Sites of Nucleic Acid Binding in Type I−IV Intermediate Filament Subunit Proteins†

Qiang Wang,‡ Genrich V. Tolstonog, Robert Shoeman,* and Peter Traub*

Max-Planck-Institut für Zellbiologie, Rosenhof, 68526 Ladenburg, Germany

Received April 23, 2001; Revised Manuscript Received June 20, 2001

ABSTRACT: A combination of enyzmatic and chemical ladder sequencing of photo-cross-linked protein−single-stranded oligodeoxyribonucleotide complexes and analysis by MALDI-TOF mass spectrometry was employed to identify the amino acid residues responsible for the stable binding of nucleic acids in several intermediate filament (IF) subunit proteins. The IF proteins studied included the type I and type II cytokeratins K8, K18, and K19; the type III proteins desmin, glial fibrillary acidic protein (GFAP), peripherin, and vimentin; and the type IV neurofilament triplet protein L (NF-L). The site of nucleic acid binding was localized to the non-α-helical, amino-terminal head domain of all of the IF proteins tested. GFAP, which has the shortest head domain of the proteins tested, cross-linked via only two amino acid residues. One of these residues was located within a conserved nonapeptide domain that has been shown to be required for filament formation. One or more cross-linked residues were found in a similar location in the other proteins studied. The major binding site for nucleic acids for most of the proteins appears to be localized within the middle of the head domain. The two exceptions to this generalization are GFAP, which lacks these residues, and NF-L, in which a large number of cross-linked residues were found scattered throughout the first half of the head domain. Control experiments were also done with two bacteriophage ssDNA-binding proteins, as well as actin and tubulin. The single sites of cross-linkage observed with the bacteriophage proteins, Phe_{183} for the T4 gene 32 protein and Phe_{73} for the M13 gene 5 protein, were in good agreement with literature data. Actin and tubulin could not be cross-linked to the oligonucleotide. Aside from the insight into the biological activity of IF proteins that these data provide, they also demonstrate that this analytical method can be employed to study a variety of protein−nucleic acid interactions.

Intermediate filaments (IFs)[1] are ubiquitous substructures of vertebrate cells and, together with microfilaments and microtubules, are the major structural elements of the cytoskeleton (*1*, *2*). Unlike the two other filamentous networks, which are composed of only one or a few isoforms of identical subunit proteins in all of the cells of an organism, IFs are assembled from one or more members of a large family (more than 50 proteins to date in humans) of subunit proteins in a developmental and tissue-specific fashion. This fact has led to the proposal that the IFs or their subunit proteins may function in other roles (*3*) in addition to their cytoskeletal functions (*1*). While a detailed discussion is beyond the scope of this paper, it is important to note that all of the IF subunit proteins possess a highly conserved, relatively long (310−350 aa), central, α-helical rod domain responsible for filament formation and that the major differences seen among and between the proteins belonging to the various types of IF subunit proteins are found within the non-α-helical head and tail domains (*1*). Since it has been previously shown that the type III IF subunit protein vimentin binds to single-stranded nucleic acids via two imperfect repeats of a *β*-ladder DNA-binding wing located in the middle of its head domain (*4*, *5*) and since other IF subunit proteins such as desmin and glial fibrillary acidic protein (GFAP) selectively bind nucleic acids in a fashion similar but not identical to vimentin (*6*), we decided to employ a method (*5*) recently developed for determining the sites of DNA binding in vimentin to a panel of seven other IF proteins to see whether any common determinants in DNA-binding sites exist. Furthermore, this method was additionally employed to investigate the cross-linkage of DNA to control proteins, including the bacteriophage T4 gene 32 protein and the bacteriophage M13 gene 5 protein, both of which, like IF proteins, bind to single-stranded DNA (ssDNA).

EXPERIMENTAL PROCEDURES

Materials. Recombinant mouse desmin, mouse NF-L protein, mouse peripherin, and mouse vimentin were produced and purified essentially as previously described for vimentin (*4*, *5*). An expression plasmid for mouse desmin was produced using conventional PCR and cloning methods;

† A part of these studies was performed in partial fulfillment of the requirements for a Ph.D. degree at the University of Heidelberg (Q.W.).

* Corresponding authors. Tel: +49-6203-106-200. Fax: +49-6203-106-122. E-mail: rshoeman@zellbio.mpg.de; ptraub@zellbio.mpg.de.

‡ Present address: T. H. Morgan School of Biological Sciences, University of Kentucky, Lexington, KY 40506-0225.

[1] Abbreviations: DABITC, dimethylaminoazobenzene isothiocyanate; FITC, fluorescein isothiocyanate; GFAP, glial fibrillary acidic protein; HPLC, high-pressure liquid chromatography; IF, intermediate filament; MALDI TOF, matrix-assisted laser desorption/ionization time of flight; MAR, matrix attachment region; *m*/*z*, mass/charge; NF-L, neurofilament triplet protein L; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; ssDNA, single-stranded DNA.

10.1021/bi0108305 CCC: $20.00
Published on Web 08/01/2001

Table 1: Summary of the Results of the Cross-Linking of a BrdU-Containing Oligonucleotide to IF Subunit Proteins in Vitro[a]

Vimentin, mouse:
STRSVSSSSY RRM**F**$_{14}$GGSGTS SRPSSNRS**Y**$_{29}$V TTSTRT**Y**$_{37}$SLG SALRPSTSRS
L**Y**$_{52}$SSSPGGAY VTRSSAVRLR SSVPGVRLLQ DSVDFSLADA INTEFKNTRT NE*KVELQELN*
DRFANY

Desmin, mouse:
MSQAYSSSQR VSS**Y**$_{14}$RRT**F**$_{18}$GG APG**F**$_{24}$SLGSPL SSPV**F**$_{35}$PRAGF GTKGSSSSMT
SRV**Y**$_{54}$QVSRTS GGAGGLGSLR SSRLGTTRAP SYGAGELLDF SLADAVNQEF LATRTNE*KVE*
LQELNDRFAN Y

Peripherin, mouse:
MPSSASMSHH HSSGLRSSIS STSYRRT**F**$_{28}$GP PPSLSPGA**F**$_{39}$S YSSSSR**F**$_{47}$SSS
RLLGSGSPSS SARLGS**F**$_{67}$RAP RAGALRLPSE RLD**F**$_{84}$SMAEAL NQEFLATRSN E*KQELQELND*
RFANF

