cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins

(lamin sequence identity/divergent carboxyl termini/ α -helical domains/coiled coils/nuclear localization sequence)

DANIEL Z. FISHER, NILABH CHAUDHARY, AND GUNTER BLOBEL

Laboratory of Cell Biology, The Rockefeller University, New York, NY ¹⁰⁰²¹

Contributed by Gunter Blobel, May 13, 1986

ABSTRACT The amino acid sequences. deduced from cDNA clones of human lamin A and lamin C show identity between these two lamins except for an extra 9.0-kDa carboxylterminal tail that is present only in lamin A. Both lamins A and C contain an α -helical domain of approximately 360 residues that shows striking homology to a corresponding α -helical rod domain that, is the structural hallmark of all intermediate filament proteins. However, the lamin α -helical domain is 14% larger than that of the intermediate filament proteins. In addition to the extensive homology to intermediate filament proteins as reported [McKeon, F., Kirschner, M. & Caput, D. (1986) Nature (London) 319, 463-468], a different 82-amino acid residue stretch at the carboxyl terminus of lamin A has been deduced and verified by amino acid sequencing. This region contains sequence homology to amino- and carboxylterminal domains of type ^I and type II epidermal keratins. Implications of the presence of these and other domains in lamins A and C for the assembly of the nuclear lamina are discussed.

The nuclear lamina is a polymeric structure that is intercalated between chromatin and the inner membrane of the nuclear envelope. In mammalian cells the lamina consists primarily of three proteins $(1-4)$ that have been termed lamins A (70 kDa), B (67 kDa), and C (60 kDa) (5). Peptide maps of lamins A and C are similar, but they are different from those of lamin B (6). However, certain monoclonal antibodies have been shown to crossreact with all three lamins, indicating that some epitopes are shared by all three lamins (7, 8).

The lamins have been shown to be synthesized throughout the cell cycle (9). In vitro translation of mRNA (10) showed that lamins A and C are synthesized from separate mRNAs with lamin A being made as ^a precursor approximately ² kDa larger than mature lamin A. The precursor is converted into mature lamin A only after incorporation into the lamina structure, as shown by pulse-chase experiments (9).

The lamina is depolymerized during mitotic prophase by a process that may involve transient hyperphosphorylation (5). Concomitant with a gradual dephosphorylation, the lamins repolymerize in telophase (3, 5). How the regular protein meshwork (2) that constitutes the lamina is assembled from the three lamins is not known. The lamins must contain domains that are involved in this polymerization and that interact with chromatin on the nucleoplasmic face of the lamina, with the inner nuclear membrane on the other face of the lamina, and, most likely, with nuclear pore complexes (1, 2). As such domains could be revealed from the amino acid sequences of the lamins, we decided to clone their cDNAs. We report here the deduced amino acid sequences for lamins A and C. While this manuscript was in preparation, McKeon et al. (11) reported lamin sequence data that our findings

corroborate and extend. A difference in predicted amino acid sequence of the carboxyl-terminal ⁸² residues of the lamin A precursor was noted and confirmed by amino acid sequencing. This carboxyl-terminal domain showed sequence homology to the head and tail domains of keratins, thereby extending lamin A sequence homologies with intermediate filament proteins to regions outside the α -helical rod domain. Furthermore, a 762-base-pair (bp) 3'-untranslated region including the polyadenylylation signal for lamin A is now reported.

MATERIALS AND METHODS

Screening cDNA Libraries. Guinea pig antibodies to rat liver lamins A, B, and C (gift of L. Gerace) were used to screen a rat liver $\lambda g111$ cDNA library (gift of R. Hynes) according to established protocols (12, 13). Eight cDNA clones were obtained, the largest of which [1.3 kilobases (kb)] was used to screen (14) the simian virus 40-transformed human fibroblast cDNA library of Okayama and Berg (15) (gift of H. Okayama). A 2.6-kb partial cDNA clone of lamin A was obtained and sequenced (see below). A 22-bp oligonucleotide, TGGCCCCGCAGATCATGCAGCT, complementary to a ⁵' portion of the 2.6-kb clone was synthesized (Applied Biosystems 380A DNA Synthesizer) and used to rescreen the fibroblast library. A 1.8-kb partial cDNA clone of lamin C was obtained. A 21-bp oligonucleotide, GCATC-CCCGAGCTCGGCCTCG, complementary to a 5' portion of the 1.8-kb clone was synthesized and used to screen (16) a human hepatoma λ gtll cDNA library (gift of M. Mueckler) for full length cDNA clones. A 1962-bp cDNA clone of lamin C was obtained.

