

Mass Spectrometric Analysis of the N Terminus of Translational Initiation Factor eIF4G-1 Reveals Novel Isoforms*

Received for publication, November 20, 2001, and in revised form, January 10, 2002
Published, JBC Papers in Press, January 30, 2002, DOI 10.1074/jbc.M111134200

Christopher A. Bradley[‡], Júlio César Padovan[§], Timothy L. Thompson[¶], Clint A. Benoit[¶],
Brian T. Chait[§], and Robert E. Rhoads^{‡||}

From the [‡]Department of Biochemistry and Molecular Biology and [¶]Research Core Facility, Louisiana State University Health Sciences Center, Shreveport, Louisiana 71130-3932 and the [§]Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, New York, New York 10021-6399

In eukaryotes, translation initiation factor 4G (eIF4G) acts as the central binding protein for an unusually large number of proteins involved in mRNA metabolism. Several gene products homologous to eIF4G have been described, the most studied being eIF4G-1. By its association with other initiation factors, eIF4G-1 effects mRNA cap and poly(A) recognition, unwinding of secondary structure, and binding to the 43S initiation complex. Multiple electrophoretic isoforms of eIF4G-1 are observed, and multiple cDNAs have been reported, yet the relationship between the two is not known. We report here a new cDNA for eIF4G-1, present as a previously unidentified human expressed sequence tag, that extends the long open reading frame, provides a new in-frame initiation codon, and predicts a longer form of eIF4G-1 than reported previously. eIF4G isoforms from human K562 cells were cleaved with recombinant Cocksackievirus 2A protease and the N-terminal domains purified by m⁷GTP-Sepharose chromatography and polyacrylamide gel electrophoresis. Proteins were digested with proteolytic enzymes and peptide masses determined by matrix-assisted laser desorption ionization-time of flight mass spectrometry. In selected cases, peptides were sequenced by electrospray-mass spectrometry fragmentation. This identified the N termini of the three most abundant eIF4G-1 isoforms, two of which had not previously been proposed. These proteins appear to have been initiated from three different AUG codons.

The three stages of protein synthesis are catalyzed by three groups of proteins: initiation, elongation, and termination factors (1). Initiation is characterized by formation of a series of initiation complexes, each catalyzed by a different subset of initiation factors. Recruitment of mRNA to the 43 S initiation complex to form the 48 S initiation complex involves eIF3,¹

PABP, and the eIF4 proteins. eIF3 is a 520-kDa multimer that is required for both Met-tRNA_i and mRNA binding. PABP is a 70-kDa protein that specifically binds poly(A) and homo-oligomerizes. The eIF4 factors consist of: eIF4A, a 46-kDa RNA helicase; eIF4B, a 70-kDa RNA-binding and -annealing protein that stimulates eIF4A; eIF4E, a 25-kDa cap-binding protein; and eIF4G, a group of proteins of 154–180 kDa that form specific complexes with all of the other proteins known to be involved in mRNA recruitment.

Proteins sharing eIF4G homology represent the products of at least three different genes in mammalian cells (2–4). One of these has been mapped to the chromosomal location 3q27-qter (5). In accordance with the nomenclature system proposed for translation factors (6), the protein product of the gene at 3q27-qter is referred to as eIF4G-1, whereas its two homologues are eIF4G-2 (also known as p97, eIF4G2, DAP5, and NAT1; Refs. 3, 7, and 8) and eIF4G-3 (also known as eIF4G3, and eIF4GII; Ref. 4). Most studies reported to date have concerned eIF4G-1.

eIF4G appears to serve as a nucleation site for co-localization of an unusually large number of proteins involved in mRNA metabolism. Combining data from yeast, plant, and mammalian eIF4Gs, these include (in approximate order of binding sites on eIF4G from N to C termini): the influenza protein NS1 (9), the cytoplasmic poly(A)-binding protein PABP (10, 11), the rotavirus protein NSP3 (12), the decapping protein Dcp1 (13), the cytoplasmic cap-binding protein eIF4E (14–16), the nuclear cap-binding protein CBP80 (17, 18), the initiation factor eIF3 (14, 19), the RNA helicase eIF4A at two distinct sites (14, 20, 21), RNA itself (22–24), the heat shock proteins hsp27 (25) and hsp70 (26), and the eIF4E kinase Mnk1 (27–29). In some cases it has been shown that eIF4G-1 not only binds proteins but also affects their activities or binding of other proteins (19, 21, 30).

The mRNA recruitment step is rate-limiting for initiation under normal cellular conditions and appears to be highly regulated (31, 32). The best studied regulatory mechanism involving eIF4G is its cleavage by 2A protease of entero- and rhinoviruses and L protease of foot-and-mouth-disease virus. Upon infection of mammalian cells with these picornaviruses, most host protein synthesis is shut off coincident with the appearance of viral proteins (33). This is thought to be mediated by a switch from cap-dependent to cap-independent translation. Complexes containing eIF4G restore cap-dependent translation in lysates of poliovirus-infected cells (34–36). eIF4G was discovered as a result of its proteolysis coincident with the loss of cap-dependent initiation during poliovirus infection (37). eIF4G is separated into two functional domains, an N-terminal cleavage product (cp_N) that binds eIF4E and PABP, and a C-terminal cleavage product (cp_C) that binds eIF4A and eIF3 (14, 38). Initiation of picornaviral mRNA translation is via an internal ribosome entry site (39, 40). Cleavage of eIF4G

* This work was supported by Grants GM20818 and RR00862 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

|| To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Louisiana State University Health Sciences Center, 1501 Kings Highway, Shreveport, LA 71130-3932. Tel.: 318-675-5161; Fax: 318-675-5180; E-mail: rrhoad@lsuhsc.edu.

¹ The abbreviations used are: eIF, eukaryotic initiation factor; aa, amino acid; EST, expressed sequence tag; MALDI-TOF-MS, matrix-assisted laser desorption ionization-time of flight mass spectrometry; MOWSE, molecular weight search; nt, nucleotide residues; PABP, poly(A)-binding protein; MOPS, 4-morpholinepropanesulfonic acid; LC-ESI, liquid chromatography-electrospray ionization; ACTH, adrenocorticotropic hormone.

drastically inhibits translation of capped mRNAs *in vitro*, whereas internal initiation and initiation of uncapped mRNAs are either unaffected or even stimulated (41–43).

Multiple isoforms of eIF4G-1 are observed by SDS-PAGE, but the origin of these is not known. However, an analysis using 2A protease cleavage, SDS-PAGE, and antibodies directed against different domains of eIF4G suggested that the heterogeneity was attributable to the ~50-kDa cp_N domain (38). Digestion with L protease further delimited the source of heterogeneity to the N-terminal ~30 kDa cp_{N1} domain (14). Multiple cDNAs for human eIF4G-1 have also been reported. The initial cloning, based on two overlapping cDNAs from fetal and adult brain, predicted a protein of 154 kDa (2). Subsequent cloning revealed a cDNA corresponding to a protein with an additional 156 aa at the N terminus (44). An initiation codon was proposed based on alignment with eIF4G-3 and lack of any further upstream cDNA sequence. Further cloning revealed a cDNA that was 42 nt longer (45), indicating either that multiple mRNAs exist or that Imataka *et al.* (44) had not reached the 5'-end of the same mRNA. Finally, a fourth and fifth cDNA were reported (46). One of these independently confirmed some of the sequence reported by Johannes and Sarnow (45) but did not extend the cDNA. The other was collinear with the other cDNAs up to a point, upstream of which it deviated, suggesting a splice variant. None of these cDNAs provided a new in-frame AUG. Thus, the most upstream AUG codon for all five eIF4G-1 cDNAs reported to date is that originally proposed by Imataka *et al.* (44).

The present study was motivated by the fact that the relationship between the multiplicity of electrophoretic forms of the protein and the multiplicity of cDNAs is not known. Importantly, despite speculation based on cDNA sequences, the N terminus has not been established for any eIF4G-1 isoform. It is important to establish the actual protein structures, since eIF4G-1 isoforms differing at the N terminus may contain different binding sites for proteins involved in translational control. This study identifies an EST in the public databases as corresponding to a longer form of eIF4G-1 mRNA. Because it predicts a new upstream, in-frame AUG, it could encode an even longer isoform of eIF4G-1 than those predicted from previous cDNA sequences. Mass spectrometric analysis confirmed that this was the case and also established the structure of two other eIF4G-1 isoforms.

EXPERIMENTAL PROCEDURES

Materials—Porcine trypsin was obtained from Promega (catalog number V5111). Endoproteinase Lys-C (catalog number P3428) and bovine α -chymotrypsinogen A (catalog number C4879) were obtained from Sigma. Arg-C from *Clostridium histolyticum* was purchased from Worthington (catalog number LS001641). Asp-N (catalog number 1420488) and Complete, Mini, Protease Inhibitor Cocktail Tablets (catalog number 1836153) were obtained from Roche Molecular Biochemicals. m⁷GTP-Sepharose 4B was purchased from Amersham Biosciences.

Identification of a New EST for eIF4G-1—As noted above, Bushell *et al.* (46) reported a novel cDNA related to eIF4G-1, GenBankTM nucleotide accession number AF002815. The hypothetical protein encoded by this cDNA was assigned GenBankTM protein accession number AAC78442. We compared the cDNA sequence corresponding to aa residues -24 to +49 of this protein² to the est_human database using the program blastn found under the NCBI BLAST suite of programs. We obtained a new EST derived from an adult melanoma: GenBankTM nucleotide accession number AL120751. The 215-nt segment from nt

25–240 of AL120751 exactly matched the 215-nt segment from nt 4–219 of AF002815.

The sequence reported for AL120751 covered only the most 5'-terminal 627 nt, even though the complete eIF4G-1 cDNA is predicted to contain 5510 nt (2, 44, 45). We obtained the plasmid corresponding to this EST, termed *Homo sapiens* cDNA clone DKFZp762O191, from the German Genome Project and determined additional DNA sequence information using the DNA sequencing facility at Iowa State University. The 3'-end of AL120751 (nt 4647–5306, using a composite numbering system)³ was determined using an oligo(dT) 3'-A primer. A 3'-terminal poly(A) tract of 111 nt was observed using a primer corresponding to the SP6 promoter in the vector. A sequence covering nt 439–1155 was obtained using the sense primer 5'-AACACGCCCTTCT-CAGCCCCGC-3'. This corrected an entry of "N" at nt 456 to G in the GenBank record of AL120751. A sequence covering nt 710–1399 was obtained using the antisense primer 5'-GGGGCAAGCTGGGGGAG-GAGC-3'. This sequence exactly matched nt 328–1018 of cDNA AF104913, which encodes protein AAC82471. Finally, sequences covering nt 1–300 and 97–793 were obtained using the sense primer 5'-CGCCACGGCCGAAGCAGCTAG-3' and antisense primer 5'-AACACGCCTTCTCAGCCCCGC-3', respectively. Overall, this extended the sequence information for cDNA clone DKFZp762O191 by 1544 nt.