GFAP, pig:
MERRRVTSAA RRS**Y**$_{14}$VSSLET VGGGRRLGPG PRLSLARMPP PLPARVD**F**$_{48}$SL
AGALNTGFKE TRASE*RAEMME LNDRFASY*

NF-L, mouse:
GS**F**$_{3}$G**Y**$_{5}$DP**Y**$_{8}$**F**$_{9}$S TS**Y**$_{13}$KRR**Y**$_{17}$VET PRV**H**$_{24}$ISSVRS G**Y**$_{32}$STARSA**Y**$_{39}$S
S**Y**$_{42}$SAPVSSSL SVRRSYSSSS GSLMPSLENL DLSQVAAISN DLKSIRTQEK
AQ*LQDLNDR***F**$_{100}$ *ASFIERVHEL*

Keratin K8, human:
MSIRVTQKSY KVSTSGPRA**F**$_{20}$ SSRS**Y**$_{25}$TSGPG SRISSSS**F**$_{38}$SR VGSSN**F**$_{46}$RGGL
GGGYGGASGM GGITAVTVNQ SLLSPLVLEV DPNIQAVRTQ *EKEQIKTLNN KFASFIDKVR*

Keratin K18, human:
MSFTTRST**F**$_{9}$S TN**Y**$_{13}$RSLGSVQ APSYGARPVS SAASVYAGAG GSGSRISVSR
STS**F**$_{54}$RGGMGS GGLATGIAGG LAGMGGIQN*E KETMQSLNDR LASYLDRVRS LETENRRLES*

Keratin 19, human:
MTSYSYRQSS ATSS**F**$_{15}$GGLGG GSVR**F**$_{25}$GPGVA **F**$_{31}$RAPSIHGGS GGRGVSVSSA
RFVSSSSSGG YGGGYGGVLT ASDGLLAGN*E KLTMQNLNDR LASYLDKVRA LEAANGELEV*

[a] Those residues found to be cross-linked to dUMP are indicated by bold font. Selected residue positions are indicated by subscript numbers. Those residues within the rod domain are indicated by italic font. Underlined residues indicate those found at a molar ratio of ≤15% of the other cross-linked residues. The molar yield of the cross-linkage at Phe$_{14}$ in vimentin was very low and was missed in an earlier study (*5*). For vimentin, the cross-linkage to Phe$_{14}$ was observed only in complexes formed in the absence of added salt, whereas the other cross-linkages (also with the other proteins) were observed at 75−150 mM added NaCl.

the expression plasmids for mouse peripherin and NF-L were kind gifts of Drs. W. E. Mushynski (McGill University, Montreal, Quebec, Canada) and F. Beaumont-Landon (CNRS, Paris, France), respectively. Porcine GFAP was isolated from spinal cord as described (*7*). Human cytokeratins (a mixture of keratins K8, K18, and K19) were isolated from human breast carcinoma MCF-7 cells, as described (*8*). The protein sequences were obtained from the National Center for Biotechnology Information Entrez browser (http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Protein) under the following accession numbers: desmin, P31001; GFAP, 1003182A; keratin K8, P05787; keratin K18, P05783; keratin K19, P08727; NF-L, P08551; peripherin, P15331; and vimentin, P20152. The primary sequences of the recombinant proteins are identical to those of the native proteins, except for NF-L. As a result of the cloning procedure employed (and confirmed by DNA sequencing), the recombinant NF-L protein had a Gly residue at position 2, rather than the Ser$_2$ characteristic of the native mouse protein. The amino acid residue numbering scheme in Table 1 reflects the fact that the N-terminal formyl-Met was observed to be removed from the recombinant vimentin and NF-L proteins. The bacteriophage T4 gene 32 protein was purchased from Amersham Pharmacia Biotech (Freiburg, Germany), and the bacteriophage M13 gene 5 protein (wild type) was a kind gift of Dr. C. W. Hilbers, University of Nijmegen, The Netherlands. Bovine tubulin and human actin were purchased from Cytoskeleton (supplied by TEBU, Frankfurt, Germany), BSA was from Sigma-Aldrich (Deisenhofen, Germany), and sequencing grade chymotrypsin was from Roche Molecular Biochemicals (Mannheim, Germany). Other materials were as described (*5*).

Methods. Analytical and preparative scale photochemical cross-linking reactions were performed exactly as previously described for vimentin (*5*), except for the cytokeratins. This protein preparation, which consists of a mixture of the cytokeratins K8, K18, and K19 reflective of their individual concentrations in the MCF-7 cells, was first mixed with the oligonucleotide in the presence of 6 M urea. The urea was removed by dialysis, and the clear solution was employed for photochemical cross-linking, as per the standard protocol. This step was necessary to prevent the formation of undefined keratin aggregates. Unless otherwise indicated, all cross-linking experiments were performed in the presence of 75−150 mM NaCl. For experiments of cross-linking specificity, additional proteins, unlabeled oligonucleotides, salts, or

polyanions were either altered in concentration (salts) or added to the vimentin−oligonucleotide reaction mixtures prior to illuminating with UV light. The remaining methods, digestion with trypsin, isolation of oligonucleotide cross-linked peptides, enzymatic and manual Edman degradation, peptide ladder sequencing, and MALDI-TOF analysis using a Kratos Maldi IV instrument (Shimadzu Deutschland, Duisburg, Germany), were performed exactly as described previously for vimentin (*5*). The determination of the degree of cross-linking to any one amino acid residue was based on a comparison of the total OD_{214nm} of the individual peptides obtained from any one protein. In some cases, the peptide−dUMP complexes were additionally treated with chymotrypsin. Specific, purified tryptic peptide−dUMP complexes, derived from 5.4 mg of starting protein material, were incubated for 1 h at 30 °C together with 1 μg of chymotrypsin in 100 mM Tris-HCl, pH 7.4. The chymotryptic/tryptic peptide−dUMP complexes were purified by HPLC and analyzed as described for the tryptic peptide−dUMP complexes.

Computer predictions of secondary structure either were done as previously described (*9*) or were performed using the World Wide Web-based program Jpred[2], which can be accessed at http://jura.ebi.ac.uk:8888/index.html.