RNA Blot Analysis. Nitrocellulose blots of HeLa cell $poly(A)^+$ RNA (prepared according to refs. 17-19) were hybridized with nick-translated rat liver cDNA or ³²P-labeled oligonucleotides. Nick-translated cDNA clones of the heat shock protein hsp70 (2.9 kb) and actin (2.0 kb) (gift of H. Kao) were hybridized as internal M_r markers. Blots were washed for 1 hr at 22°C, air dried, and exposed to Kodak XAR-5 film for 1-14 hr.

DNA Sequence Analysis. Restriction fragments of the plasmid and phage cDNA inserts were subcloned into M13 vectors (20) and sequenced in both orientations by the dideoxy chain termination method (21, 22). The rapid deletion subcloning procedure of Dale et al. (23) was used to sequence large fragments and to sequence across restriction sites. Maxam and Gilbert sequencing (24) was used to sequence DNA regions not adequately resolved by the dideoxy method.

Lamin Sequencing. Rat liver nuclear pore complex-lamina fraction was purified according to established procedures (2). Lamins were solubilized in 6 M urea/10 mM Tris HCl, pH 8.0/20 mM dithiothreitol, loaded onto ^a DEAE-cellulose

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.

Abbreviations: bp, base pair(s); kb, kilobase(s).

column equilibrated in the above buffer, and eluted with a 0-200 mM NaCl gradient (S. Georgatos and G.B., unpublished observations).

For acid hydrolysis of Asp-Pro bonds only, eluted fractions enriched in lamins A and C were electrophoresed on 7.5-15% NaDodSO4/polyacrylamide gradient gels and electroblotted onto trifluoroacetic acid-treated glass fiber filters (Whatman GF/C) according to the low pH procedure of Aebersold et al. (25). Protein bands were visualized with Coomassie Blue, and lamin A and lamin C bands were cut from the filters. Asp-Pro cleavage was carried out by spotting 88% (vol/vol) formic acid (30 μ l/cm² of filter strip) and incubating for 20 hr at 44°C in a sealed chamber saturated with 88% (vol/vol) formic acid (D. Atherton, personal communication). Filters were then air dried for 15 min.

For V8 protease digestion, fractions off the DEAE-cellulose column containing lamins A and C were pooled and precipitated with 10% (wt/vol) trichloroacetic acid; precipitates were resuspended in 0.2% NaDodSO4/125 mM Tris-HCl, pH 6.8/20 mM dithiothreitol, and proteins were cleaved with V8 protease (Miles) as described (26). Samples were electrophoresed and electroblotted onto glass fiber filters as above. Protein bands were visualized with Coomassie Blue.

The lamin fragments on glass fiber filters were subjected to Edman degradation with the Applied Biosystems model 470A gas-phase sequencer. The phenylthiohydantoin amino acid derivatives were identified and quantitated using the Hewlett Packard ¹⁰⁸⁴ HPLC system.

RESULTS

cDNA Cloning. A full-length lamin C clone (1962 bp) from a human hepatoma library and a partial lamin C clone (1.8 kb) and partial lamin A clone (2.6 kb) from ^a transformed human fibroblast library were obtained. The size of full-length lamin A and C cDNAs, 2.9 kb and 2.0 kb, respectively, was determined by RNA-blot analysis (Fig. 1).

DNA Sequence Analysis. Analysis of the two cloned cDNA sequences of lamin C (from both the fibroblast library and the hepatoma library) (Fig. 2) and the partial cDNA sequence of lamin A (from the fibroblast library) (Fig. 2) revealed complete identity in nucleotide sequences of lamins A and C with the exception of their ³' ends. In addition, the ⁵' end of

FIG. 1. Blot-hybridization analysis of HeLa cell poly(A)+ RNA. A single lane of RNA (25 μ g) was hybridized with ³²P-labeled nick-translated 1.3-kb rat liver cDNA alone (lane 1), with 1.3-kb rat liver cDNA, and then with actin cDNA (lane 2) or with 1.3-kb rat liver cDNA followed by actin cDNA and hsp70 cDNA (lane 3). The actin (2.0-kb) and hsp70 (2.9-kb) probes were used as internal molecular weight markers. Lamin A and C mRNA bands are seen in lane 1, the lamin A mRNA band is still visible in lane 2, but both lamin mRNAs are covered by the two molecular weight probes in lane 3. Similar results were obtained with 32P-labeled oligonucleotide probes (data not shown).

full-length lamin A is identical to that of lamin C (11).

The predicted M_r of the encoded lamin C protein is 65,145, slightly larger than its apparent M_r of 60,000 as estimated by NaDodSO₄/PAGE (3) and in agreement with McKeon et al. (11).

The complete lamin A sequence (Fig. 2) has a predicted M_r of 74,152. This size is in close agreement with the apparept molecular size of ⁷² kDa of the lamin A precursor as estimated by NaDodSO4/PAGE (9) and is about 5.3 kDa smaller than the size predicted by McKeon et al. (11).