Generation of K562 Cell Lysates—Human K562 cells were grown in RPMI medium (Invitrogen) containing 10% fetal bovine serum (Atlanta Biologicals, Norcross, GA). Cultures were grown to confluence in 175-cm² tissue culture flasks maintained in a humidified, 5% CO₂ environment at 37 °C. A standard preparation of lysate was derived from 3.2 × 10⁹ cells (16 T-175 flasks). Cells were resuspended on ice for 30 min in an equal volume of Buffer A (1% Triton X-100, 10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM EDTA, 1 tablet of protease inhibitor mixture per 20 ml buffer) and then centrifuged at 25,000 × *g* for 20 min. The supernatant was frozen in liquid N₂ and stored at -80 °C.

Immunological Procedures—Two antibodies against different regions of eIF4G-1 were used.⁴ Anti-Peptide 7 antibodies were obtained as described previously (2). Anti-Peptide 10 antibodies were produced and affinity purified using the peptide CRAQPPSSAASR, which corresponds to aa residues 55–65 of the consensus eIF4G-1 sequence² with an added N-terminal Cys residue, as described previously (2). Immunoblotting was carried out as described previously (47). Incubation with the anti-Peptide 7 antibodies (1:1000) was carried out at room temperature for 1 h. Incubation with affinity-purified anti-Peptide 10 antibodies (1:50) was at 4 °C overnight. Both antibodies were visualized with alkaline phosphatase-conjugated goat anti-rabbit IgG antibodies (Vector Laboratories, Burlingame, CA) at 1:1000.

Purification of cp_N —Recombinant 2A protease from Coxsackievirus serotype B4 was prepared as described previously (41). K562 cell lysates were cleared after thawing by spinning at 25,000 × *g* for 20 min at 4 °C. The supernatant was incubated with recombinant 2A protease at a final concentration of 50–100 μg/ml for 30 min on ice. (As 2A protease preparations differed somewhat in activity, the amount was adjusted to produce complete cleavage of eIF4G-1, based on western blotting with anti-Peptide 7 antibodies.) The 2A protease-treated lysate was then subjected to m⁷GTP-Sepharose affinity chromatography as described previously (48) but with the following modifications. One standard batch of K562 lysate was gently rotated with 2 ml of m⁷GTP-Sepharose for 1 h at 4 °C. The slurry was then transferred to a column and the flow-through fraction was reappplied to the column. The column was washed first with 12 volumes of Buffer B₅₅ (20 mM MOPS, pH 7.6, 10% (w/v) glycerol, 0.5 mM EDTA, 0.25 mM dithiothreitol, 25 mM NaF, 55 mM KCl) and then with 3 volumes of 100 μM GTP in Buffer B₅₅. Proteins were eluted with 4 volumes of 200 μM m⁷GTP in Buffer B₅₅. The eluate was frozen in liquid N₂ and stored at -80 °C.

Electrophoretic Separation of cp_N Isoforms—Linear polyacrylamide was prepared (49) and added to each cp_N sample at a final concentration of 120 μg/ml as carrier. Ice-cold 100% trichloroacetic acid was added to

² Various numbering systems have been used in previous reports dealing with the human eIF4G-1 proteins. To allow comparison between sequences, we have performed an alignment of the proteins predicted from all eIF4G-1 cDNAs. The aa numbers used throughout the present work are based on this alignment, rather than the numbers used in the original publications, and are shown in Fig. 1.

³ The sequence deposited as GenBankTM nucleotide accession number AL120751 contains only 627 nt of the cDNA represented by clone DKFZp762O191. The current report confirms the entire EST sequence and extends the cDNA sequence. Alignment with other eIF4G-1 cDNAs predicts a 5510-nt mRNA corresponding to clone DKFZp762O191. The nucleotide numbers used in this report, unless otherwise noted, are based on this composite mRNA in which nt 1 is the first nt of AL120751.

⁴ Names of eIF4G-1 peptides used for designing oligonucleotide probes and making site-specific antibodies in earlier studies (2, 52) are not related to the names for eIF4G-1 peptides used in the present report (Fig. 3). To avoid confusion, the earlier peptide names are always preceded by the term "anti-" in the present work.

a final concentration of 10% and the sample allowed to stand on ice for 30 min. The precipitate was collected by centrifugation at 25,000 \times g for 20 min, washed four times with ice-cold 80% aqueous acetone, and dissolved in 1 \times SDS-loading buffer (adjusted to pH 8.2; Ref. 50). Alkylation of Cys residues was performed as described previously (50). The sample was then adjusted to pH 6.8 with HCl, and the cp_N isoforms were separated by electrophoresis on a 16 \times 20-cm gradient gel (8–15% acrylamide:*N,N'*-bisacrylamide, 30:0.8) with a 4% stacking gel. Electrophoresis was carried out a constant current of 16 mA for 4 h followed by 24 mA for 16–20 h. Protein bands were stained with Coomassie Blue, excised, and stored at –80 °C.

Protease Digestion—Gel pieces were minced and further destained with three washes (400 μ l each) of 50% acetonitrile, 25 mM ammonium bicarbonate, pH 8.0. The polyacrylamide was dehydrated for 5 min with 100% acetonitrile followed by vacuum centrifugation in a Savant Speed-vac for 30 min. In-gel digestion with trypsin was performed overnight at 37 °C by the addition of 15 μ l of enzyme (10 μ g/ml) in 25 mM NH₄HCO₃, pH 8.0, to each gel piece. For Lys-C digestion, a stock enzyme solution of 50 μ g/ml was made in 0.1 M Tris-HCl, pH. 8.0, and 15 μ l were added to each gel piece. Digestion was performed overnight at 37 °C. For Arg-C digestion, the enzyme was reconstituted at 2 mg/ml in 1 mM CaCl₂, 2.5 mM dithiothreitol. In-gel digestion was carried out at room temperature overnight in 15 μ l of 50 mM sodium phosphate, pH 7.6, 2.5 mM dithiothreitol by the addition of 3 μ g enzyme per gel slice. Finally, for Asp-N digestion, the enzyme was reconstituted in H₂O to give a concentration of 40 μ g/ml in 10 mM Tris, pH 7.5. In-gel digestion was carried out overnight at 37 °C in 15 μ l of 50 mM sodium phosphate, pH 8.0, by the addition of 4 μ l of enzyme per gel slice.

Mass Spectrometry—Mass spectrometric analysis was performed at both the LSUHSC-S Research Core Facility or the Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University. At LSUHSC-S, MALDI-TOF-MS was performed on a PerSeptive Biosystems Voyager-DE PRO Biospectrometry work station. At Rockefeller University, two instruments were used. Peptide mapping by MALDI-TOF-MS was performed on a PerSeptive Biosystems Voyager-DE STR Biospectrometry Work station. Sequence information was obtained by LC-ESI-MS/MS on a Finnigan LCQ-DECA ion trap mass spectrometer.

Peptides were prepared for MALDI-TOF-MS after proteolytic digestion by extraction from gel pieces twice with 50- μ l portions of 50% acetonitrile, 5.0% trifluoroacetic acid. Peptides in the extract were dried, dissolved in 15 μ l of 0.1% trifluoroacetic acid, and purified on a ZipTip (Millipore). They were eluted with 2 μ l of the appropriate organic acid (matrix) dissolved in 50% acetonitrile, 0.1% trifluoroacetic acid, and spotted on a MALDI plate. For masses in the 800–6000 Da range, the matrix was a 0.01 mg/ml solution of α -cyano-4-hydroxy-*trans*-cinnamic acid (Sigma, catalog number C-2020). For masses in the 6000–26,000 Da range, the matrix was a 0.01 mg/ml solution of sinapinic acid (3,5-dimethoxy-4-hydroxy-*trans*-cinnamic acid; Aldrich, catalog number D13,460-0). When internal calibration was used, the eluting solution also included peptide mass standards. Data were summed over 50–100 acquisitions in delayed extraction mode.

At LSUHSC-S, data analysis was performed using the Data Explorer software, version 3.5–4.0. At Rockefeller University, analysis was performed using the software program M over Z from Proteometrics, LLC. In both cases, spectra were subjected to algorithms for base-line correction and noise removal at two S.D. values. Throughout the current report, monoisotopic masses are reported for <2500-Da peptides, whereas average masses are reported for >2500-Da peptides. Theoretical masses were determined using the Peptide Mass tool at the ExPASy Proteomics website (ca.exPASy.org/tools/). Peaks seen in both the sample of interest and also in a blank gel treated identically were eliminated from further consideration. For automatic matching of observed to predicted peptide masses, we used Auto-MS Fit, version 1.2.18 (PerSeptive Biosystems). For manual matching, we considered that peptides matched if their masses were within 1 Da for the range 800–10,000 Da. Because of unresolved microheterogeneity, caused by Met oxidation for example, peptides with molecular masses above 10,000 Da were considered to be matches if the experimentally determined masses were within 0.6% of the calculated values.

LC-ESI-MS/MS analysis was carried out using a Smart System (Amersham Biosciences) equipped with 10-ml syringe pumps and a pre-column flow splitter (Michrom Bioresources, Auburn, CA). The chromatographic eluate was monitored by on-line mass spectrometry using an electrospray ion trap mass spectrometer, model LCQ-DECA (Finnigan ThermoQuest, San Jose, CA). Peptide mixtures were diluted 10-fold with 0.01% trifluoroacetic acid (v/v) in water/methanol/acetic acid (949:50:1, v/v/v) and loaded on a reverse phase Magic C18 column (50 \times 0.2 mm inside diameter; pore size, 100 Å; particle size, 5 μ m) from Michrom

Bioresources (Auburn, CA). Peptide separation was performed at room temperature with a fast-rising methanol gradient (in 5 min) at a flow rate of 2.8 μ l/min. The eluate was transferred through a 50- μ m inside diameter fused silica capillary from the column to the ion source of the mass spectrometer and electrosprayed at 2.8–3.2 kV. The transport capillary in the mass spectrometer was kept at 130–150 °C in order to assist desolvation.

RESULTS

Identification of a New EST for eIF4G-1—As noted above, several cDNAs for eIF4G-1 have been reported (2, 44–46). Four of the polypeptides theoretically encoded, NP_004944, AAC82471, AAC78443, and eIF4G-I_{ext}, are collinear, but one of them, AAC78442, deviates from the others upstream of aa 50 (Fig. 1). In some cases, the polypeptide sequences deposited in GenBankTM represent those encoded by the longest open reading frame following an AUG codon (NP_004944, AAC82471), but in other cases, the polypeptide sequence continues in the N-terminal direction despite the absence of an AUG codon (AAC78442, AAC78443, eIF4G-I_{ext}). We used sequence information from the cDNA with the longest open reading frame (GenBankTM nucleotide accession number AF002815, which encodes GenBankTM protein accession number AAC78442; see Fig. 1) to search for additional ESTs. The result was a new EST entered in GenBankTM under nucleotide accession number AL120751 (Fig. 1). The EST sequence reported (627 nt) was not previously identified as corresponding to eIF4G-1 mRNA, but a comparison with other eIF4G-1 cDNAs indicated an identical sequence over 215 nt.