RESULTS

The procedures previously employed for vimentin (*5*) worked quite well for all of the proteins studied (see Table 1 and below), with the exception of the cytokeratins. Since these proteins aggregate or polymerize into IFs even at low ionic strength, it was necessary to mix the oligonucleotide [oligo(dG·BrdU)$_{12}$·dG·3′-FITC] together with the cytokeratins in 6 M urea and then to remove the urea by dialysis prior to cross-linking with UV light. The binding of the oligonucleotide to the cytokeratins prevented IF polymerization or protein aggregation, since the solution viscosity remained low. If the urea was removed prior to the addition of the oligonucleotide, the cytokeratin mixture polymerized/aggregated into an intractable gel. The total amount of IF protein (about 100 nmol) utilized in the large-scale cross-linking reactions yielded about 10−15 nmol of protein cross-linked to the oligonucleotide (*5*; data not shown). Although this was far more material than that required for the subsequent steps [since the MALDI-TOF analysis requires only a few picomoles of analyte (or less for ideal samples)], it provided a comfortable excess of peptide−oligonucleotide complexes to permit repeat analyses, multiple peptidase treatments, etc.

The derived masses of the 41 tryptic peptide−dUMP complexes generated from the 10 proteins studied were in good agreement with those predicted on the basis of the known amino acid sequences and the cleavage specificity of trypsin as -Arg-↓-X- or -Lys-↓-X- (where X is not Pro), except for two peptides derived from peripherin, one from vimentin, and two partially digested peptides, one each from GFAP and the T4 gene 32 protein. For GFAP, roughly equal amounts of $GFAP_{11-25}$−dUMP and $GFAP_{12-25}$−dUMP complexes were found, as a result of incomplete cleavage at the Arg_{11}−Arg_{12} bond in the -Arg_{11}-Arg_{12}-Ser_{13}-Tyr_{14}−dUMP- sequence. A trace amount (about 5%) of a partial cleavage product derived from complexes of the T4 gene 32 protein was found due to failure of trypsin to cleave after Lys_{178} in the sequence Val_{177}-Lys_{178}-Gln_{179}...Phe_{183}−dUMP. Given the cross-linkages found in the conserved nonapeptide sequence of the other type III proteins (see Table 2), the vimentin−dUMP complexes were reanalyzed. An unusual vimentin−dUMP complex (vim_{13-22}−dUMP) was found in low yield (<5% of the total) from reaction mixtures incubated without added salt and came about as a result of anomalous cleavage by trypsin at -Arg_{22}-Pro_{23}-. Because of the low yield, unexpected mass, and presence only in complexes formed at low salt concentration, this peptide and binding site were missed in an earlier study (*5*). Two peripherin peptides, each with a dUMP bound to Phe_{28}, were generated as a result of an unusual cleavage after either Phe_{39} or Tyr_{41}. Since TPCK-treated, modified, sequencing grade trypsin was used in these experiments, it is unlikely that this cleavage was a result of the production or contamination of the reactions with pseudotrypsin (*10*). No other unusual or unexpected cleavages were observed with the 10 proteins tested. Trypsin was capable of cleaving the IF proteins cross-linked to oligo(dG·BrdU)$_{12}$·dG·3′-FITC even when the oligonucleotide was attached to one of the residues forming the scissile bond (the X in the sequence -Arg-↓-X- or -Lys-↓-X-; i.e., $peripherin_{47-51}$, NF-$L_{100-106}$, and cytokeratin $K19_{25-32}$). Each peptide−oligonucleotide complex was found to have only a single site of cross-linking, even for those whose peptide moiety contained several possible sites (i.e., dUMP was bound to either Phe_{18}, Phe_{24}, or Phe_{35} for $desmin_{17-37}$ or to either Phe_3, Tyr_5, Tyr_8, Phe_9, or Tyr_{13} in NF-L_{1-14}).

Peptide ladder sequencing with aminopeptidase M and carboxypeptidase Y was quite useful in identifying the peptides; however, neither enzyme was capable of removing an amino acid residue containing a bound dUMP molecule. Aminopeptidase M removed all unmodified residues (except Pro or Asp) from the N-terminus of the peptides, leaving the dUMP-cross-linked residue. Carboxypeptidase Y failed to remove the C-terminal residue of peptides in which the second from last residue contained the bound dUMP. Combination of these enzymatic treatments with one or more

Table 2: Sites of Cross-Linking within or Adjacent to the Nonapeptide Conserved Sequence Located near the Amino-Terminal End of the Head Domain of the Type III and Type IV IF Subunit Proteins[a]

Vimentin	++ * SSSSYRRMFGGS$_{17}$
Peripherin	++ * SSTSYRRTFGPP$_{31}$
Desmin	+ *++ * RVSSYRRTFGGA$_{20}$
GFAP	+++ ++ * RRRVTSAARRSYVSS$_{17}$
NF-L	* *- ** *+++* - GSFGYDPYFSTSYKRRYVET$_{20}$

[a] The * symbol indicates those residues found to be cross-linked to dUMP. Charged residues are indicated by a (+) for basic and a (−) for acidic residues. Residue positions are indicated by subscript numbers. For vimentin, the cross-linkage to Phe_{14} was observed only in complexes formed in the absence of added salt, whereas the other cross-linkages were observed at 75−150 mM added NaCl.

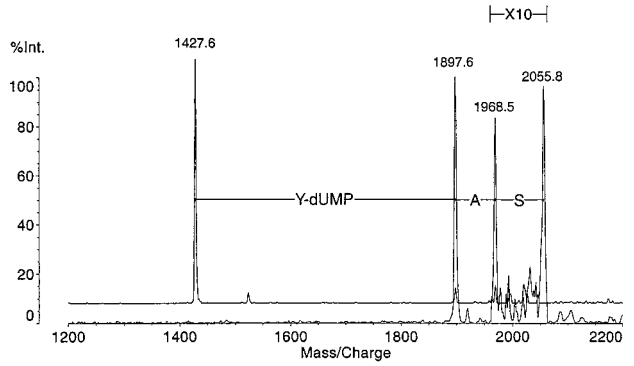

FIGURE 1: Sequencing of the N-terminus of the NF-L_{37-53}−Y_{39}−dUMP complex. The peptide ladder generated by aminopeptidase M treatment yields the N-terminal sequence NH_2-Ser-Ala- (lower spectrum), and the digestion terminates at *m*/*z* 1897. One cycle of Edman degradation with this sample results in the cleavage of one Tyr and dUMP together (upper spectrum). As indicated, the *y*-axis scale has been expanded to emphasize the peaks in the *m*/*z* range of 1920−2080.