Protein Purification and Sequencing. To confirm the identities of the cloned cDNAs, we attempted amino-terminal sequencing of purified rat liver lamin A and C polypeptides and found that both proteins were blocked. By V8 protease digestion we purified a 14-kDa fragment shared by both rat liver lamins. Fifteen amino acids sequenced from this rat liver fragment corresponded to residues 449-463 in human lamins A and C (Fig. 2).

Both lamins were also subjected to mild acid hydrolysis to cleave between Asp-Pro residues. Only one sequence (residues 477-486) was obtained from lamin C as predicted by the nucleotide sequence (Fig. 2). This same sequence was obtained from hydrolyzed lamin A; in addition, a second sequence was obtained from lamin A (residues 576-585) corresponding to the only other Asp-Pro site in the predicted lamin A sequence. The last three residues of this sequence (583-585) (Fig. 2) confirmed the predicted lamin A sequence where it diverged from that of McKeon et al. (11), who reported an additional nucleotide (cytosine) following codon 582.

These protein sequence data clearly established the identity of the two lamin clones.

Lamin Homology to Intermediate Filament Proteins. When the lamin A and C sequences were compared with other known protein sequences, a striking sequence homology was found with all types of intermediate filament proteins. In Fig. 3, the sequences of lamins A and C are compared to that of hamster vimentin (28) (serving as a representative of the family of intermediate filament proteins). An overlap of 366 residues between lamins A and C and of ³¹⁸ residues of vimentin showed 27.6% identity.

Chou and Fasman analysis (30) to predict the secondary structure of lamins A and C revealed ^a 360-residue stretch of α -helical conformation extending from residue 31 to residue 390 (Fig. 2). Most strikingly, this region of predicted α -helical conformation coincides closely with the region of sequence homology between lamins A and C and vimentin (Fig. 3) and that of other intermediate filament proteins (29) (data not shown). An approximately 310-residue central domain of about 30% sequence homology and α -helical conformation is shared by all intermediate filament proteins (reviewed in ref. 31). The lamins contain a repeating heptad sequence within this α -helical domain of the form $(a-b-c-d-e-f-g)_n$ where residues a and d are usually hydrophobic residues (Fig. 2). Three regions within the α -helical domain, containing these heptad repeats, are separated by short stretches of amino acid residues that break up this repeating pattern. The a-helical domain of all intermediate filament proteins contains the same repeating heptad organization as the lamins. The three regions of heptad repeats are labeled Coil la, Coil lb, and Coil 2 in Figs. 2 and 3, since such domains are thought to be capable of forming coiled coils with analogous domains of other molecules (32).

Fig. 3a shows alignment of these coiled coil forming regions between lamins A/C and vimentin with the exception of Coil lb, which is about 50% longer in the lamins than in vimentin. Fig. 3b depicts this structural organization schematically.