We obtained the plasmid corresponding to this GenBankTM entry from the German Genome Project. Sequencing portions of the plasmid insert from both the 5' and 3' termini revealed: (i) agreement over the entire EST sequence for AL120751 except for a single nucleotide sequencing ambiguity at nt 456; (ii) preservation of reading frame between the 3'-end of the EST and other eIF4G-1 cDNA sequences; (iii) the presence of a 111-nt poly(A) tract; (iv) collinearity with the cDNAs encoding AAC82471, AAC78443, and eIF4G-I_{ext} (Fig. 1); and (v) partial collinearity with the cDNAs encoding NP_004944 and AAC78442 (see "Experimental Procedures"). If the new cDNA is collinear with these cDNAs throughout its entire length, it corresponds to an mRNA of 5510 nt.³

Surprisingly, the polypeptide encoded by AL120751 deviates from the AAC78442 polypeptide upstream of aa 50 (Fig. 1), even though sequences from the cDNA encoding it, AF002815, were used to find AL120751. Further comparison of AL120751, AF002815, and the corresponding human genomic sequence (GenBankTM nucleotide accession number AC078797.8; gi number 15887175; nt 122674–133235) indicates that AF002815 lacks two exons that are present in AL120751 (data not shown). The most 5' exon, however, is common to AF002815 and AL120751. The absence of these two internal exons from AF002815 results in a different reading frame upstream of aa 50. On this basis, we suggest that AF002815 represents a splice variant that lacks internal exons present in the AL120751 cDNA.

The new polypeptide encoded by AL120751 is collinear with AAC82471, AAC78443, eIF4G-I_{ext}, and NP_004944, except for the Arg in the first position of AAC78443 (aa 30 using the common numbering system²), which is Pro in eIF4G-I_{ext} and AL120751 (Fig. 1). We assume this difference is due to an incomplete codon at the junction between the AAC78443 cDNA and the vector or a sequencing error. AL120751 predicts the longest eIF4G-1 protein reported to date, adding 22 aa residues to the eIF4G-I_{ext} sequence. Furthermore, it provides a new potential translational initiation site. Upstream of the AUG that encodes Met-1 of AL120751 are both in-frame and out-of-

TABLE I
Proposed names, descriptions, and assignments of hypothetical eIF4G-1 isoforms

eIF4G-1 isoform name	Starting aa no. ^a	Starting aa sequence of nascent polypeptide	Calculated mass of eIF4G-1	Calculated mass of cp _N	Assignment ^b	N terminus found
				<i>Da</i>		
eIF4G-1f	1	MNK	175,541	72,124	Band 1	N ^o AcNK
eIF4G-1e	41	MNT	171,564	68,146	Band 2	ND ^c
eIF4G-1d	88	MMI	166,669	63,252	Not found	
eIF4G-1c	89	MIP	166,538	63,121	Band 3	N ^o AcMIP
eIF4G-1b	165	MNQ	158,597	55,180	Not found	
eIF4G-1a	197	MSG	154,885	51,468	Band 4?	MSG and N ^o AcMSG

^a The numbering system is given in Fig. 1.

^b Electrophoretic bands are shown in Fig. 2.

^c Not determined because of the high mass (15,387 Da) and microheterogeneity of Peptide d1C1.

tides making up the various electrophoretic bands. Initially, we attempted to identify them by N-terminal sequence analysis. The various cp_N isoforms were separated by SDS-PAGE, blotted onto a polyvinylidene difluoride membrane, and analyzed by Edman degradation at the Macromolecular Structure Analysis Facility of the University of Kentucky. However, this failed to yield interpretable data, presumably due to blocked N termini.

We therefore turned to sequence information provided by cDNAs and ESTs. Five of the cDNA sequences (AAC82471, AL120751, AAC78443, eIF4G-I_{ext}, and NP_004944) are theoretically translated into collinear polypeptide sequences. Due to the existence of multiple AUG codons, six eIF4G-1 isoforms can be predicted that correspond to alternative translational initiation sites (Fig. 1). We termed these hypothetical isoforms eIF4G-1a, -1b, -1c, -1d, -1e, and -1f (Table I). The proteolytic peptides of each eIF4G-1 isoform can be predicted from the composite cDNA sequence (Fig. 3). It is also possible to determine the masses of peptides actually present in proteolytic digests from the various electrophoretic bands by MALDI-TOF-MS (54). Putting information on theoretical and actual peptides together, we can deduce which hypothetical isoforms are consistent with the peptide pattern of each electrophoretic band. The masses of “diagnostic” peptides, *i.e.* those that are unique to the various hypothetical eIF4G-1 isoforms, are shown in Table II.

Initial Characterization of Protein Bands in cp_N Preparations by MALDI-TOF-MS—To determine which bands present in cp_N preparations were actually related to eIF4G-1, we subjected a preparation of cp_N to electrophoresis and staining with Coomassie Blue (Fig. 4). This pattern differs from that of Fig. 2, *lane 1*, for two reasons. First, higher resolution was achieved on the preparative gel (Fig. 4) than the analytical gel (Fig. 2). Second, staining with Coomassie Blue (Fig. 4) is roughly proportional to protein quantity, but staining with silver (Fig. 2) is disproportionately strong for cp_N polypeptides (38), perhaps because of several polyglutamic acid stretches in the sequence. Protein bands were excised, digested with endoproteases, and the resulting peptides analyzed by MALDI-TOF-MS (see “Experimental Procedures”). eIF4G-1 peptides were found in *band b*, *band c* plus *band d* (partially resolved), *band e*, and *band g* only, which correspond to *Bands 1, 2, 3, and 4* in Fig. 2, respectively. These assignments, based on observed eIF4G-1 peptides, also agree with immunoblotting results (Fig. 2 as well as Refs. 14, 38, 51, and 52).

Some non-eIF4G-1 proteins present in the cp_N preparation were potentially of interest. In addition to eIF4G-1, *band c* contained peptides that matched a LINE1 reverse transcriptase homologue. *Bands d* and *n* matched the p110 and p36 subunits of eIF3, respectively (55). As noted above, several heat-shock proteins bind to eIF4G, so it is interesting to note that *band i* is the endoplasmic reticulum-resident chaperone

BiP, and *Band l* is hsp70-1. *Band j* is PABP, known to associate with eIF4G (10, 11). Finally, whereas the 46-kDa DEAD-box helicase eIF4A is known to bind eIF4G-1 at two sites in cp_C (14, 20, 21), it is interesting to note that *band k* is a 72-kDa DEAD-box protein, and *band m* is a 68-kDa RNA helicase.

Band 1 Corresponds to eIF4G-1f—To correlate hypothetical eIF4G-1 isoforms (Table I) with electrophoretic bands (Fig. 2), we examined the eIF4G-1-containing bands in more detail with four different proteolytic enzymes over several mass ranges. A typical spectrum for Band 1 (which corresponds to *band b* in Fig. 4) in the range $m/z = 500-6000$ is shown in Fig. 5. Of the tryptic peptides predicted to arise from eIF4G-1 (Fig. 3), those with masses similar to peaks observed in Fig. 5 are listed in Table III. The average deviation of observed *versus* theoretical masses was 0.19 Da. The major peaks (other than calibrants) matched predicted eIF4G-1 peptides, indicating that the sample was not grossly contaminated. Peptides below 800 Da were excluded from further consideration because of high noise levels and paucity of significant sequence information. In addition to the tryptic peptides listed in Table III, we detected in other spectra Peptides 11-12 (missed cleavage), 13-14, 16, 29-30-31, 30-31, and 37-38, some of which were either singly or doubly oxidized in Met residues (see below).

Diagnostic tryptic peptides for Band 1 are shown in Fig. 6A (*upper panel*). Band 1 has peaks at $m/z = 4534.7$ and 4550.8 , which are within 0.5 Da of the calculated masses of Peptide 2 (Table II) and the Met-oxidized form of Peptide 2 (+16.0 Da). This “one-Met signature” agrees with the presence of one Met residue in Peptide 2 (Fig. 3 and Table II) and contributes to the confidence of this assignment. Other peaks in Fig. 6A for Band 1 at $m/z = 4629.7$, 4645.9 , and 4662.1 are within 0.6 Da of the predicted masses of Peptide 5 (Table II) and its derivatives containing either one or two oxidized Met residues, respectively. This “two-Met signature” agrees with the presence of two Met residues in Peptide 5 (Fig. 3 and Table II). Finally, the peak at $m/z = 4728.6$ matches Peptide 28-29 (Fig. 3 and Table III). In contrast, the spectrum of Band 2 contains peaks corresponding to Peptide 5 but not Peptide 2 (Fig. 6A, *lower panel*), nor were peaks corresponding to Peptides 2 or 5 observed in Bands 3 or 4 (not shown). Peptide 5 can come from eIF4G-1e or eIF4G-1f, but not from eIF4G-1a, -1b, -1c, or -1d (Fig. 3 and Table II). However, Peptide 2 can *only* come from eIF4G-1f (or an undescribed, longer form not predicted from the cDNA sequence). Thus, the presence of Peptide 2 in Band 1 but not in Bands 2, 3, or 4 is consistent with eIF4G-1f being present only in Band 1.

Confirmation that eIF4G-1f is in Band 1 was obtained with endoprotease Lys-C. When this enzyme was used with Band 1, we detected Peptides k2, k4, k7, k12, k13, k15, and k21 (Fig. 3). A diagnostic peak at $m/z = 17,770$ was observed in Band 1 but not Band 2 (Fig. 6B). Peptide k2 has a predicted, unmodified mass of 17,695 Da (Table II), which is 75 Da less than the peak