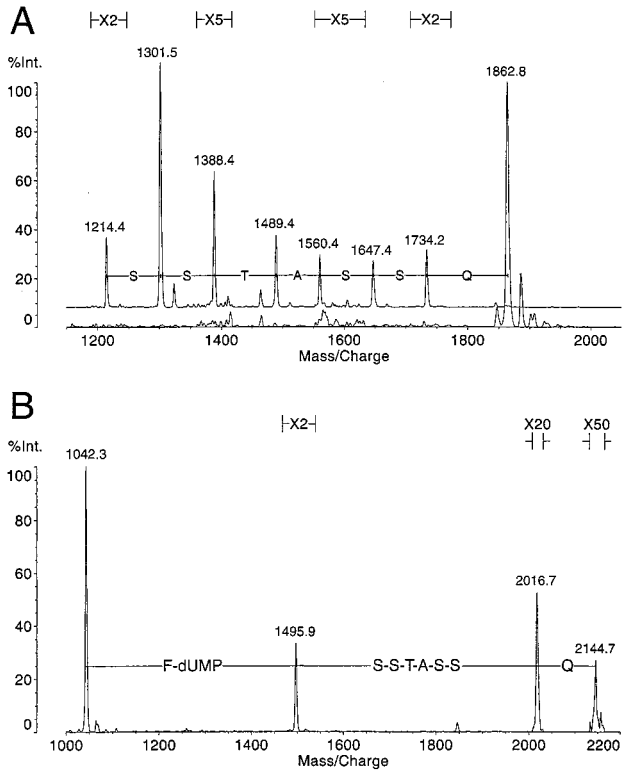

FIGURE 2: (A) Peptide ladder sequencing of the N-terminus of the $K19_{8-24}$−dUMP complex with aminopeptidase M. The $K19_{8-24}$−dUMP complex has a mass of *m*/*z* 1863 (lower spectrum). The peptide ladder generated by aminopeptidase M treatment yielded the N-terminal sequence NH_2-Gln-Ser-Ser-Ala-Thr-Ser-Ser (upper spectrum), which matches the first seven residues of $K19_{8-24}$. The enzyme digestion terminated before Phe_{15}. (B) Peptide ladder sequencing of the N-terminus of the $K19_{8-24}$−dUMP complex by a combination of aminopeptidase treatment and Edman degradation. The aminopeptidase M-treated sample was subjected to one cycle of Edman degradation, and the peptides were modified with DABITC. The modified peptide ladder yielded an N-terminal sequence, NH_2-Gln-(Ser,Ser,Ala,Thr,Ser,Ser)-Phe-dUMP, which matches the sequence of $K19_{8-24}$ with dUMP uniquely bound to Phe_{15}. The remaining mass (*m*/*z* 1042) corresponds to native $K19_{16-24}$ modified with DABITC. As indicated, the *y*-axis scale has been expanded to emphasize several of the minor peaks.

cycles of manual Edman degradation was generally sufficient to unequivocally identify the site of cross-linking within the peptide−oligonucleotide complexes. Examples of the type of results obtained are shown in Figure 1 for NF-L_{37-53}−Y_{39}−dUMP and in Figure 2 for cytokeratin $K19_{8-24}$−dUMP. Given the low intensities of several of the peptide−dUMP complexes derived from cytokeratin $K19_{8-24}$−dUMP (Figure 2A), it was necessary to derivatize the mixtures with DABITC to enhance the signals seen in the MALDI-TOF analysis. While this approach was successful for the final analysis of the $K19_{8-24}$−dUMP complex (Figure 2B) and other peptides, it must be remembered that the masses of the complexes are increased by that of the added DABITC. For two of the largest tryptic peptides, $desmin_{17-37}$ and $peripherin_{27-46}$, smaller peptide−dUMP complexes were produced by cleavage with chymotrypsin. The analysis of the chymotryptic complexes confirmed the results obtained with the tryptic complexes.

The single sites of cross-linking found for the control bacteriophage proteins encoded by T4 gene 32 and M13 gene 5 (Figure 3) were Phe_{183} and Phe_{73}, respectively. As for $K19_{8-24}$−dUMP (Figure 2), it was necessary to use DABITC-modified reaction mixtures for the final analysis of the M13 gene 5_{70-80}−dUMP complex (Figure 3B).

In control experiments not shown here, no cross-linking was observed when tubulin or actin was employed in place of the IF subunit proteins. At low salt concentrations (≤50 mM), trace amounts of BSA−oligonucleotide complexes were formed. However, no complexes were formed with BSA at NaCl concentrations higher than 50 mM. In mixing experiments, none of these three proteins had any influence whatsoever, individually or together, on the ability of vimentin to cross-link to the oligonucleotide (results not shown). However, the addition of unlabeled $oligo(dG)_{25}$ efficiently inhibited complex formation between vimentin and the $oligo(dG{\cdot}BrdU)_{12}{\cdot}dG{\cdot}3'$-FITC. For all of the proteins observed to bind the oligonucleotide, cross-linking experiments were performed in buffer containing NaCl at various concentrations ranging from 0 to over 500 mM. All of the IF and bacteriophage proteins readily bound and were cross-linked to the oligonucleotide at NaCl concentrations up to and including 250 mM (500 mM for the T4 gene 32 protein), although the total yield of cross-linked complexes declined at the higher salt concentrations. Since the MALDI-TOF spectra from IF protein−oligonucleotide complexes formed in buffer containing 75−150 mM NaCl were much cleaner than those formed in the absence of added NaCl, all of the results reported in this study were obtained at 75−150 mM added NaCl.

