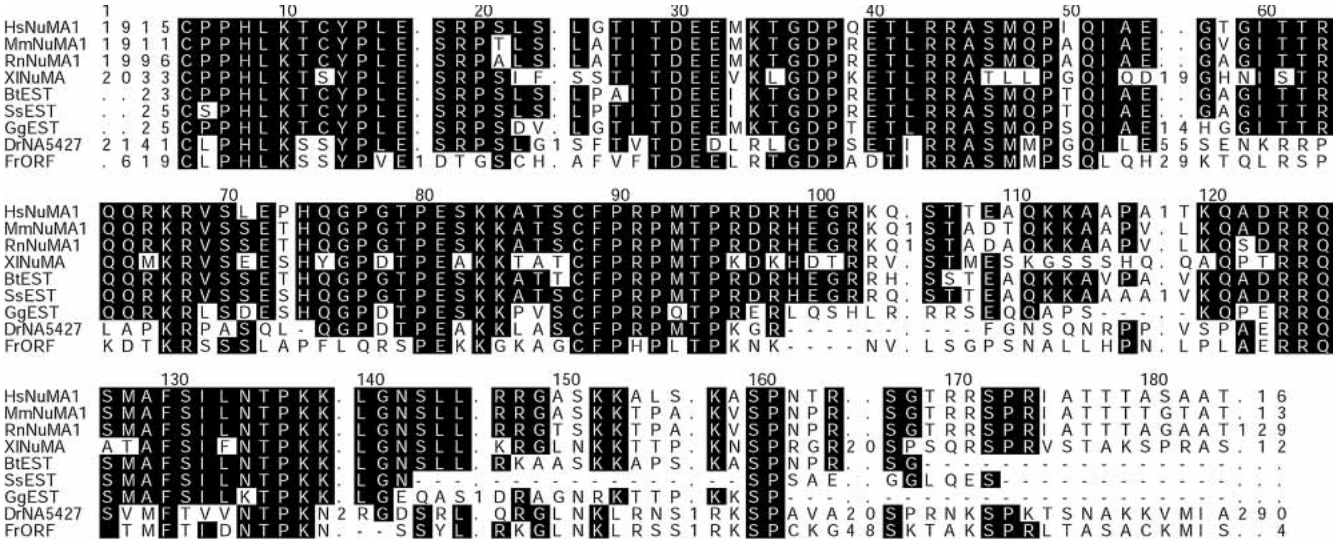

Figure 2. Multiple sequence alignment of the C-terminal non-coiled-coil segment of NuMA family members. The sequences shown are HsNuMA1 (*Homo sapiens* NuMA; databank code NP_006176; human), MmNuMA1 (*Mus musculus* NuMA; NP_598708; mouse), RnNuMA1 (*Rattus norvegicus* NuMA; XP_218972.2; rat), XlNuMA (*Xenopus laevis* NuMA; T30336; frog), BtEST (*Bos taurus* EST, BI680620; GI: 15633534; cow), SsEST (*Sus scrofa* EST; BX673330.1, GI: 37986836; pig), GgEST (*Gallus gallus* EST; ChEST630g3, [http://chick.umist.ac.uk/] accession no.: 603737612F1; chicken), DrNA5427 (*Danio rerio* Ensembl GENSCAN predicted peptide; NA5427.1.260.43185; zebrafish), and FrORF (*Takifugu rubripes* GENSCAN predicted peptide in *Fugu* scaffold 307; pufferfish). Positions conserved in at least five of the nine sequences are highlighted; the number of residues not displayed explicitly is listed.

be present at or near the cell membrane. Indeed, full-length NuMA is observed in the cytoplasm of cells expressing oncogenic fusion protein NuMA-RAR (Hummel et al. 2002). Furthermore NuMA binds to protein 4.1R, a protein that participates in tethering the cytoskeleton to the plasma membrane in addition to being located at the poles of the mitotic spindle (Mattagajasingh et al. 1999). Although NuMA could be detected in the cytoplasm of S1 cells differentiated into acini, as seen on electron micrographs (not shown), it was not located at the cell membrane and was not present in preparations of crude membrane extracts that include the plasma membrane, Golgi apparatus, and rough endoplasmic reticulum (Fig. 3). To rule out the possibility that the absence of NuMA in membrane fractions was a characteristic of acinar differentiation, we performed the same analysis using collagen I culture of S1 cells, which induces the formation of incorrectly polarized acini, and HMT-3522 T4–2 malignant cells derived from S1 cells. T4–2 cells form disorganized tumor nodules when cultured in Matrigel (Weaver et al. 1997). NuMA was still absent from crude membrane fractions prepared under these conditions (Fig. 3). Thus, it seems unlikely that NuMA interacts with proteins at the cell surface.