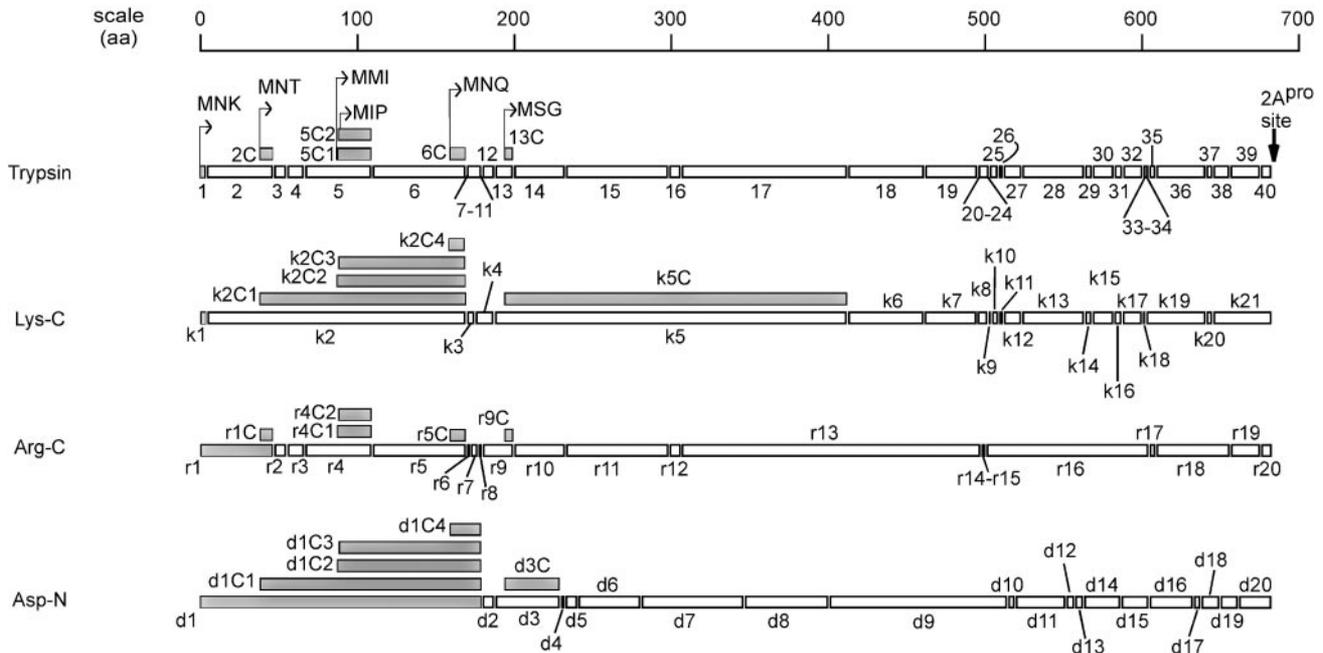


FIG. 3. Predicted peptides from digestion of hypothetical eIF4G-1 isoforms with trypsin, Arg-C, Lys-C, and Asp-N. Peptides shown cover the cp_N region of eIF4G-1 only. Tryptic peptides are numbered from N to C terminus using normal Arabic numerals (Peptides 1, 2, etc.). Lys-C peptide names are preceded by “k” (Peptides k1, k2, etc.). Arg-C peptide names are preceded by “r” (Peptide r1, r2, etc.). Asp-N peptide names are preceded by “d” (Peptide d1, d2, etc.). Peptides that form the N terminus of one of the hypothetical eIF4G-1 isoforms (Table I) are shaded. The suffixes C, C1, C2, etc. denote truncated N-terminal peptides (e.g. Peptide 2C is the C-terminal portion of Peptide 2). The initial aa sequences of predicted N-terminal peptides, prior to any posttranslational modification, are shown above the corresponding tryptic peptides (MNK, MNT, etc.).

TABLE II
Predicted masses of diagnostic peptides that distinguish between hypothetical forms of eIF4G-1

Peptide name ^a	Diagnostic of ^b	Starting aa no. ^c	Ending aa no. ^c	Calculated mass of unmodified peptide ^d	Number of Met residues ^e
				<i>Da</i>	
1	eIF4G-1f	1	3	392.20	1
k1	eIF4G-1f	1	3	392.20	1
r1	eIF4G-1f	1	48	4908.6	2
d1	eIF4G-1f	1	183	19,278	5
2	eIF4G-1f	4	48	4535.1	1
k2	eIF4G-1f	4	174	17,695	4
2C	eIF4G-1e	41	48	930.45	1
k2C1	eIF4G-1e	41	174	14,091	4
r1C	eIF4G-1e	41	48	930.45	1
d1C1	eIF4G-1e	41	183	15,300	4
5	eIF4G-1e or -1f	66	110	4630.3	2
5C1	eIF4G-1d	88	110	2516.9	2
k2C2	eIF4G-1d	88	174	9196.4	3
r4C1	eIF4G-1d	88	110	2516.9	2
d1C2	eIF4G-1d	88	183	10,406	3
5C2	eIF4G-1c	89	110	2384.17	1
k2C3	eIF4G-1c	89	174	9065.2	2
r4C2	eIF4G-1c	89	110	2384.17	1
d1C3	eIF4G-1c	89	183	10,275	2
6C	eIF4G-1b	165	174	1123.59	1
k2C4	eIF4G-1b	165	174	1123.59	1
r5C	eIF4G-1b	165	175	1279.69	1
d1C4	eIF4G-1b	165	183	2333.81	1
13C	eIF4G-1a	197	201	521.25	1
k5C	eIF4G-1a	197	426	23,754	4
r9C	eIF4G-1a	197	201	521.25	1
d3C	eIF4G-1a	197	232	3575.0	1

^a See legend of Fig. 3 for peptide nomenclature.

^b Found only in the indicated hypothetical isoform of eIF4G-1 (Table I).

^c The numbering system and aa sequences of these peptides are given in Fig. 1.

^d Masses in the range of 1–2500 Da are reported as monoisotopic values, whereas those in the range of 2500–30,000 Da are reported as average values. The masses given do not reflect oxidation of Met residues to the sulfoxide, which adds 16.00 Da.

^e Number of Met residues in the predicted peptide.

observed in Fig. 6B. Since Peptides 2 and 5 exist primarily in the oxidized forms (Fig. 6A), it is likely that the peak at 17,770 consists of a mixture of Peptide k2 and derivatives oxidized at

one to four Met residues (calculated masses: 17,711, 17,727, 17,743, and 17,759 Da). The observed peak is within 0.06% of the latter calculated mass. These results indicate that Band 1

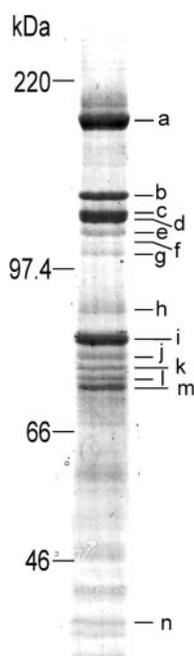


FIG. 4. Identification of eIF4G-1 and other affinity-purified polypeptides by matching peptide masses to data bases. 2A protease-treated K562 lysate was subjected to m^7 GTP-Sepharose affinity chromatography as described under "Experimental Procedures." Bound proteins were then resolved by preparative gradient gel SDS-PAGE and stained with Coomassie Blue. Protein bands (arbitrarily labeled *a* thru *n*) were excised, digested with trypsin, and subjected to MALDI-TOF-MS. The spectra were automatically matched to the NCBI protein data base by the Auto-MS Fit program, as described under "Experimental Procedures." Identities and confidence scores (MOWSE; Ref. 68) were generated for each band. The human proteins that best matched the indicated bands are: *band a*, hypothetical protein, GenBankTM protein accession number T17345, MOWSE = 2.84×10^3 ; *band b* (same as Band 1 of Fig. 2), eIF4G-1, GenBankTM AAC78443, MOWSE = 6.85×10^2 ; *band c* (same as Band 2 of Fig. 2), LINE1 reverse transcriptase homologue, GenBankTM P08547, MOWSE = 3.19×10^2 ; subsequent analysis of this band revealed eIF4G-1 peptides as well; *band d*, p110 subunit of eIF3, GenBankTM NP_003743, MOWSE = 2.61×10^5 ; *band e* (same as Band 3 of Fig. 2), no protein matched by Auto-MS Fit, but subsequent analysis revealed eIF4G-1 peptides; *band f*, no protein matched by Auto-MS Fit; *band g* (same as Band 4 of Fig. 2), no protein matched by Auto-MS Fit, but subsequent analysis revealed eIF4G-1 peptides; *band h*, no protein matched by Auto-MS Fit; *band i*, BiP, GenBankTM AAF13605, MOWSE = 1.09×10^4 ; *band j*, PABP, GenBankTM NP_002559, MOWSE = 8.29×10^3 ; *band k*, DEAD-box protein p72, GenBankTM NP_006377, MOWSE = 1.63×10^3 ; *band l*, HSP70, GenBankTM NP_005336, MOWSE = 3.22×10^5 ; *band m*, RNA helicase p68, GenBankTM NP_004387, MOWSE = 5.94×10^3 ; *band n*, p36 subunit of eIF3, GenBankTM NP_003748, MOWSE = 7.82×10^4 .

contains Peptide k2, which can only arise from eIF4G-1f (or a hypothetical larger form), whereas Bands 2, 3, and 4 do not contain Peptide k2.

Unfortunately, the tryptic and Lys-C peptide predicted to represent the extreme N terminus of eIF4G-1f, MNK, is too small to be unambiguously identified by our approach. Thus, we cannot conclude from the data in Fig. 6, A and B, whether Peptides 2 and k2 were contributed by eIF4G-1f or a hypothetical larger form. We therefore turned to a different enzyme, Arg-C, which produced a diagnostic peak at $m/z = 4819.3$ (Fig. 6C). The mass of Peptide r1, which is predicted to result from Arg-C digestion of eIF4G-1f, is 4908.6 Da (Table II). However, the mass of a derivative of Peptide r1 in which the N-terminal Met is removed and the resulting N-terminal Asp is N^α -acetylated (+42.0 Da) is 4819.4, within 0.1 Da of the observed peak. This agrees with our observation that the same band was refractory to analysis by Edman degradation (see above). A

second peak at 4835.6 matches the mass of Met-oxidized, N^α -acetylated Peptide r1 in which the N-terminal Met is removed. This one-Met signature is consistent with the postulated removal of the N-terminal Met-1 and oxidation of Met-41 (Fig. 1). The peaks at $m/z = 4819.3$ and 4835.6 were observed in Band 1 but not Bands 2–4 (data not shown), which agrees with the hypothesis that eIF4G-1f is in Band 1 but not in the other bands.

To obtain additional evidence for eIF4G-1f in Band 1, we digested with Asp-N. A major peak with $m/z = 19,371$ was observed in Band 1 (Fig. 6D, upper panel) but not Band 2 (Fig. 6D, lower panel), Band 3, or Band 4 (data not shown). This is similar to the mass of Peptide d1, which is 19,278 Da (Table II). Peptide d1 contains five Met residues, although based on the data with Arg-C (Fig. 6C), the N-terminal Met may have been removed on at least some of the polypeptides. Subtracting one Met residue, adding an acetyl group, and oxidizing four Met residues gives a mass of 19,253. Although this does not agree exactly with the m/z of the observed peak, it is within 0.6% of this value. As noted above, exact matching of predicted and observed peptide masses can be compromised by microheterogeneity (partial removal of N-terminal Met, acetylation, oxidation). Nonetheless, the peak at $m/z = 19,371$ is at least consistent with the presence of Peptide d1 in Band 1, which is, in turn, indicative of eIF4G-1f.

Band 2 Represents a Mixture of Proteins, Including eIF4G-1e—Three lines of evidence suggest that Band 2 is a mixture of proteins. First, the band is broader than Band 1, Band 3, or other non-eIF4G bands in Coomassie Blue-stained gels (Fig. 4, *band c* + *band d*), suggesting that proteins are unresolved. Second, Band 2 is, in fact, partially resolved into a doublet or triplet in some gels (data not shown). Third, the Auto MS-Fit program identified a LINE1 reverse transcriptase homologue and an eIF3 subunit in *bands c* and *d* (Fig. 4), despite the fact that they also contain numerous eIF4G-1 peptides (see below). Nonetheless, we were able to obtain structural information about the isoform of eIF4G-1 present in Band 2.