A summary of all of the amino acid residues found to be cross-linked to the IF subunit proteins is presented in Table 1. Estimates of the relative proportions of each peptide−oligonucleotide complex were made on the basis of the $OD_{214\ nm}$ observed during HPLC purifications. The residues indicated in Table 1 were found in roughly equimolar amounts within each protein, except for vimentin Phe_{14} (which was present only in complexes formed at low salt concentration), GFAP Phe_{48}, peripherin Phe_{84}, and NF-L Phe_{100}, all of which were found at levels of ≤15% of the other cross-linked residues. The sites of cross-linking within or adjacent to the nonapeptide conserved sequence (*11*; see also ref *12*) located near to the amino-terminal end of the head domain of the type III and type IV IF subunit proteins are presented in Table 2. A comparison of a conserved sequence within the amino

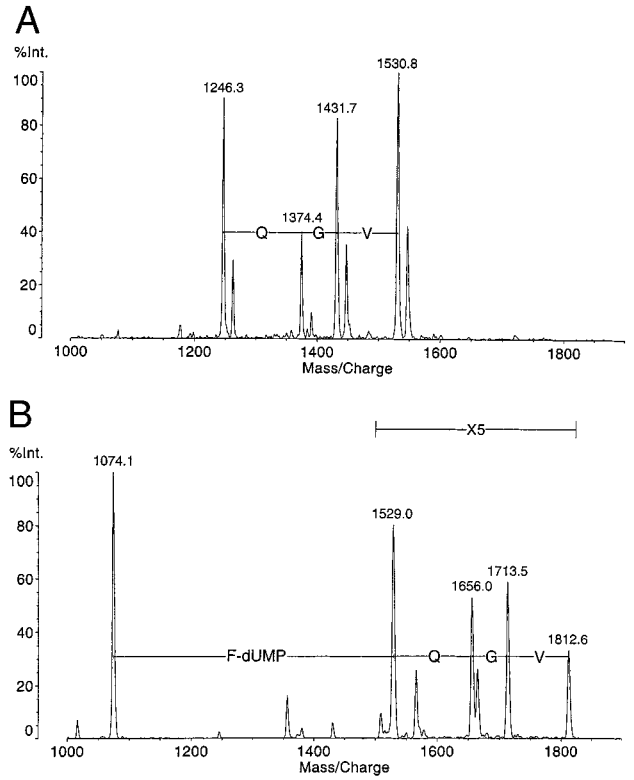

FIGURE 3: (A) Peptide ladder sequencing of the N-terminus of the P_{70-80}−dUMP complex of gene 5 protein with aminopeptidase M. The generated peptide ladder yields the N-terminal sequence of NH_2-Val-Gly-Gln-. The digestion terminates at Phe_{73}. The additional peaks at m/z +16 relative to the major peaks probably represent peptides containing an oxidized Met_{77}. (B) Peptide ladder sequencing of the N-terminus of the P_{70-80}−dUMP complex of gene 5 protein by a combination of aminopeptidase M treatment and Edman degradation. The aminopeptidase M-treated sample was subjected to one cycle of Edman degradation, and the peptide ladder was then modified with DABITC. The modified peptide ladder yielded an N-terminal sequence, NH_2-Val-Gly-Gln-Phe-dUMP, which matches the sequence of gene 5 protein peptide P_{70-80}. Since the remaining mass matches that of DABITC-modified P_{74-80}, Phe_{73} is the single site of cross-linking to DNA in the gene 5 protein. As indicated, the *y*-axis scale has been expanded to emphasize the peaks in the m/z range of 1480−1850.

terminus of selected cytokeratins shown to be involved in DNA binding and potentially important in IF polymerization is presented in Table 3. The sites of cross-linking of oligonucleotide to peripherin and GFAP within the short α-helical region immediately preceding the rod domain of the type III IF subunit proteins studied are presented in Table 4.

Analysis of the primary sequence of the head domains of all of the eight IF proteins studied revealed that all have several stretches of amino acids with a propensity for β-sheet formation (data not shown).

DISCUSSION

The methods previously employed for vimentin (*5*) proved to be appropriate for the study of the sites of nucleic acid cross-linkage in the other IF proteins and also two bacteriophage ssDNA-binding proteins. No protein−oligonucleotide complexes were obtained in experiments with actin or tubulin, and competition experiments with these and other macromolecules (BSA) failed to diminish or compete with the formation of specific IF protein−oligonucleotide complexes. The results obtained with the bacteriophage proteins are in agreement with those previously obtained (*13*) or expected from other types of experiments (*14*−*18*). These results, together with the fact that only a limited number of the aromatic amino acid residues within the IF proteins studied were found to be cross-linked, suggests that the method is specific and appropriate for the study of protein−nucleic acid interactions in general (see also ref *5*).

The amino-terminal head domains of all of the IF proteins studied were found to possess both a wealth of arginine residues and multiple aromatic residues which can be cross-linked to an oligonucleotide. Cross-linkage was limited to specific, multiple Tyr and/or Phe residues and, in the case of NF-L, included a single His residue. No cross-linking was observed to any residues within the tail or rod domains, except for NF-L. A low yield of cross-linking was found to NF-L residue Phe_{100}, which is generally accepted as being incorporated into a coiled-coil structure at the beginning of the rod domain (*19*). However, crystallization experiments

Table 3: Comparison of a Conserved Sequence within the Amino Terminus of Selected Cytokeratins Shown, for K8, K18, and K19, To Be Involved in DNA Binding and Potentially Important in IF Polymerization[a]

Cytokeratin	Sequence	Cytokeratin	Sequence
	+ * + *		
K8	RA**F**SSRS**Y**$_{25}$		
		K9	RG<u>F</u>SS-S<u>Y</u>$_{10}$
		K10	RY<u>Y</u>SSSK<u>H</u>$_{10}$
	* *		
K18	ST**F**ST-N**Y**$_{13}$	K14	RQ<u>F</u>TSSSSM$_{13}$
		K16	RQ<u>F</u>TSSSSM$_{14}$
	*		
K19	SSATS-S**F**$_{15}$	K17	RQ<u>F</u>TSSSSI$_{14}$
		K20	MD<u>F</u>SRRS<u>F</u>$_{8}$

[a] The * and bold symbol indicates those residues found to be cross-linked to dUMP. Charged residues are indicated by a (+) for basic and a (−) for acidic residues. Residue positions are indicated by subscript numbers. Residues predicted to cross-link to the oligonucleotide, based on homology to the proteins studied, are underlined.