The lamin A tail has ^a glycine- and serine-rich stretch of

10
met glu thr pro ser gln arg arg ala thr arg ser gly ala gln ala ser ser thr pro leu ser pro thr arg ile
GCCAACCTGCCGGCC ATG GAG ACC CCG TCC CAG CGG GCG ACC GGC AGC GGG GCG AGC GGC AGC TCC ACT CCG CTG TCG CCC ACC CGC ATC ³⁰ ⁴⁰ . 50 * thr arg ieu gin giu lye giu asp ieu gin glu ,.eU asn asp arg leu ala val toyr lie asp arg voal arg ser leu glu thr giu aesn ACC COG CTG CAG GAG AAG GAG GAC CTG GAG GAG CTC AAT GAT COG TTG GCG GTC TAC ATC GAC CGS, GTG CGC TCG CTG GAA AGO GAG AAC $~\sim~$ 10 $~\sim~$ and the property and argument of the serves of the ser ala gly leu arg leu arg ile thr glu ser glu glu val val ser arg glu val ser gly ile lys ala ala tyr glu ala glu leu gly
GCA GGG CTG CGC CTT CGC ATC ACC GAG TCT GAA GAG GTG GTC AGC CGC GAG GTG TCC GGC ATC AAG GCC GCC TAC GA CCA GGG CTG CGC CTT CGC ATC ACC GAG TCT GAA GAG GTG GTC AGC CGC GAG GTG TCC GGC ATC AAG GCC GCC TAC GAG CTC GGG
As an all a sep 198 thr 180 asp ser values and a separate and arg 180 gin leu glu 180 ser lys values and lyse 130
and 120 set of the U.S. 110
and CCC COC ANT ACC AAG AAG GAG GGT GAG CTG ATA GCT CGT CAG GCT CGG CTG AAG GAC CTG GAG GCT CTG CTG AAC TCG CTG
ala leu ser glu 139 serg thr 120 glu gly glu 120 his asp 120 gly gln value is 150
ala leu ser that a leu ser glu log arg thr 18u glu glu leu his asp leu arg gly gln val als leu glu ala ala leu
GCA CTG AGC ACT GCT CTC AGT GAG AAG CGC ACG CTG GAG GGC GAG CTG CAT GAT CTG CGG GGC CAG GTG GCC AAG CTT GAG COT GAO 0CC AAG AAG CAA CTT CGA OAT GAG ATG GTG COG COG GTG GAT OCT GAG AAC ACO CTG GAG ACG ATG AAG GAG GAA CTG GAG TTC 220
glu du leu arg glu thr lys arg arg his glu thr arg leu val glu ile asp asn gly lys gln arg glu
GAG GAG CTG CGT GAG ACC AAG CGC CGT CAT GAG ACC CGA CTG GTG GAG ATT GAC AAT GGG AAG CAG CGT GAG gin lys asn ile tyr ser glu giu leu arg glu thr lys arg arg his glu thr arg leu val glu ile asp asn gly lys gln arg glu
CAG AAG AAC ATC TAC AGT GAG GAG CTG CGT GAG ACC AAG CGC CGT CAT GAG ACC CGA CTG GTG GAG ATT GAC AAT GG Coil 2 240
phe glu ser arg leu als asp als ales and a leu arg als gln his glu asp gln val glu gln type glu leu glu leu glu
TTT GAG AGC CGG CTG GGG GAT GGG CTG CAG GAA CTG CGG GCC CAG CAT GAG GAC CAG GTG GAG CAG TAT AAG GAG CTG GA 230
Evasp as nata arg gin ser alla glan ser as heuval gly ala alla his glu glu leu gin gin ser arg and the gin gin
TAT TCT GCC AAG CTG GAC AAT GCC AGG CAG TCT GCT GAG AGG AAC AGC AAC CTG GTG GGG GCT GCC CAC GAG GAG CTG CAG The arg are seen to the see th The same and the same of the s . ³⁶⁰ * ⁰ . ³⁷⁰ 3& asep glu tyr gin gin leu leu asp ils lye leu ala leu asp met giu jim his ale toyr erg lye eun e giu gly ? As a la arg of the set arg and level to the set arg and the set arg of the set arg and arg of the set arg and the set arg and the set are of the set are of the set are of the set are set are of the set are set are conside 240

The control of the state of the sta CG GC AC AG CG GC CT GA AMGAGATC AC CC AC GC AG TCTO GAG GOG GAO GAG GAG ACO 390

390 and the set of the service service service of the service Tys are the given the given are ser of the given and the ser given and the ser given and the ser given and the ser given and the method on both strands of all the chemical method on the serve and the ser given are one conc $\frac{1}{1/5}$ $\frac{1}{1/5}$ $\frac{120}{1/5}$ $\frac{120}{1/5}$ ser thr glu ser arg ser ser phe ser gln his ala arg thr ser gly arg val ala val glu glu val asp **nation method on both Strands Of a**
AAA AAG CSC AAA CTG GAG ACC CTG GAG ^P ^L ^M ^T ^Y ^X ^F ^P ^X ^K ⁴⁹⁰ 500kb weesq ncda ^w sth pril $\frac{1}{2}$ is the thr leu lys ala gly gln val val thr ile trp ala gly ala gly ala thr his lamin A cDNA clone (2.6 kb). The pro lys pha thr his control ala gly ala thr his lamin A cDNA clone (2.6 kb). The change of the stat ser pro pro thr asp leu val trp lys ala gln asn thr trp gly cys gly asn ser leu arg thr ala leu ile asn ser thr gly glu
AGC CCC CCT ACC GAC CTG GTG TGG AAG GCA CAG AAC ACC TGG GGG CGGG AAC ACC CGGG CCT CTC CTCC ATC CACC TG 560 CDNA sequence, and residue 566
glu val als met arg lys leu val arg ser val thr val val glu asp asp glu asp glu asp gly asp asp leu leu his his his his bis (closed arrowhead) is the last amino
GAA GTG GCC ATG GCC ATG GC GAA GTG GCC ATG CGC AAG CTG GTG GCC TCA GTG ACT GTG GTT GAG GAC GAC GAG GAT GAG GAT GAG ATT GAC CTG CAT CAC CAC

570 572 **before their sequences diverge. Coils**

val ser gly ser arg arg op

and 2 denote a-helical domains VALUE AGT AGT AGT COG COG TGA GGCGGAGCCTGCACTGGGGCAGGGAGGGAGGGAGCGGGGGGGAGCCTCCCGGAANATCTTTTGATTAMAGAATGTTTTG 1a, 1b, and 2 denote a-helical domains vith heptad repeats capable of forming