Of the tryptic peptides in cp_N with masses >800 Da, we detected Peptides 3, 4, 5, 13, 30, and 38 in Band 2, either as single peptides or composite peptides resulting from missed cleavages. A set of peaks with $m/z = 4630.0$, 4646.3, and 4662.1 was observed for Band 2 (Fig. 6A, lower panel), corresponding to the two-Met signature of Peptide 5 (Table II). The presence of Peptide 5 is diagnostic of a protein being initiated with a Met upstream of eIF4G-1d, *i.e.* either eIF4G-1e or eIF4G-1f (Fig. 3). We also detected Peptide 13, which indicates the eIF4G-1 isoform present is initiated upstream of eIF4G-1a (data not shown). Yet Band 2 lacks Peptide 2 (Fig. 6A, lower panel), indicating it does not contain eIF4G-1f. It is clear from the results with Band 1 that Peptide 2 would have been detected if present (Fig. 6A, upper panel). The presence of Peptide 5 but absence of Peptide 2 indicates that Band 2 contains eIF4G-1e.

With Lys-C, we detected Peptides k3, k4, k5, k7, k8, k9, k10, k13, k14, k15, k16, k18, k19, k20, and k21 as either single or composite peptides (Fig. 3). Unfortunately, none of these peptides gave insight into the form of eIF4G-1 present, other than it was upstream of eIF4G-1a. Specifically, no evidence for the diagnostic peptide k2C1 was seen.

Digestion with Asp-N revealed a peak at $m/z = 15,387$ in Band 2 (Fig. 6D, lower panel). This m/z value is similar to that of Peptide d1C1 (15,300), which represents the N terminus of eIF4G-1e (Table II). Peptide d1C1 contains four Met residues. If each Met were oxidized, it would add 64 Da to the mass of this peptide, bringing the m/z value to 15,364 Da. The resulting difference between the observed and calculated values is

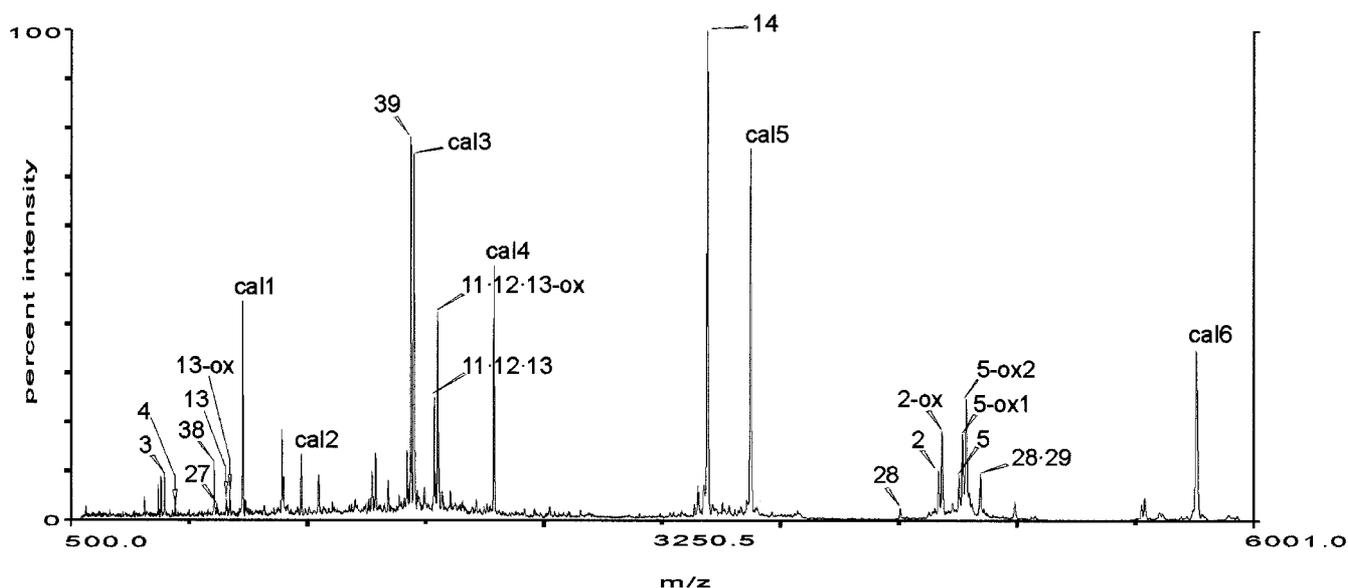


FIG. 5. MALDI-TOF-MS analysis of tryptic peptides from Band 1 of a cp_N preparation. Band 1 (see Fig. 2) from a preparative gradient gel (same as band *b* in Fig. 4) was subjected to trypsin digestion and MALDI-TOF-MS as described under "Experimental Procedures." Peaks attributable to eIF4G-1 peptides are indicated using the numbering system in Fig. 3. Composite peptides resulting from missed tryptic cleavages are shown with a center dot, e.g. Peptide 11·12·13. Peptides in which Met residues are oxidized have the suffixes ox, ox1, and ox2 (see Table III). Peaks resulting from peptide standards used for internal calibration are indicated as cal1, cal2, etc. For each peak, *m/z* values are compared with predicted mass values in Table III.

TABLE III
Matching of peaks from Band 1 in the range of *m/z* = 800–6000 to theoretical eIF4G-1 tryptic peptides

Peptide name ^a	Observed peak ^b	Predicted mass	Difference
	<i>m/z</i>		
Peptide 3	934.43	934.45	0.02
Peptide 4	971.46	971.49	0.03
Peptide 38	1165.67	1165.67	0.00
Peptide 27	1177.60	1177.61	0.01
Peptide 13	1221.58	1221.57	0.01
Peptide 13-ox	1237.68	1237.58	0.10
Peptide 39-acr	2078.04	2078.01	0.03
Peptide 11 · 12 · 13	2187.18	2187.08	0.10
Peptide 11 · 12 · 13-ox	2203.21	2203.08	0.13
Peptide 14	3458.9	3458.8	0.1
Peptide 31 · 32 · 33 · 34	3642.9	3643.0	0.1
Peptide 28	4355.9	4356.5	0.6
Peptide 2	4534.7	4535.1	0.4
Peptide 2-ox	4550.8	4551.1	0.3
Peptide 5	4629.7	4630.3	0.6
Peptide 5-ox1	4645.9	4646.3	0.4
Peptide 5-ox2	4662.1	4662.3	0.2
Peptide 28 · 29	4728.6	4728.9	0.3

^a Peptide names are given in Fig. 3, but with the following exceptions: ox indicates a single Met oxidation; ox1 means a peptide containing two Met residues but with only one oxidized; ox2 means both are oxidized; acr means Cys acrylamidization; and center dots designate composite peptides resulting from missed tryptic cleavages. Peptide sequences are given in Fig. 1.

^b Data are from Fig. 5.

0.1%, which is considerably less than the difference for any other peptide predicted from the six hypothetical isoforms of eIF4G-1. This peak was not observed in Band 1 (Fig. 6D, upper panel), Band 3, or Band 4 (data not shown), providing additional evidence that eIF4G-1e is present only in Band 2.

Band 3 Contains eIF4G-1c—The only tryptic peptides with masses >800 Da detected in Band 3 were Peptides 5C2, 12, 14, and 15. This is considerably fewer than for Band 1, presumably due to the lower amount of protein typically seen in this band (Fig. 2). A diagnostic peak was found in Band 3 at *m/z* = 2426.20 that was not present in Band 1 (Fig. 7A, lower panel). This is 42.03 Da more than the mass of Peptide 5C2 (Table II), which is the postulated N terminus of eIF4G-1c. This peak is

therefore consistent with an N^α-acetylated form of Peptide 5C2, which would agree with our inability to obtain information for this band from Edman degradation (see above). A second peak was detected for Band 3 at *m/z* = 2442.23 (Fig. 7A, lower panel). This one-Met signature agrees with the presence of one Met residue in Peptide 5C2 (Table II) and suggests that, unlike eIF4G-1f, the N-terminal Met is *not* removed before acetylation of eIF4G-1c. Alternatively, the nascent polypeptide may have been eIF4G-1d from which Met-88 was removed, followed by N^α-acetylation of Met-89 (Fig. 1).

Additional evidence for diagnostic peptides in Band 3 was sought using Lys-C. Peptides k2C3, k3, k4, k5, k7, k13, k14, k15, k16, k17, k18, k19, k20, and k21 were detected as either single or composite peptides. Diagnostic peptides were found at *m/z* = 9124.1 and 9139.5 (Fig. 7B, lower panel). These are within 1 Da of the predicted masses for the N^α-acetylated form of Peptide k2C3 containing either one (9123.3 Da) or two (9139.3 Da) oxidized Met residues. This two Met signature is consistent with the presence of two Met residues in Peptide k2C3 (Table II). The peak at *m/z* = 8883.3 of Band 1 corresponds to doubly charged Peptide k2 (predicted *m/z* = 8847.9 for the unoxidized peptide; see Fig. 6B). Its breadth is presumably due to microheterogeneity from multiple oxidized Met residues.

Finally, we analyzed the Asp-N peptides of Band 3 (Fig. 7C). A unique peptide at *m/z* = 10,317 matched the predicted mass for N^α-acetylated Peptide d1C3 (10,317 Da). Similarly, the observed peaks at *m/z* = 10,333 and 10,349 matched the masses of the singly and doubly oxidized forms of this acetylated peptide (predicted at 10,333 and 10,349 Da). These peptides provide additional evidence for the presence of eIF4G-1c in Band 3.

In a separate experiment, MS fragmentation data (Fig. 8) confirmed our assignment of the tryptic peptide detected at *m/z* = 2426.20 in Fig. 7A as N^α-acetylated Peptide 5C2 (cDNA-derived sequence given in Fig. 1).