Table 4: Sites of Cross-Linking of Oligonucleotide to Peripherin and GFAP within the Short α-Helical Region Immediately Preceding the Rod Domain of the Type III IF Subunit Proteins[a]

	-+ -* - -
Peripherin	ERLD<u>F</u>SMAEALNQEF$_{94}$
	+ -*
GFAP	ARVDFSLAGALNTGF$_{58}$
	- - - -
Vimentin	DSVDFSLADAINTEF$_{95}$
	- - - -
Desmin	ELLDFSLADAVNQEF$_{90}$

[a] Vimentin and desmin, which lack the Arg residue close to the site of cross-linking, had no observable cross-links to dUMP within this region. The * symbol indicates those residues found to be cross-linked to dUMP. The yield of cross-linkage to peripherin Phe_{84} (underlined) was <15% of that observed for the other cross-linked residues in peripherin (see Table 1). Charged residues are indicated by a (+) for basic and a (−) for acidic residues. Residue positions are indicated by subscript numbers.

with vimentin peptides corresponding to the amino-terminal end of the rod domain have shown that these peptides form monomeric α-helices (*20*), which raises the possibility that this region of the type III IF proteins is not sequestered in a rigid coiled-coil structure and thus, by analogy, NF-L Phe_{100} may also be accessible for cross-linking to the oligonucleotide in our experiments.

The equivalent cross-links observed in peripherin and GFAP within the short α-helical region immediately preceding the rod domain of IF subunit proteins (see Table 4) suggest that the local or neighboring amino acid residues play a determining role in the binding reaction. In contrast to vimentin and desmin, which have a Phe residue at the same location but cannot be cross-linked to DNA, peripherin and GFAP have an Arg residue in close proximity to the Phe residue, which apparently is required for binding, perhaps through at least partial neutralization of the two neighboring acidic Asp residues. Most, but not all, and particularly with the type III subunit proteins (see Tables 1 and 2), sites of cross-linkage have one or more neighboring Arg residues. Particularly striking examples of this phenomenon, shown in Table 2, are seen within the sequences containing or overlapping the nonapeptide sequence of the head domain shown to be important in IF polymerization (*11*). Previous experiments have demonstrated the importance of the Arg residues (*4*, *21*) and aromatic residues (*4*, *22*) of the head domain of vimentin in its ability to bind nucleic acids. The cytokeratins differ somewhat from the type III IF proteins (Table 3): while cross-linkages are observed in a conserved sequence at the beginning of the head domain of cytokeratin K8 which are also proximal to Arg residues, these same homologous Phe residues are also cross-linked in cytokeratins K18 and K19, which lack the aforementioned Arg residues. Thus, the site of binding of DNA by these proteins is, in many cases, more complicated than a simple arrangement of aromatic and Arg residues. Based on the results obtained with the three cytokeratins, it seems reasonable to predict that other cytokeratins would also bind DNA, employing the homologous residues indicated in Table 3.

NF-L, the single type IV IF protein investigated in this study, can be readily distinguished from the other proteins on the basis of two of its unique properties. First of all, many more residues (eleven) were found to be cross-linked in the NF-L head domain than in the other proteins (two−five; see Table 1). Second of all, four aromatic residues within the first nine residues of the head domain, in the absence of any adjacent Arg residues, were cross-linked to the oligonucleotide, suggesting that a high local concentration of aromatic residues can also serve as a binding motif in these proteins. None of the other proteins studied contain such a locally high concentration of aromatic residues (see Table 1).

The present results underscore the fact that while the general principles of the DNA-binding sites of these proteins are similar, the details are different. GFAP and NF-L represent the extremes, with relatively few and relatively many residues in direct position to be photochemically cross-linked to a bound oligonucleotide. The other IF proteins, with vimentin as a model example, all possess nucleic acid binding sites consisting of multiple aromatic residues generally in close proximity to one or more Arg residues. In this regard, it should be remembered that the experimental approach taken in the present study is capable of identifying only those amino acid residues in perfect register with the BrdU-containing oligonucleotide—the total number of residues comprising the actual nucleic acid binding site is undoubtedly more than those few identified here by photochemical cross-linking. Fluorescent polarization binding studies with vimentin deletion mutant proteins have demonstrated that amino acid residues 25−68 comprise the major part of the nucleic acid binding site (*4*). This peptide region is largely dispensable for filament formation and stability and probably loops out from the filament surface (*4*). It is likely that other IF subunit proteins, particularly desmin, peripherin, and the cytokeratins, have a similar configuration.

The cross-linkage of amino acid residues within regions of the head domain known or thought to be important for IF polymerization or stability provides a ready explanation for the observation that (excess) nucleic acids can prevent IF formation and even disassemble preexisting IFs (*23*). These two reactivities, binding of nucleic acids and involvement in polymerization into IFs, are two different activities of the head domain and are mutually exclusive at extreme reaction conditions. At more moderate conditions, up to about 25% of the head domains of vimentin are dispensable for IF formation (*24*) and presumably may be available for binding to nucleic acids (or for other interactions). Modification of the head domain of IF proteins by posttranslational modification (*25*−*27*) may represent a cellular fine-tuning mechanism to regulate these interactions. In a pathological situation, disruption of vimentin IF organization following liberation of the head domain by the action of the HIV-1 protease in vivo results in the entry of the vimentin head domain peptides into the nucleus and perturbation of the organization of nuclear chromatin (*28*).

The ssDNA-binding site of the bacteriophage proteins contains a twisted β-sheet, made up of five β-strands and, specifically, has a positively charged surface arranged in parallel to a series of hydrophobic pockets composed of clusters of aromatic amino acid residues (*14*). While it remains to be experimentally addressed, on the basis of the similar amino acid sequences and number and spacing of cross-linked residues, it is attractive to speculate that the type III IF proteins all make use of a β-ladder DNA-binding wing (*15*−*18*) in their interaction with single-stranded nucleic acids, as shown for vimentin (*4*, *5*). As for vimentin (*9*), computer predictions suggest that the middle of the head domain of these proteins contains several stretches of amino acids with a propensity for β-sheet formation, making such a scenario plausible. In support of these predictions, infrared spectroscopy has shown that a major fraction of the head domain of desmin has a β-sheet structure (*29*). GFAP, which naturally lacks the residues corresponding to the middle of the head domain of the other proteins, is an exception to this proposed rule. However, it should be noted that the location of the two cross-linked amino acid residues found in GFAP are exactly equivalent to the positions of the first and last cross-linked amino acid residues in peripherin, suggesting that these two proteins share some, but not all, nucleic acid binding motifs or modalities. Detailed inspection shows that Phe_{24} of desmin and Phe_{39} of peripherin are not proximal to any Arg residues. If the structure of the DNA-binding site of the other IF proteins is indeed similar to that of vimentin, it is possible that these residues are actually located near to Arg residues in the final structure formed

after (nucleic acid induced) folding of the N-terminus; i.e., the proximity is not evident in the primary sequence but is present in the secondary structure. The N-terminus of vimentin (*22*), and presumably the other IF proteins as well, exhibits considerable structural flexibility and can adopt a variety of configurations, especially in response to posttranslational modifications and upon interaction with other molecules such as nucleic acids (*22*), lipids (*30*), etc.