The presence of NuMA-CTDP family members in vertebrates including species with extreme phylogenetic separation (i.e., mammals and teleost fish), together with their likely absence in invertebrates, suggests that NuMA-CTDP may have a highly conserved vertebrate-specific role. The investigation of the three-dimensional structure of NuMA-CT is a key tool to help further decipher NuMA functions. It has been proposed that the proximal portion of NuMA-CT, which contains binding sites for tubulin and LGN and is important for the function of NuMA during mitosis, may consist of an α helix structure (Haren and Merdes 2002). In our study, we describe a potential structural similarity between NuMA-CTDP and the extracellular domain of β3-integrin, which has a largely β structure. The β strand is considered to bring maximum exposure for ligand binding and may lead to intermolecular linkage (Chow et al. 2004). This secondary structure has been identified as a critical recognition element in physiological processes (Glenn and Fairlie 2002) and is present in numerous proteins involved in signal transduction. These data suggest that NuMA may also function to orchestrate cellular events.

Given that NuMA has a role in differentiation (Lelièvre et al. 1998; Sukhai et al. 2004), one hypothesis is that NuMA-CTDP may be associated with the control of gene expression. This hypothesis is supported by the fact that expression of NuMA truncated at its C terminus and antibodies directed against NuMA-CT induce alterations in chromatin organization (Gueth-Hallonet et al. 1998; Lelièvre et al. 1998). Recently proteins bearing actin-binding domains have been proposed to play a critical role in the control of gene expression by providing a structural framework that facilitates and integrates molecular cross-talk within the nucleus (Shumaker et al. 2003). Thus, with its calponin homology domain at the N terminus and a possible β structure for NuMA-CTDP, NuMA may provide a structural

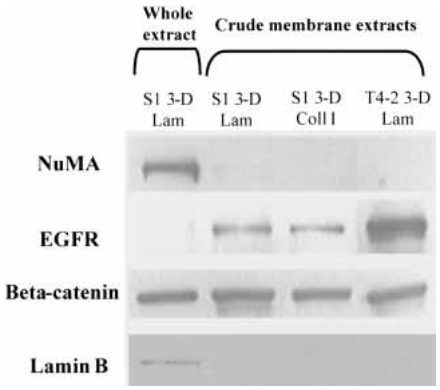

Figure 3. NuMA is absent from the plasma membrane of HMT-3522 cells. Non-neoplastic S1 and malignant T4–2 HMT-3522 cells were cultured for 10 d in the presence of Matrigel (S1 3-D Lam and T4–2 3-D Lam) or collagen I (S1 3-D Coll I). Western blots of whole-cell extracts and crude membrane fractions are shown for NuMA, plasma membrane region markers epidermal growth factor receptor (EGFR) and β-catenin, and nuclear marker lamin B.

platform for the transduction and coordination of signals involved in the regulation of gene expression.

Materials and methods

Sequence analysis

We used *H. sapiens* NuMA residues 1915–2095 as the query for TBLASTN searches against the complete genomes or ESTs (protein vs. translated DNA) of *B. taurus*, *S. scrofa*, *R. norvegicus*, *M. musculus*, *G. gallus*, *X. laevis*, *T. rubripes*, *D. rerio*, *C. intestinalis*, *C. elegans*, *D. melanogaster*, and *A. thaliana*. Statistical modeling, alignment, and database searching were performed using HMMs as implemented in the SAM software (Hughey and Krogh 1996). TBLASTN and all other programs used in this work were run with default parameter settings.

Cell culture

HMT-3522 nonneoplastic (S1) and malignant (T4–2) cells were cultured in H14 medium (Weaver et al. 1997). To induce differentiation into acini and the formation of tumor-like nodules, we cultured S1 and T4–2 cells, respectively, for 10 d on 40 $\mu L/cm^2$ Matrigel (BD Biosciences)-coated surfaces in the presence of culture medium containing 5% Matrigel. Collagen I culture of S1 cells was performed as previously described (Lelièvre et al. 1998).

Crude membrane extract preparation

Cellular structures were collected as described earlier (Lelièvre et al. 1998) and crude membrane fractionations were performed according to standard procedures. Briefly, after dissociation of nuclei and cytoplasms, cytoplasms were deposited on top of 10 mL of a buffer containing 10 mM HEPES (pH 7.4), 1 mM ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid, and 2 mM $MgCl_2$ and centrifuged (4°C, 114,000*g*) for 75 min to separate cytoplasmic (supernatant) from crude membrane (pellet) proteins. Crude membrane fractions were analyzed by Western blot using antibodies against lamin B and NuMA (clone 204.4; Oncogene Research Products) and β-catenin and epidermal growth factor receptor (BD Transduction Laboratories).

Acknowledgments

We thank Jun Xie for helpful discussion. This work was supported by grants from the Department of Defense/Breast Cancer Research Program and the Walther Cancer Institute to S.A.L., California Breast Cancer Research Program to I.S.M., Howard Hughes Undergraduate Summer Research Fellowship to A.N., and the Department of Energy, Office of Health and Environmental Research to I.S.M.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Borges, E., Jan, Y., and Ruoslahti, E. 2000. Platelet-derived growth factor receptor β and vascular endothelial growth factor receptor 2 bind to the β 3 integrin through its extracellular domain. *J. Biol. Chem.* **275:** 39867–39873.