Band 4 Represents a Mixture of Proteins That May Include eIF4G-1a—The amounts of eIF4G-related material in Band 4 were insufficient to make a definitive assignment. However,

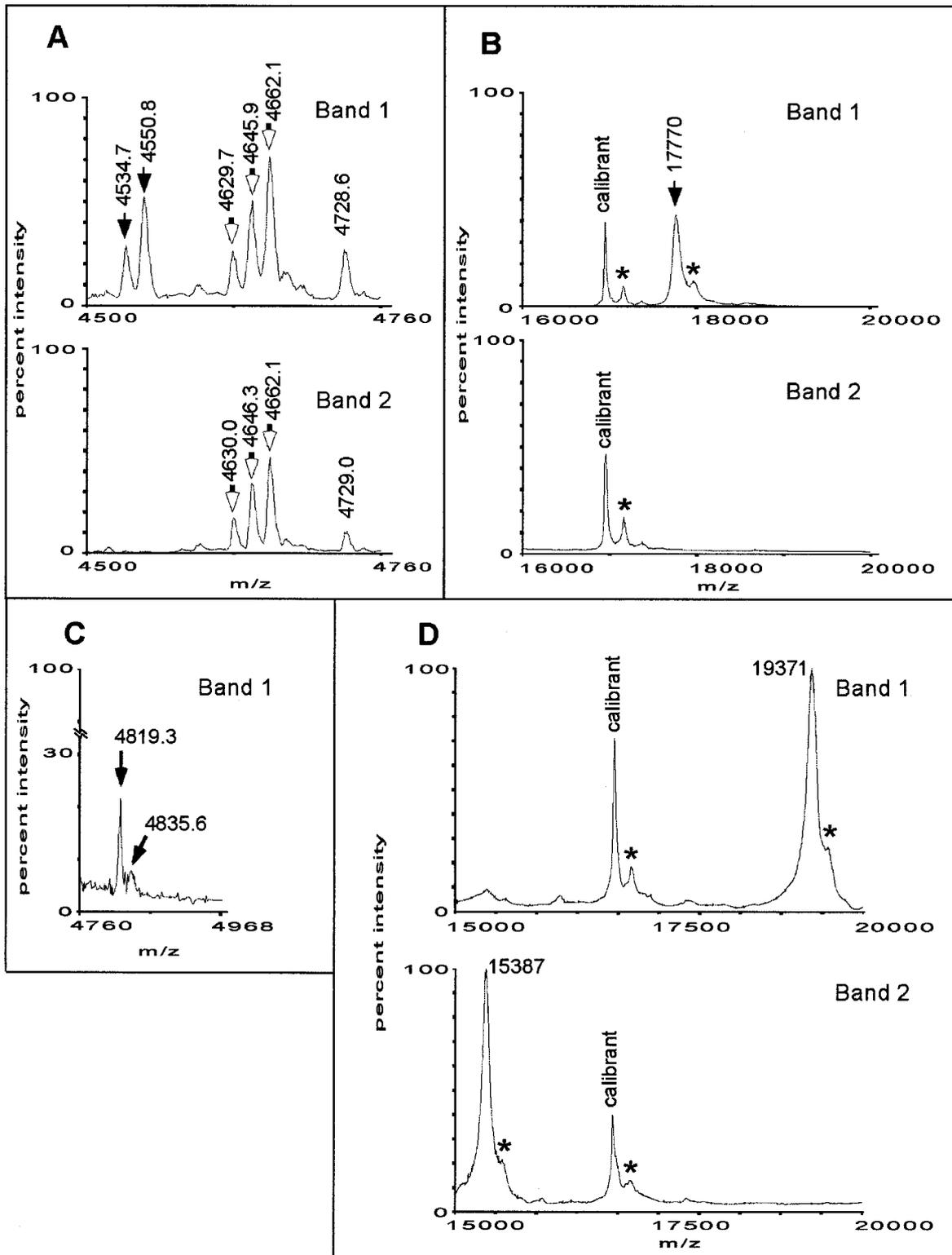


FIG. 6. Identification of diagnostic peptides in Bands 1 and 2. Peptides produced from *Bands 1* and *2* (Fig. 2) with four endoproteases were analyzed by MALDI-TOF-MS. Peaks marked with an *asterisk* arise from MALDI matrix adduction (69). *A*, peaks for tryptic peptides from Bands 1 and 2. *Closed arrows* (upper panel) indicate unmodified and Met-oxidized forms of Peptide 2. *Open arrows* (upper and lower panels) indicate unmodified and Met-oxidized forms of Peptide 5. *B*, peaks for Lys-C peptides from Bands 1 and 2. The *closed arrow* (upper panel) indicates Peptide k2. *C*, peaks for Arg-C peptides from Band 1. The *arrows* denote the N^{α} -acetylated, des-Met derivative of Peptide r1 and the Met-oxidized form. *D*, Asp-N peptides from Bands 1 and 2. Peptide d1 is present only in Band 1 (upper panel), while Peptide d1C1 is present only in Band 2 (lower panel). An internal calibrant peak for the protein standard horse apomyoglobin (16,952 Da) is present in both upper and lower panels of *B* and *D*.

peaks were detected that suggest the presence of eIF4G-1a. Specifically, we detected a peak in tryptic digests at 538.24 Da that is within 1 Da of Peptide 13C with one Met oxidized (537.25 Da). We also detected a peak in tryptic digests at

580.25 that is within 1 Da of the N^{α} -acetylated, Met-oxidized form of Peptide 13C (579.26 Da). In Asp-N digests, we detected a peak at $m/z = 3574.1$ that is within 1 Da of Peptide d3C (3575.0 Da; Table II). We also detected a peak at $m/z = 3616.1$

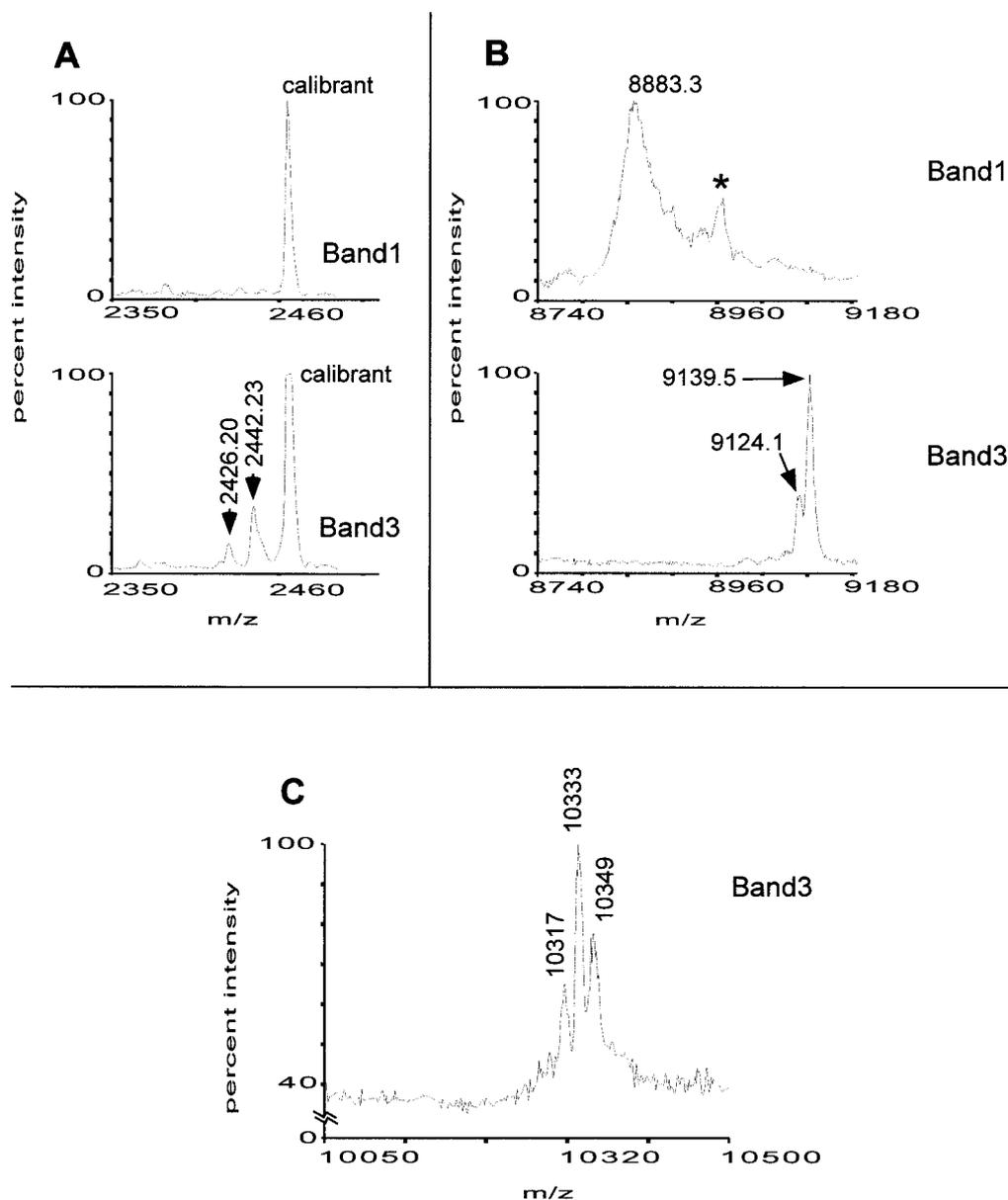


FIG. 7. **Identification of diagnostic peptides in Band 3.** A, peaks for tryptic peptides from Bands 1 and 3. Arrows (lower panel) indicate N^α -acetylated Peptide 5C2 and its Met-oxidized derivative. A calibrant peak (ACTH clip 18–39, calculated mass 2465.20 Da) is present in both samples. B, peaks for Lys-C peptides from Bands 1 and 3. Arrows (lower panel) denote Met-oxidized derivatives of N^α -acetylated Peptide k2C3, which are unique to Band 3. C, peaks for Asp-N peptides from Band 3. The group of peaks match masses of Met-oxidized and unoxidized forms of N^α -acetylated Peptide d1C3.

that is similar to N^α -acetylated Peptide d3C (3617.0 Da). All of these were only slightly above the background, making assignment of structure uncertain. Band 4 often occurs as a diffuse doublet, indicating heterogeneity, and immunoblotting demonstrates that the eIF4G-1 isoforms in Band 4 are much less abundant than in Bands 1–3 (Fig. 2).

DISCUSSION

The data presented here permit the identification of several eIF4G-1 primary structures and their assignment to electrophoretic bands. Band 1 consists of a novel isoform, here termed eIF4G-1f. This is also the most abundant form, as measured by Coomassie Blue staining and immunoreactivity to both anti-Peptide 7 and anti-Peptide 10 antibodies. Using trypsin or Lys-C, we observed Peptides 2 and k2, which can only arise from eIF4G-1f or a hypothetical protein initiated upstream of it. With Arg-C we detected the N-terminal peptide, Peptide r1, and its mass indicated that it was modified in two ways: the

N-terminal Met was removed and the resulting Asn was N^α -acetylated. Asp-N produced Peptide d1, in agreement with the assignment of eIF4G-1f to Band 1, but it was too large (19,371 Da) to allow determination of the exact N-terminal structure. Thus, all of the data from Band 1 are consistent with its identity as eIF4G-1f. This isoform of eIF4G-1 was not predicted from any of the previously reported cDNA sequences. It is the longest isoform, with 22 additional aa residues at the N terminus compared with eIF4G-I_{ext} (Fig. 1), the longest isoform previously postulated.