The ability of the cytokeratins K8, K18, and K19 to be cross-linked to the oligonucleotide is in good agreement with a previous study (*8*), in which it was demonstrated that a mixture of these three proteins bound to various model phosphodiester and phosphorothioate (antisense) oligonucleotides in vitro in a manner and with affinities similar to that of vimentin. In vivo experiments have shown that the cytokeratins of Novikoff ascites hepatoma cells can be selectively chemically cross-linked to DNA (*31*). In other studies (*32*, *33*), cytokeratins K8, K18, and K19 were not only found to be bound to nuclear DNA in human breast cancer cells in vivo but were also observed to bind nuclear DNA in an estrogen-dependent manner and participate in the organization of nuclear chromatin. These authors (*32*) point out that the increased expression of cytokeratins K8 and K18 correlates with the development of a malignant phenotype in epithelial cells (*34*); they furthermore suggest that the cytokeratins are authentic nuclear matrix attachment region (MAR) binding proteins and participate in the organization and regulation of function of nuclear DNA. Also, vimentin was found to be cross-linked to nuclear DNA of a hormone-independent human breast cancer cell line with high metastatic potential (*33*). An earlier study has demonstrated that vimentin is tightly bound to ssDNA in the nucleus of Chinese hamster ovary cells (*35*). Desmin and GFAP exhibit properties grossly similar to those of vimentin in the binding in vitro of mouse genomic DNA fragments, particularly of repetitive and mobile sequence elements (*6*). Interestingly, SDS-stable cross-linkage products of vimentin with repetitive and mobile DNA sequences, as well as with MAR and mitochondrial DNA sequences, could be isolated from actively proliferating fibroblast cells (*36*). Last, disruption of neurofilaments in cultured dorsal root ganglion neurons induces an increase in the motion of nuclear chromatin, possibly due to a perturbation of a neurofilament−chromatin (DNA) complex (*37*). Thus, the data presented in this paper directly support our hypothesis (*3*) that IF proteins may play an active role in genome-based events as a result of their ability to bind to nucleic acids.

Our hypothesis (*3*, *5*) proposes that IFs play a supportive or fine-tuning role in genome-based events, which would also be compatible with the observed lack of lethality in vimentin knockout mice (*38*). In this context, it should be noted that the vimentin knockout mice (*38*) do indeed lack vimentin IFs, but instead express a fusion protein containing the first 59 amino acids of the vimentin head domain (i.e., the bulk of the DNA-binding domain; refs *4* and *5* and this study) in the nucleus of normally vimentin-positive cells. This fusion protein is also capable of binding DNA (*39*). Thus, these mice are vimentin IF knockouts but not vimentin DNA-binding domain knockouts. The repetitive arrangement of multiple nucleic acid binding sites on the filament surface probably allows IFs to enter into cooperative interactions with, preferentially repetitive, genomic DNA sequence elements (*6*, *23*, *40*). Likewise, these interactions are responsible for the ability of IFs in vivo in the cytoplasm to bind nucleic acids (*8*, *41*) and nucleoprotein complexes, such as ribosomes (*42*, *43*). This might be a particularly important function of NF-L, which is unique among the IF subunit proteins studied in that it has more than twice as many aromatic amino acid residues that can be cross-linked to nucleic acid and may thus enter into either more or different types of interactions than the other IF subunit proteins. NF-L is found not only in the soma of neurons, where it might participate in the organization of nuclear chromatin, but also in the very long axons, where its ability to bind nucleic acids might be of paramount importance in the transport and storage of nucleic acids and nucleoproteins.

The ability of IF subunit proteins to participate in a plethora of diverse cellular activities ranging from the mechanical integration of cellular space (*44*), the strengthening of the cytoskeleton (*1*), providing a solid-state matrix for enzymatic reactions (*45*), transport processes (*46*), and signal transduction (*47*), and, finally, interaction with mitochondrial and nuclear DNA/chromatin (*36*) arises as a natural consequence of their multidomain structure: each domain contributes its unique reactivities to the individual proteins. The rod domains of the IF subunit proteins have the highest homology and are primarily responsible for the fact that all IF subunit proteins polymerize into structurally similar 10 nm IFs that provide a common framework for multisite interactions with other cellular constituents. Much of the uniqueness of the activities of these IFs is found in the non-α-helical head and tail domains, which are largely exposed on the filament surface. As shown in the present study, while the reactivities of the head domains of all of these proteins are grossly similar (they bind to nucleic acids), the exact details are unique in that they make use of different numbers of and distributions of specific amino acids. Dependent on the wealth of arginine residues for facilitated target acquisition and stabilization of complexes with nucleic acids via electrostatic forces, the specificity of complex formation is determined by amino acid residues capable of stacking interactions and hydrogen bonding. This should enable each of these proteins to interact with different, specific nucleic acids and their higher order structures (ref *6* and references cited therein) or regions of the nuclear chromatin, enabling them to specifically participate in the regulation of genomic events.

ACKNOWLEDGMENT

We thank Mrs. Ulrike Traub and Mrs. Annemarie Scherbarth for the provision of the cultured cells and Mrs. Monica Berthel for excellent technical assistance.