arg lys leu val arg ser val thr val du val du asp asp glu asp glu asp gly asp leu leu his his his his his **mark hydrophobic and nonpolar resi-**
CGC AAG CTG GTG GTG GTG ACT GTG GTT GAG GAC GAC GAC GAC GA GAT GGA GAT GAC CTG 570 **P A E Y N L R S X T**
gly ser his cys gly thr cys gly thr cys gly throw ala asp **idues. Open boxes (residues 267 and**
gac rece car cyc arc as asp pro ala glu tyr as also arc cyclose throw the cys given by the cys give gly ser his cys ser ser ser gly asp pro ala glu tyr asn leu arg ser arg thr val leu cys gly thr cys gly gln pro ala asp
GGC TCC CAC TGC AGC TCG GGG GAC CCC GCT GAG TAC AAC CTG CGC TCG CGC GCG CTGC GGG GGC GGGCGCGGG CACC GA gln val gly gly pro ile ser ser gly ser ser ala ser cad thr val thr val thr val thr ear ear thr ear thr ear thr e AAG OGA TCT 0CC AGCGGCO TCA OGA 0CC GAG GOG GCOGOGA CCC ATC TCC TOT GCO TGT TCT 0CC TCG AGT OTG AGO GTC ACT COG AOG TAG ⁱ cdhdoye ^a ^a ⁱ ⁿ Ser gly ser gly ala gln val gly gly pro ile ser ser gly ser ser ala ser ser val thr val thr arg ser tyr
AGC GGC TCA GGC CAG GTGC GGG GGA CCC ATC TCC TCT GGC TCT TCT GCC TCC AGT GTC AGT GTC AGC TACC AGC TAC
gly ser gly gly So ser all the ser ser all the ser ser and ser ser all the ser ser and the ser ser all the ser ser was the contract the sequences. Amino acid residues and for the ser ser was the service of the service of the service of th AGC CCC CGA AAC TOG AGC ATO ATG; TAA TCTGGGACCTGCCGGCOAGGGGTGGGGGOGGAGCOTTCCTGCGTCCTCCTOACCTCATGCCCACCCGCGTOGCTOGACGTGATO human cDNA sequences. "X" denotes $\begin{minipage}{16cm} \textbf{GAGTCGAAAAGGGAAAAGGTGAGTGTTGAGCTCTCCTAGCTTTTAGCCTGGGTGGGTGGAGGGTGGAGTGGAGGGTTGGGATGGGGTGGCGGGTGGCTGCCTGGGGGGGGAGGG & \textbf{CATDOXyl} and of the lamin & \textbf{GATDOXyl} domain of the lamin & \textbf{GATDOX} of the lamin & \textbf{GATDAX} and the form \textbf{GATDAX} is the \textbf{GATDAX} of the lamin & \textbf{GATDAX} is the \textbf{GATDAX} of the lamin & \textbf{GATDAX} is the \textbf{GATDAX} of the lamin & \textbf{GATDAX$ __TGTGC __TGTTTGGAAAGprotein)CTGCCCareA highlyin ar acidic.iic C!!ACT-ACCTTTCCCAACAAATGTTT

quence homology with amino- and carboxyl-terminal por-
tions of a human type II epidermal keratin (67 kDa) (33) (Fig. Nuclear Localization Site. Lamins A and C contain a tions of a human type II epidermal keratin (67 kDa) (33) (Fig. Nuclear Localization Site. Lamins A and C contain a
4) with a mouse type II keratin (60 kDa) (36), and with mouse sequence of 6 amino acids (Lys-Lys-Arg-Lys-Le 4), with a mouse type II keratin (60 kDa) (36), and with mouse

performed by the dideoxy chain-termilamin C cDNA clones $(1.8 \text{ kb}$ and 1.9 k kb) were sequenced as was the partial 3'-polvadenvlylation signals are under-530 lined. Residue 150 is the first deduced
act even we see the gly g_{AA}^{11} amino acid of the incomplete lamin A with heptad repeats capable of forming
coiled coils. At the first and fourth dues, and open circles mark polar resheptad sequences. Amino acid residues
the valid the range serior sequenced from V8-digested and forfragments are indicated by single letter amino acid symbols above the corre-
sponding amino acids predicted by the amino acids not detected. A stretch of six amino acid residues (residues 417-422, broken line), similar to the nuclear localization signal of the simian virus 40 large T (tumor) antigen (27), is also indicated. Residues 551-560 at the within the carboxyl domain of the lamin
A protein) are highly acidic.

residues (571–652) that shares approximately 20%-30% se- (59 kDa) and human (52 kDa) type I keratins (34, 35) (data not quence homology with amino- and carboxyl-terminal por- shown).

selected regions. Both of the isolated
 $\frac{470}{10}$ served as the strategies of $\frac{470}{10}$ serv Lamin A Carboxyl Terminus