Several observations point to Band 2 being a mixture of proteins. Nonetheless, three types of data demonstrate that Band 2 contains eIF4G-1e. First, it reacts with anti-Peptide 10 antibodies, which are directed against an epitope in only eIF4G-1e and eIF4G-1f. Second, tryptic digests contain Peptide 5 but not Peptide 2, despite the fact that Peptide 2 is readily detectable in Band 1. Third, Asp-N digests contain Peptide

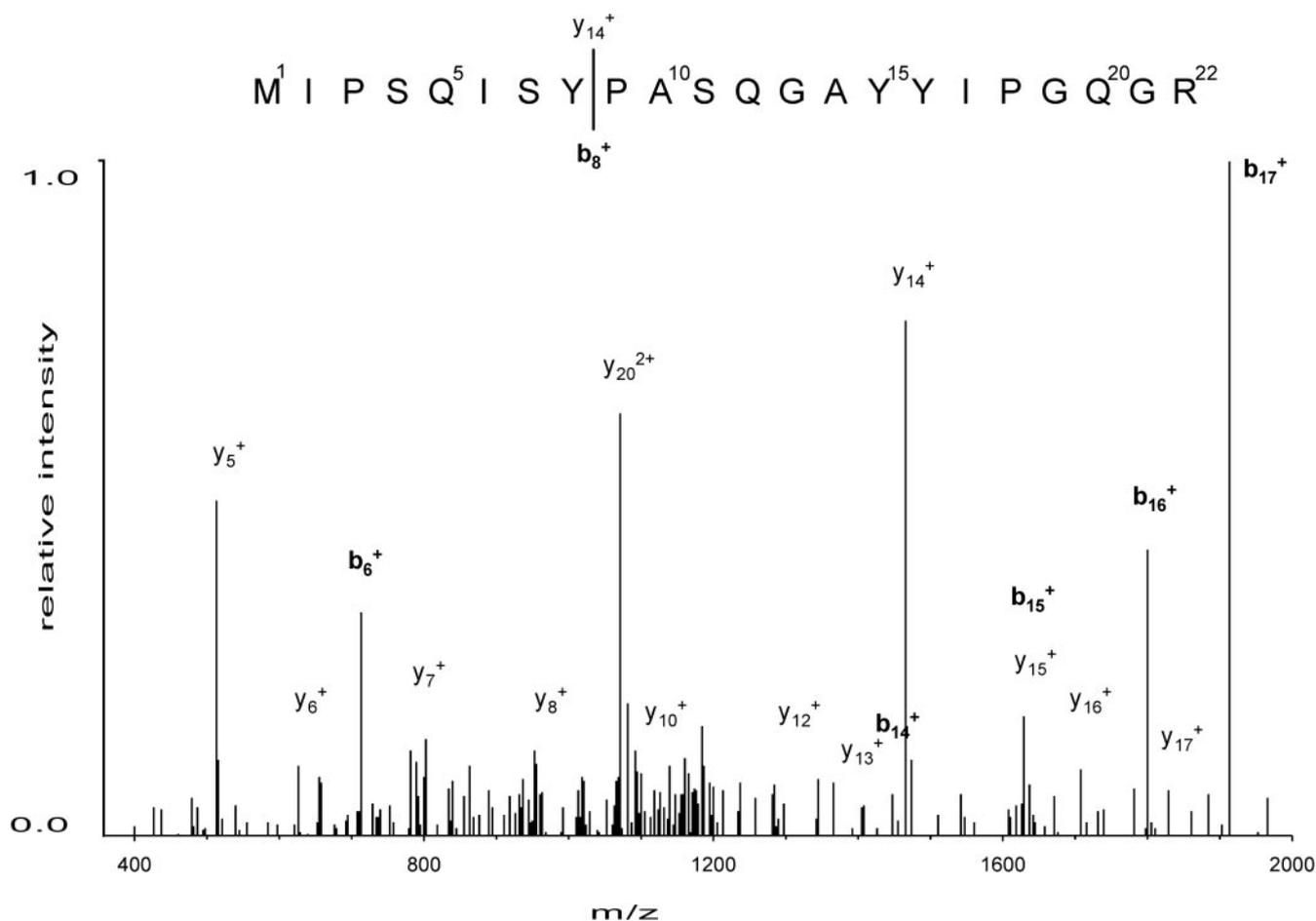


FIG. 8. LC-ESI/MS spectrum of Peptide 5C2 in Band 3. Fragmentation analysis of N^α -acetylated Peptide 5C2 shows ions of the b (N-terminal) and y (C-terminal) series. The y fragments (in the range y_3^+ to y_{20}^+) all match their predicted masses, whereas the detected b fragments give values that are shifted by +42 Da, indicating an acetylated Met residue at position 1.

d1C1 in Band 2 but not Band 1. Due to the size and microheterogeneity of Peptide d1C1, however, we were unable to determine the exact structure at the N terminus. eIF4G-1e is the isoform predicted from several of the cDNAs published before the present report (44–46), although its existence had not actually been demonstrated. Comparison of eIF4G-1e to eIF4G-1f by immunoreactivity with two different antibodies indicates that eIF4G-1e is less abundant (Fig. 2 and data not shown).

Both peptide matching and fragmentation data support the assignment of Band 3 as a N^α -acetylated form of eIF4G-1c. In digests with trypsin, Lys-C, and Asp-N, we detected the predicted N-terminal peptides and their Met-oxidation products, *viz.* N^α -acetylated Peptide 5C2, N^α -acetylated Peptide k2C3, and N^α -acetylated Peptide d4C3, respectively. These peptides were found in Band 3 but not other bands. Finally, the sequence of N^α -acetylated Peptide 5C2 was established by MS fragmentation. Based on the agreement between Coomassie Blue staining and immunoreactivity in comparison with Band 1, Band 3 appears to be homogeneous for eIF4G-1c. This form of eIF4G-1 is clearly present in K562 cells but is less abundant than eIF4G-1f and -1e (Fig. 2). eIF4G-1c is another novel isoform, not previously predicted from cDNA sequences.

Additional forms of eIF4G-1 may exist but were too low in abundance to be identified in the present study. Several peptides unique to eIF4G-1a were detected in Band 4, but we do not judge the evidence to be compelling. It is interesting to note that many bands *besides* Bands 1–4 react with anti-Peptide 7 antibodies (Fig. 2, lane 4). One might consider them to repre-

sent nonspecific binding of the antibodies to non-eIF4G-1 proteins except for one additional feature: they are not present in a preparation pretreated with 2A protease (Fig. 2, lane 3). The 2A protease of entero- and rhinoviruses is quite specific for a consensus aa sequence (38, 52, 56, 57). A few cellular proteins other than eIF4G are cleaved by 2A proteases (58–60), but two-dimensional electrophoresis reveals that the overwhelming majority of cellular proteins are unaffected by picornavirus infection (61). The likelihood that a non-eIF4G-1 protein would both react with anti-Peptide 7 antibodies and also be a substrate for 2A protease is quite remote. Thus, the numerous weak bands detected in anti-Peptide 7 immunoblots (Fig. 2, lane 4) may represent additional eIF4G-1 isoforms.

While there may well be additional isoforms of eIF4G-1, it is unlikely that there are any larger than eIF4G-1f. Even if longer cDNAs are found, three considerations make it unlikely that they will encode proteins containing an extension of the eIF4G-1f polypeptide sequence. First, we have shown that the structure of the major form (eIF4G-1f) is initiated from the AUG that is furthest upstream of the known cDNAs. Second, there are termination codons, both in-frame and out of frame, upstream of this AUG. It is possible that other forms of eIF4G-1 may be encoded by mRNAs arising by alternative splicing that contain different N-terminal aa sequences, *e.g.* AAC78442 (Fig. 1). However, we detected no peaks corresponding to either eIF4G-3 (4) or AAC78442. Third, eIF4G-1f is the principle constituent of Band 1, which is the slowest migrating immunoreactive band (Fig. 2). Drawing conclusions about eIF4G structure from electrophoretic mobility is complicated by the fact that the migration of eIF4G-1

on SDS-PAGE is aberrantly slow. The slowest eIF4G-1 isoform migrates at 220 kDa and the fastest, at 205 kDa (62), yet the calculated molecular masses range from 176 to 155 kDa (Table I). Thus, something besides the additional aa sequence reported here must account for the disparity. Despite the aberrant mobility, we observed that the *order* of bands follows the molecular mass of the isoforms, *e.g.* eIF4G-1f is both the largest form (Table I) and the slowest eIF4G band (Fig. 2), *etc.* Larger forms of eIF4G-1 may exist but not be purified by the method used in this study; it is worth noting that our method of affinity purification on m⁷GTP-Sepharose would exclude any hypothetical eIF4G-1 isoforms that lack an eIF4E-binding site such as the eIF4G homologue eIF4G-2 (3, 7, 8).

Although it is clear that the various isoforms of eIF4G-1 contain different N termini, the origin of these proteins cannot be determined from the results presented. Several formal possibilities can be considered. The first is alternative translation initiation of the same mRNA by leaky scanning. The optimal consensus sequence for initiation in animals is A/GXXAUGG (63). A "strong" initiation codon is considered to be one containing either the purine at -4, the G at +1, or both. The corresponding sequences for the various hypothetical eIF4G-1 isoforms are: eIF4G-1f, AAAAUGA; eIF4G-1e, CAAAUGA; eIF4G-1d, GUAAUGA; eIF4G-1c, AUGAUGA; eIF4G-1b, UUGAUGA; and eIF4G-1a, AUCAUGU. Thus, the strong initiation codons are in mRNAs for eIF4F-1f, -1d, -1c, and -1a, perhaps explaining the absence of eIF4G-1b. The second possibility for the multiplicity of eIF4G-1 isoforms is internal initiation of translation. Two separate sequences derived from the 5'-portion of eIF4G-1 mRNA can function as internal ribosome entry sites in cultured cells (45, 64, 65). A single mRNA corresponding to the longest cDNA reported to date (this report; 5510 nt) could be initiated from the 5'-end, giving rise to eIF4G-1f, and internally, giving rise to eIF4G-1c. The nucleotide sequence representing the 5'-untranslated region of an mRNA that would encode eIF4G-1d has already been shown to direct internal initiation of translation (45). This may provide a mechanism for eIF4G-1c expression, allowing for the N-terminal processing of the nascent polypeptide (removal of Met-88 and N^α-acetylation of Met-89). Finally there may be several mRNAs differing in their 5'-terminal sequences that encode different eIF4G-1 isoforms, some of which exclude upstream AUG codons. These could arise from alternative splicing (*e.g.* NP_004944 and AAC78442) or alternative promoter usage. Resolution of this question will require an investigation of structure of eIF4G-1 mRNA(s) present in the cell and their translational properties. The fact that we never observed peptides diagnostic for slower bands in faster bands argues against the faster bands representing proteolytic breakdown products, or polypeptides missing internal exons. The same conclusion can be drawn from the absence of any bands running faster than Bands 1 and 2 that bind anti-Peptide 10 antibodies (Fig. 2).