REFERENCES

1. Fuchs, E., and Weber, K. (1994) *Annu. Rev. Biochem. 63*, 345.
2. Herrmann, H., and Aebi, U. (2000) *Cur. Opin. Cell Biol. 12*, 79.
3. Traub, P., and Shoeman, R. L. (1994) *Int. Rev. Cytol. 154*, 1.
4. Shoeman, R. L., Hartig, R., and Traub, P. (1999) *Biochemistry 38*, 16802.
5. Wang, Q., Shoeman, R., and Traub, P. (2000) *Biochemistry 39*, 6645.
6. Tolstonog, G. V., Wang, X., Shoeman, R., and Traub, P. (2000) *DNA Cell Biol. 19*, 647.
7. Vorgias, C. E., and Traub, P. (1986) *Z. Naturforsch. 41c*, 897.

8. Shoeman, R. L., Hartig, R., Huang, Y., Grüb, S., and Traub, P. (1997) *Antisense Nucleic Acid Drug Dev. 7*, 291.
9. Shoeman, R. L., Mothes, E., Kesselmeier, C., and Traub, P. (1990) *Cell Biol. Int. Rep. 14*, 583.
10. Smith, R. L., and Shwa, E. (1969) *J. Biol. Chem. 244*, 4704.
11. Herrmann, H., Hofmann, I., and Franke, W. W. (1992) *J. Mol. Biol. 223*, 637.
12. Beuttenmüller, M., Chen, M., Janetzko, A., Kühn, S., and Traub, P. (1994) *Exp. Cell Res. 213*, 128.
13. Shamoo, Y., Williams, K. R., and Konigsberg, W. H. (1988) *Proteins: Struct., Funct., Genet. 4*, 1.
14. Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg, W. H., and Steitz, T. A. (1995) *Nature 376*, 362.
15. De Jong, E. A. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Tesser, G. I., Konings, R. N. H., and Hilbers, C. W. (1989) *J. Mol. Biol. 206*, 119.
16. De Jong, E. A. M., van Duynhoven, J. P. M., Harmsen, B. J. M., Tesser, G. I., Konings, R. N. H., and Hilbers, C. W. (1989) *J. Mol. Biol. 206*, 133.
17. Folmer, R. H., Folkers, P. J., Kaan, A., Jonker, A. J., Aelen, J. M., Konings, R. N., and Hilbers, C. W. (1994) *Eur. J. Biochem. 224*, 663.
18. Folmer, R. H. A., Nilges, M., Papavoine, C. H. M., Harmsen, B. J. M., Konings, R. N. H., and Hilbers, C. W. (1997) *Biochemistry 36*, 9120.
19. Quinlan R., Hutchison C., and Lane, B. (1995) *Intermediate filament proteins. Protein Profile 2*, pp 795−952, Academic Press, New York.
20. Strelkov, S. V., Herrmann, H., Geisler, N., Lustig, A., Zimbelmann, R., Burkhard, P., and Aebi, U. (2000) *Mol. Biol. Cell 11*, 285a.
21. Traub, P., and Vorgias, C. E. (1983) *J. Cell Sci. 63*, 43.
22. Traub, P., Mothes, E., Shoeman, R., Kühn, S., and Scherbarth, A. (1992) *J. Mol. Biol. 228*, 41.
23. Traub, P., Mothes, E., Shoeman, R. L., Schröder, R., and Scherbarth, A. (1992) *J. Biomol. Struct. Dyn. 10*, 505.
24. Andreoli, J. M., and Trevor, K. T. (1994) *Exp. Cell Res. 214*, 177.
25. Skalli, O., Chou, Y.-H., and Goldman, R. D. (1992) *Trends Cell Biol. 2*, 308.
26. Inagaki, M., Matsuoka, Y. Tsujimura, K. Ando, S., Tokui, T., Takahasi, T., and Inagaki, N. (1996) *BioEssays 18*, 481.
27. Yuan, J., Huiatt, T. W., Liao, C. X., Robson, R. M., and Graves, D. J. (1999) *Arch. Biochem. Biophys. 363*, 314.
28. Shoeman, R. L., Hüttermann, C., Hartig, R., and Traub, P. (2001) *Mol. Biol. Cell 12*, 143.
29. Heimberg, T., Schuenemann, J., Weber, K., and Geisler, N. (1996) *Biochemistry 35*, 1375.
30. Perides, G., Harter, C., and Traub, P. (1987) *J. Biol. Chem. 262*, 13742.
31. Wedrychowski, A., Schmidt, W. N., Ward, W. S., and Hnilica, L. S. (1986) *Biochemistry 25*, 1.
32. Spencer, V. A., Coutts, A. S., Samuel, S. K., Murphy, L. C., and Davie, J. R. (1998) *J. Biol. Chem. 273*, 29093.
33. Spencer, V. A., Samuel, S. K., and Davie, J. R. (2000) *Cancer Res. 60*, 288.
34. Trask, D. K., Band, V., Zajchowski, D. A., Yaswen, P., Suh, T., and Sager, R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2319.
35. Cress, A. E., and Kurath, K. M. (1988) *J. Biol. Chem. 263*, 19678.
36. Tolstonog, G. V., Mothes, E., Shoeman, R. L., and Traub, P. (2001) *DNA Cell Biol.* (in press).
37. Hay, M., and DeBoni, U. (1991) *Cell Motil. Cytoskel. 18*, 63.
38. Colucci-Guyon, E., Portier, M.-M., Dunia, I., Paulin, D., Pournin, S., and Babinet, C. (1994) *Cell 79*, 679.
39. Tolstonog, G. V., Shoeman, R. L., Traub, U., and Traub, P. (2001) *DNA Cell Biol.* (in press).
40. Wang, X., Tolstonog, G., Shoeman, R. L., and Traub, P. (1996) *DNA Cell Biol. 15*, 209.
41. Hartig, R., Huang, Y., Janetzko, A., Shoeman, R., Grüb, S., and Traub, P. (1997) *Exp. Cell Res. 233*, 169.
42. Bauer, C., and Traub, P. (1995) *Eur. J. Cell Biol. 68*, 288.
43. Traub, P., Bauer, C., Hartig, R., Grüb, S., and Stahl, J. (1998) *Biol. Cell 90*, 319.
44. Lazarides, E. (1980) *Nature 283*, 249.
45. Sarria, A. J., Panini, S. R., and Evans, R. M. (1992) *J. Biol. Chem. 267*, 19455.
46. Gyoeva, F. K., and Gelfand, V. I. (1991) *Nature 353*, 445.
47. Maniotis, A. J., Chen, C. S., and Ingber, D. E. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 849.

BI0108305