Lamino access the set of the set

coiled coils. At the first and fourth

the head repeates divergences divergences

So Umin A Carboxyl Terminus

Lamin A Carboxyl Terminus

the space of the heptads, closed circles

the first and fourth

positions of the heptads, closed circles

mark hydrophobic and nonpolar resi-

mark hydrophobic and nonpo

		10	20	30	Coil 1a	50
LAMINS					METPSQRRATRSGAQASSTPLSPTRITRLQEKEDLQELNDRLAVYIDRVR	
<u>VIMENTIN</u>		80	90	100	SMPGVRLLQDSVDFSLADAINTEFKNTRTNEKVELQELNDRFANYIDKVR Δ 110	120
60		70	Coil 1b $80\frac{1}{2}$	90	100 SLETENAGLRLRITESEEVVSREVSGIKAAYEAELGDARKTLDSVAKERARLQLELSKVR	110
re de	130	Δ 140	4250	160	a sa totala a a conto a cittata cantana e a a a a a teatra a gall FLEQQNK---ILLAELEQLKGQGKSRLGDLYEEEMRELRRQVDQLTNDKARVEVERDNLA 170	
120		130	140	150	160 EEFKELKARNTKKEGDLIAAQARLKDLEALLNSKEAALSTALSEKRTLEGELHDLRGQVA	170
<i>Execution</i> EDIMRLREK-- 180						$2.11 - 1.1$ -LOEEMLORE 190
180 KLEAALGEAKKQLQDEMLRRVDAENRLQTMKEELDFQKNIYSEELRETKRRHETRLVEID		190	200	210	-20	230
EAESTLQSFRQDVDNASLARLDLERKVESLQEEIAFLKKLHDEEIQELQAQIQEQHVQID 200	210	220	230		240 Δ	. 250
240 NGKQREFESRLAĎALQELRAQHEDQVEQYKKELEKTYSAKLDNARQSAERNSNLVGAAHE	Coil	2_{50}	260	270	280	290
VDVSK---PDLTAALRDVRQQYESVAAKNLQEAEEWYKSKFADLSEAANRNNDALRQAKQ 260	270		280	290	300	310
300 ELQQSRIRIDSLSAQLSQLQKQLAAKEAKLRDLEDSLARERDTSRRLLAEKEREMAEMRA		310	320	330	340	350
t av travitis solitar i svi tropolitisk tropis solitar ESNEYRRQVQSLTCEVDALKGTNESLERQMREMEENFALEAANYQDTIGRLQDEIQNMKE 320	330		350 340		360	Bandar 370
360 RMQQQLDEYQELLDIKLALDMEIHAYRKLLEGEEERLRLSPSPTSQRSRGRASSHSSQTQ		370	380	₩0	400	410
EMARHLREYQDLLNVKMALDIEIATYRKLLEGEESRISLPLPNFSSLNLRETNLESLPLV 380	390		400	410	420	430
LAMIN	NH 2	1b			соон	
LAMIN						
VIMENTIN NH 2				соон		

FIG. 3. Amino acid sequence homology and secondary structural homology between lamins A/C and vimentin. (a) The deduced amino acid sequence common to lamins A and C was aligned to maximize homology with that of hamster vimentin (28). The lamins shared 27.6% identity with vimentin over a 366-amino acid residue overlap. A gap of 42 residues was introduced into the vimentin sequence to optimize this homology. This region of sequence homology coincides with predicted α -helical domains capable of forming coiled coils present in lamins A/C (between closed arrowheads) and vimentin (between open arrowheads). The first 20 and last 35 amino acids within this stretch each show greater than 70% sequence homology. (b) Amino, carboxyl, and coiled coil domains of lamins A and C were aligned with the predicted secondary structure of hamster vimentin (29). All molecules are drawn to scale. Lamins A and C have identical sequences spanning the central α -helical domain (black lines) and the carboxyl domain of lamin C. The lamin C carboxyl domain extends 6 amino acids beyond the point of divergence from lamin A whereas the lamin A carboxyl domain extends 98 amino acids further (see Fig. 2). Putative globular domains are located at both ends of all three molecules. The central α -helical domains are 14% longer in lamins A/C than in vimentin. Coils 1a and 2 and the two intercoil linkers are conserved in size between the lamins and vimentin whereas Coil 1b is about 50% longer in the lamins than in vimentin. Vertical lines mark stutters in the heptad repeats.

is similar to a stretch of amino acids in the simian virus 40 large T (tumor) antigen (Lys-Lys-Arg-Lys-Val-Glu) that contributes to nuclear localization of the viral protein (27). This sequence is located approximately 30 amino acids outside the carboxyl end of the α -helical domain (Fig. 2).

FIG. 4. Lamin A sequence homology to human type II (67 kDa) keratin head and tail domains. A 64-amino acid residue stretch (residues 589-652) in the carboxyl-tail region of lamin A, not present in lamin C (see Fig. 2), shows 20.3% and 26.6% identity with head and tail regions, respectively, of human type II (67 kDa) epidermal keratin (33). Similar homology is found to the head and tail regions of human type I (52 kDa) keratin (34) and mouse type I (59 kDa) and type II (60 kDa) keratins (35, 36) (data not shown).