Burge, C. and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268:** 78–94.

Chow, M.K., Lomas, D.A., and Bottomley, S.P. 2004. Promiscuous β-strand interactions and the conformational diseases. *Curr. Med. Chem.* **11:** 491–499.

Compton, D.A. and Cleveland, D.W. 1993. NuMA is required for the proper completion of mitosis. *J. Cell Biol.* **120:** 947–957.

Du, Q., Stukenberg, P.T., and Macara, I.G. 2001. A mammalian partner of inscuteable binds NuMA and regulates mitotic spindle organization. *Nat. Cell Biol.* **3:** 1069–1075.

Glenn, M.P. and Fairlie, D.P. 2002. Mimetics of the peptide β-strand. *Mini Rev. Med. Chem.* **2:** 433–445.

Gueth-Hallonet, C., Wang, J., Harborth, J., Weber, K., and Osborn, M. 1998. Induction of a regular nuclear lattice by overexpression of NuMA. *Exp. Cell Res.* **243:** 434–452.

Harborth, J., Weber, K., and Osborn, M. 1995. Epitope mapping and direct visualization of the parallel, in-register arrangement of the double-stranded coiled-coil in the NuMA protein. *EMBO J.* **14:** 2447–2460.

Haren, L. and Merdes, A. 2002. Direct binding of NuMA to tubulin is mediated by a novel sequence motif in the tail domain that bundles and stabilizes microtubules. *J. Cell Sci.* **115:** 1815–1824.

Hughey, R. and Krogh, A. 1996. Hidden Markov models for sequence analysis: Extension and analysis of the basic method. *Comput. Appl. Biosci.* **12:** 95–107.

Hummel, J.L., Zhang, T., Wells, R.A., and Kamel-Reid, S. 2002. The retinoic acid receptor α (RARα) chimeric proteins PML-, PLZF-, NPM-, and NuMA-RARα have distinct intracellular localization patterns. *Cell Growth Differ.* **13:** 173–183.

Hynes, R.O. 2002. Integrins: Bidirectional, allosteric signaling machines. *Cell* **110:** 673–687.

Kallajoki, M., Weber, K., and Osborn, M. 1992. Ability to organize microtubules in taxol-treated mitotic PtK2 cells goes with the SPN antigen and not with the centrosome. *J. Cell Sci.* **102:** 91–102.

Lelièvre, S.A., Weaver, V.M., Nickerson, J.A., Larabell, C.A., Bhaumik, A., Petersen, O.W., and Bissell, M.J. 1998. Tissue phenotype depends on reciprocal interactions between the extracellular matrix and the structural organization of the nucleus. *Proc. Natl. Acad. Sci.* **95:** 14711–14716.

Mattagajasingh, S.N., Huang, S.C., Hartenstein, J.S., Snyder, M., Marchesi, V.T., and Benz, E.J. 1999. A nonerythroid isoform of protein 4.1R interacts with the nuclear mitotic apparatus (NuMA) protein. *J. Cell Biol.* **145:** 29–43.

Novatchkova, M. and Eisenhaber, F. 2002. A CH domain-containing N terminus in NuMA? *Protein Sci.* **11:** 2281–2284.

Shumaker, D.K., Kuczmarski, E.R., and Goldman, R.D. 2003. The nucleoskel-

eton: Lamins and actin are major players in essential nuclear functions. *Curr. Opin. Cell Biol.* **15:** 358–366.

Sukhai, M.A., Wu, X., Xuan, Y., Zhang, T., Reis, P.P., Dube, K., Rego, E.M., Bhaumik, M., Bailey, D.J., Wells, R.A., et al. 2004. Myeloid leukemia with promyelocytic features in transgenic mice expressing hCG-NuMA-RARα. *Oncogene* **23:** 665–678.

Tang, T.K., Tang, C.J., Chen, Y.L., and Wu, C.W. 1993. Nuclear proteins of the bovine esophageal epithelium. II. The NuMA gene gives rise to multiple mRNAs and gene products reactive with monoclonal antibody W1. *J. Cell Sci.* **104:** 249–260.

Weaver, V.M., Petersen, O.W., Wang, F., Larabell, C.A., Briand, P., Damsky, C., and Bissell, M.J. 1997. Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies. *J. Cell Biol.* **137:** 231–245.

Wells, R.A., Catzavelos, C., and Kamel-Reid, S. 1997. Fusion of retinoic acid receptor α to NuMA, the nuclear mitotic apparatus protein, by a variant translocation in acute promyelocytic leukaemia. *Nat. Genet.* **17:** 109–113.

Yang, C.H., Lambie, E.J., and Snyder, M. 1992. NuMA: An unusually long coiled-coil related protein in the mammalian nucleus. *J. Cell Biol.* **116:** 1303–1317.