DISCUSSION

Lamins A and C share an α -helical region of 360 amino acid residues that in turn shows striking homology with the canonical 310-residue α -helical rod domain characteristic of all intermediate filament proteins (Fig. 3) (11). Based on this sequence homology and the similar structural organization of the molecules (Fig. 3), lamins A and C belong to the family of intermediate filament proteins.

In spite of the structural relatedness between the lamins and the intermediate filament proteins, there are some noteworthy differences. The lamin α -helical domain is approximately 14% larger than the rod domain of intermediate filament proteins due to a longer Coil 1b domain (Fig. 3). Furthermore, in mitosis the lamins undergo complete and reversible depolymerization involving transient hyperphosphorylation whereas the intermediate filaments apparently do not. The large number of serine residues outside the α -helical domain and in the lamin A tail (Fig. 2) may be substrates for such kinase activity. Moreover, although both lamins and intermediate filament proteins belong to developmentally regulated gene families (reviewed in refs. 37 and 38), different forms of lamins appear to be expressed in nondividing versus dividing cells, whereas different intermediate filament proteins are expressed in cells depending on their embryonic origin.

Lamin A contains a 9.0-kDa carboxyl-terminal domain not present in lamin C (Fig. 2). McKeon et al. (11) have reported this domain to be 14 kDa. However, amino acid sequence data of rat lamin fragments have confirmed our predicted sequence in this region (Fig. 2). Furthermore, the lamin A cDNA contains a 977-bp 3'-untranslated region of which 762 bp were not present in the reported cDNA clone (11). Whether the precursor sequence of lamin A is part of the lamin A carboxyl-terminal domain is unknown. This domain contains many glycine and serine residues and shows homology to type I and type II keratins (Fig. 4 and data not shown). Unlike the keratins, however, this lamin stretch does not contain a tandemly repeating pattern of Gly-Gly-Gly-Xaa. The functional significance of this domain remains to be determined.

The lamin A cDNA sequence diverges from that of lamin C following a CAC codon corresponding to amino acid residue 566 (Fig. 2). Interestingly, the following six nucleotides in the lamin C sequence, GTGAGT, are identical to the consensus sequence of the 5' end of introns (39). Also, the following two nucleotides in the lamin A sequence G-G represent a consensus sequence of the 5' end of spliced exons

(39). In combination with evidence from Southern analysis that there may be only one gene coding for the two lamins (11), these sequence data suggest that the two lamins may arise from alternate forms of splicing. At the ³' end of the gene transcript, the sequences containing the lamin C polyadenylylation site immediately downstream from codon 566 may be excised to splice a further downstream sequence to codon ⁵⁶⁶ to yield lamin A mRNA, or they may remain to yield lamin C mRNA.

The striking homology of lamins A and C with intermediate filament proteins suggests that at least some aspects of their assembly into a lamina may resemble that of intermediate filament proteins into intermediate filaments. Thus, coiled coil homodimers of lamins A and C may be formed by parallel interactions of their α -helical region. Such dimers might form rod-like structures with two smaller globular domains (comprising the amino terminus) at one end and two larger globular domains (comprising the carboxyl terminus) at the other end. Electron microscopy of dimers of the lamins from rat liver (U. Aebi, J. Cohn, L. Buhle, and L. Gerace, personal communication) shows that they contain two globular heads attached to an approximately 50-nm rod-like tail. Based on the sequence data, these globular head domains probably correspond to the carboxyl-terminal portions of the lamin polypeptides, which are larger than the amino-terminal portions.

How lamins A and C interact with lamin B to form ^a near-tetragonal lattice of 8- to 10-nm filaments (U. Aebi, J. Cohn, L. Buhle, and L. Gerace, personal communication) remains to be determined. As lamin B has a stronger physical association with the nuclear membrane (5), an additional domain present in lamin B may anchor the lamina to the inner nuclear membrane. However, it remains to be determined whether lamin B is also an intermediate filament-type protein.

We thank Ms. Donna Atherton of the Rockefeller University Protein Sequencing Facility for determining the partial sequence of V8 protease fragments and acid hydrolysis fragments of lamins A and C, and Ms. Beth Goldstein for the synthesis of oligonucleotides. D.Z.F. is a predoctoral trainee supported by Training Grant GM07982 from the National Institutes of Health. N.C. was supported by the Damon Runyon-Walter Winchell Postdoctoral Fellowship DRG752. This work was supported by Grant GM27155 from the National Institutes of Health.

- 1. Aaronson, R. & Blobel, G. (1975) Proc. Natl. Acad. Sci. USA 72, 1007-1011.
- 2. Dwyer, N. & Blobel, G. (1976) J. Cell Biol. 70, 581–591.
3. Gerace, L., Blum, A. & Blobel, G. (1978) J. Cell Biol.
- 3. Gerace, L., Blum, A. & Blobel, G. (1978) J. Cell Biol. 79, 546-566.
- 4. Krohne, G., Franke, W., Ely, S., ^D'Arcy, A. & Jost, E. (1978) Cytobiologie 18, 22-38.
- 5. Gerace, L. & Blobel, G. (1980) Cell 19, 277-287.
- 6. Gerace, L. & Blobel, G. (1982) Cold Spring Harbor Symp.
- 7. Burke, B., Tooze, J. & Warren, G. (1983) EMBO J. 2, 361-367.
- 8. Krohne, G., Debus, E., Osborn, M., Weber, K. & Franke, W. (1984) Exp. Cell Res. 150, 47-59.
- 9. Gerace, L., Comeau, C. & Benson, M. (1984) J. Cell Sci., Suppl. 1, 137-160.
- 10. Laliberte, J., Dagenais, A., Filion, M., Bibor-Hardy, V., Simard, R. & Royal, A. (1984) J. Cell Biol. 96, 980-985.
- 11. McKeon, F., Kirschner, M. & Caput, D. (1986) Nature (London) 319, 463-468.
- 12. Young, R. A. & Davis, R. W. (1983) Science 222, 778-782.
13. Schwarzbauer, J. E., Tamkun, J., Lemischka, I. & Hynes, I
- Schwarzbauer, J. E., Tamkun, J., Lemischka, I. & Hynes, R. (1983) Cell 35, 421-431.
- 14. Grunstein, M. & Hogness, D. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
- 15. Okayama, H. & Berg, P. (1983) Mol. Cell. Biol. 3, 280–289.
16. Benton, W. D. & Davis, R. (1977) Science 196, 180–182.
- 16. Benton, W. D. & Davis, R. (1977) Science 196, 180-182.
17. Chirewin, J., Przybyła, A., MacDonald, R. & Rutter.
- 17. Chirgwin, J., Przybyla, A., MacDonald, R. & Rutter, W. (1979) Biochemistry 18, 5194-5199.
- 18. Aviv, H. & Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
- 19. Thomas, P. S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
- 20. Messing, J. (1983) Methods Enzymol. 101, 20-78.
- 21. Sanger, F., Nicklen, S. & Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 22. Biggin, M., Gibson, T. & Hong, G. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
- 23. Dale, R., McClure, B. & Houchins, J. (1985) Plasmid 13, $31 - 40$.
- 24. Maxam, A. & Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
- 25. Aebersold, R., Teplow, D., Hood, L. & Kent, S. (1986) J. Biol. Chem. 261, 4229-4238.
- 26. Cleveland, D., Fischer, S., Kirschner, M. & Laemmli, U. (1977) J. Biol. Chem. 252, 1102-1106.
- 27. Kalderon, D., Roberts, B., Richardson, W. & Smith, A. (1984) Cell 39, 499-509.
- 28. Quax, W., Egberts, W., Hendriks, W., Quax-Jeuken, Y. & Bloemendal, H. (1983) Cell 35, 215-223.
- 29. Geisler, N., Fischer, S., Vandekerckhove, J., Plessmann, U. & Weber, K. (1984) EMBO J. 3, 2701-2706.
- 30. Chou, P. Y. & Fasman, G. D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45-148.
- 31. Fuchs, E. & Hanukoglu, I. (1983) *Cell* 34, 332-334.
32. Crick, F. H. (1953) *Acta Crystallogr.* 6, 689–697.
- 32. Crick, F. H. (1953) Acta Crystallogr. 6, 689-697.
33. Johnson, L., Idler, W., Zhou, X., Roop, D. &
- 33. Johnson, L., Idler, W., Zhou, X., Roop, D. & Steinert, P. (1985) Proc. Natl. Acad. Sci. USA 82, 1896-1900.
-
- 34. Hanukoglu, I. & Fuchs, E. (1982) Cell 31, 243-252.
35. Steinert, P., Rice, R., Roop, D., Trus, B. & Steven. 35. Steinert, P., Rice, R., Roop, D., Trus, B. & Steven, A. (1983) Nature (London) 302, 794-800.
- 36. Steinert, P., Parry, D., Racoosin, E., Idler, W., Steven, A., Trus, B. & Roop, D. (1984) Proc. Natl. Acad. Sci. USA 81, 5709-5713.
- 37. Krohne, G. & Benavente, R. (1986) Exp. Cell Res. 162, 1–10.
38. Franke, W., Schmid, E., Schiller, D., Winter, S., Jarasch, E.,
- 38. Franke, W., Schmid, E., Schiller, D., Winter, S., Jarasch, E., Moll, R., Denk, H., Jackson, B. & Illmensee, K. (1982) Cold Spring Harbor Symp. Quant. Biol. 46, 431-453.
- 39. Cech, T. R. (1983) Cell 34, 713-716.