We found a surprisingly large number of ligands co-purifying with cp_N (Fig. 4). This type of analysis does not indicate whether these ligands are bound to the cp_N fragment of all eIF4G-1 isoforms or to only a subset. These ligands include PABP and hsp70, which is consistent with previous studies showing these to be eIF4G-binding proteins (10, 11, 26). Less expected is the presence of the p110 and p36 subunits of eIF3. It is known that 11-subunit eIF3 has a high affinity for eIF4G, but the binding site has been localized in cp_C, not cp_N (14, 19). The association of eIF4G with the p110 and p36 subunits may mean that there are additional points of attachment in cp_N. The presence of a putative DEAD-box protein and an RNA helicase in cp_N is intriguing, since eIF4A, a different member of

the DEAD-box RNA helicase family, is well established to bind eIF4G-1 at two distinct sites in cp_C (14, 20, 21).

At present we have not determined whether there are differences in the biochemical properties of the different eIF4G-1 isoforms. Since eIF4G is a protein that binds an unusually large number of protein and RNA ligands, the presence of different N-terminal sequences may direct the binding of isoform-specific ligands. The various eIF4G isoforms may have different affinities for ribosomes or initiation factors that link them to ribosomes (*e.g.* eIF3), opening the possibility that eIF4G-1 isoforms participate in mRNA recruitment to different extents. The only study explicitly comparing eIF4G-1 isoforms showed that transfection of cDNAs expressing eIF4G-1e and eIF4G-1a gave similar growth efficiencies in soft agar, an assay for malignant transformation (66). Yet no *in vivo* expression studies have been performed to date on eIF4G-1f, the longest and most abundant form in K562 cells.

Acknowledgments—We gratefully acknowledge the Deutsches Ressourcenzentrum für Genomforschung GmbH and the Deutsches Krebsforschungszentrum, Germany, Heidelberg for cDNA clone DKFZp762O191 and the LSUHSC-S Research Core Facility.

REFERENCES

- Hershey, J. W. B., and Merrick, W. C. (2000) in *Translational Control of Gene Expression* (Sonenberg, N., Hershey, J. W. B., and Mathews, M. B., eds) pp. 33–88, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Yan, R., Rychlik, W., Etchison, D., and Rhoads, R. E. (1992) *J. Biol. Chem.* **267**, 23226–23231
- Imataka, H., Olsen, H. S., and Sonenberg, N. (1997) *EMBO J.* **16**, 817–825
- Gradi, A., Imataka, H., Svitkin, Y. V., Rom, E., Raught, B., Morino, S., and Sonenberg, N. (1998) *Mol. Cell. Biol.* **18**, 334–342
- Yan, R., and Rhoads, R. E. (1995) *Genomics* **26**, 394–398
- Clark, B. F. C., Grunberg-Manago, M., Gupta, N. K., Hershey, J. W., Hinnebusch, A. G., Jackson, R. J., Maitra, U., Mathews, M. B., C., M. W., Rhoads, R. E., Sonenberg, N., Spremulli, L. L., Trachsel, H., and Voorma, H. O. (1996) *Biochimie (Paris)* **78**, 1119–1122
- Yamanaka, S., Poksay, K. S., Arnold, K. S., and Innerarity, T. L. (1997) *Genes Dev.* **11**, 321–333
- Levy-Strumpf, N., Deiss, L. P., Berissi, H., and Kimchi, A. (1997) *Mol. Cell. Biol.* **17**, 1615–1625
- Aragon, T., de la Luna, S., Carrasco, L., Ortin, J., and Nieto, A. (2000) *Mol. Cell. Biol.* **20**, 6259–6268
- Tarun, S. Z., and Sachs, A. B. (1996) *EMBO J.* **15**, 7168–7177
- Le, H., Tanguay, R. L., Balasta, M. L., Wei, C. C., Browning, K., Metz, A. M., Goss, D. J., and Gallie, D. R. (1997) *J. Biol. Chem.* **272**, 16247–16255
- Piron, M., Vende, P., Cohen, J., and Poncet, D. (1998) *EMBO J.* **17**, 5811–5821
- Vilela, C., Velasco, C., Ptushkina, M., and McCarthy, J. E. (2000) *EMBO J.* **19**, 4372–4382
- Lamphear, B. J., Kirchweger, R., Skern, T., and Rhoads, R. E. (1995) *J. Biol. Chem.* **270**, 21975–21983
- Mader, S., Lee, H., Pause, A., and Sonenberg, N. (1995) *Mol. Cell. Biol.* **15**, 4990–4997
- Hershey, P. E. C., McWhirter, S. M., Gross, J. D., Wagner, G., Alber, T., and Sachs, A. B. (1999) *J. Biol. Chem.* **274**, 21279–21304
- Fortes, P., Inada, T., Preiss, T., Hentze, M. W., Mattaj, I. W., and Sachs, A. B. (2000) *Mol. Cell* **6**, 191–196
- McKendrick, L., Thompson, E., Ferreira, J., Morley, S. J., and Lewis, J. D. (2001) *Mol. Cell. Biol.* **21**, 3632–3641
- Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. C., and Rhoads, R. E. (2000) *J. Biol. Chem.* **275**, 41369–41376
- Imataka, H., and Sonenberg, N. (1997) *Mol. Cell. Biol.* **17**, 6940–6947
- Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. C., Merrick, W. C., and Rhoads, R. E. (2001) *J. Biol. Chem.* **276**, 2872–2879
- Goyer, C., Altmann, M., Lee, H. S., Blanc, A., Deshmukh, M., Woolford, J. L., Jr., Trachsel, H., and Sonenberg, N. (1993) *Mol. Cell. Biol.* **13**, 4860–4874
- Kolupaeva, V. G., Pestova, T. V., Hellen, C., and Shatsky, I. N. (1998) *J. Biol. Chem.* **273**, 18599–18604
- Pestova, T. V., Shatsky, I. N., and Hellen, C. U. (1996) *Mol. Cell. Biol.* **16**, 6870–6878
- Cuesta, R., Laroia, G., and Schneider, R. J. (2000) *Genes Dev.* **14**, 1460–1470
- Laroia, G., Cuesta, R., and Schneider, R. J. (1999) *Science* **284**, 499–502
- Waskiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. (1997) *EMBO J.* **16**, 1909–1920
- Fukunaga, R., and Hunter, T. (1997) *EMBO J.* **16**, 1921–1933
- Pyronnet, S., Imataka, H., Gingras, A.-C., Fukunaga, R., Hunter, T., and Sonenberg, N. (1999) *EMBO J.* **18**, 270–279
- Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C., and Thach, R. E. (1985) *J. Biol. Chem.* **260**, 7651–7658
- Gingras, A.-C., Raught, B., and Sonenberg, N. (1999) *Annu. Rev. Biochem.* **68**, 913–963
- Rhoads, R. E. (1999) *J. Biol. Chem.* **274**, 30337–30340
- Ehrenfeld, E. (1996) in *Translational Control* (Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., eds) pp. 549–573, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

34. Grifo, J. A., Tahara, S. M., Morgan, M. A., Shatkin, A. J., and Merrick, W. C. (1983) *J. Biol. Chem.* **258**, 5804–5810
35. Edery, I., Hümbelin, M., Darveau, A., Lee, K. A. W., Milburn, S., Hershey, J. W. B., Trachsel, H., and Sonenberg, N. (1983) *J. Biol. Chem.* **258**, 11398–11403
36. Buckley, B., and Ehrenfeld, E. (1987) *J. Biol. Chem.* **262**, 13599–13606
37. Etchison, D., Milburn, S. C., Edery, I., Sonenberg, N., and Hershey, J. W. B. (1982) *J. Biol. Chem.* **257**, 14806–14810
38. Lamphear, B. J., Yan, R., Yang, F., Waters, D., Liebig, H.-D., Klump, H., Kuechler, E., Skern, T., and Rhoads, R. E. (1993) *J. Biol. Chem.* **268**, 19200–19203
39. Pelletier, J., and Sonenberg, N. (1988) *Nature* **334**, 320–325
40. Jang, S. K., Kräusslich, H.-G., Nicklin, M. J. H., Duke, G. M., Palmenberg, A. C., and Wimmer, E. (1988) *J. Virol.* **62**, 2636–2643
41. Liebig, H.-D., Ziegler, E., Yan, R., Hartmuth, K., Klump, H., Kowalski, H., Blaas, D., Sommergruber, W., Frasel, L., Lamphear, B., Rhoads, R. E., Kuechler, E., and Skern, T. (1993) *Biochemistry* **32**, 7581–7588
42. Ohlmann, T., Rau, M., Morley, S. J., and Pain, V. M. (1995) *Nucleic Acids Res.* **23**, 334–340
43. Borman, A. M., Kirchweger, R., Ziegler, E., Rhoads, R. E., Skern, T., and Kean, K. M. (1997) *RNA (N. Y.)* **3**, 186–196
44. Imataka, H., Gradi, A., and Sonenberg, N. (1998) *EMBO J.* **17**, 7480–7489
45. Johannes, G., and Sarnow, P. (1998) *RNA (N. Y.)* **4**, 1500–1513
46. Bushell, M., Poncet, D., Marissen, W. E., Flotow, H., Lloyd, R. E., Clemens, M. J., and Morley, S. J. (2000) *Cell Death Differ.* **7**, 628–636
47. Harlow, E., and Lane, D. (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
48. Webb, N. R., Chari, R. V. J., DePillis, G., Kozarich, J. W., and Rhoads, R. E. (1984) *Biochemistry* **23**, 177–181
49. Gaillard, C., and Strauss, F. (1990) *Nucleic Acids Res.* **18**, 378
50. Sechi, S., and Chait, B. T. (1998) *Anal. Chem.* **70**, 5150–5158
51. Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H.-D., Sommergruber, W., Sobrino, F., Hohenadl, C., Blaas, D., Rhoads, R. E., and Skern, T. (1994) *J. Virol.* **68**, 5677–5684
52. Lamphear, B. J., and Rhoads, R. E. (1996) *Biochemistry* **35**, 15726–15733
53. Duncan, R., Milburn, S. C., and Hershey, J. W. B. (1987) *J. Biol. Chem.* **262**, 380–388
54. Zhang, W., Czernik, A. J., Yungwirth, T., Aebersold, R., and Chait, B. T. (1994) *Protein Sci.* **3**, 677–686
55. Milburn, S. C., Duncan, R. F., and Hershey, J. W. B. (1990) *Arch. Biochem. Biophys.* **276**, 6–11
56. Sommergruber, W., Ahorn, H., Zophel, A., Maurerfogy, I., Fessler, F., Schnorrenberg, G., Liebig, H. D., Blaas, D., Kuechler, E., and Skern, T. (1992) *J. Biol. Chem.* **267**, 22639–22644
57. Sommergruber, W., Ahorn, H., Klump, H., Seipekt, J., Zoephel, A., Fessler, F., Krystek, E., Blaas, D., Kuechler, E., Liebig, H.-D., and Skern, T. (1994) *Virology* **198**, 741–745
58. Ziegler, E., Borman, A. M., Kirchweger, R., Skern, T., and Kean, K. M. (1995) *J. Virol.* **69**, 3465–3474
59. Kerekatte, V., Keiper, B. D., Badorff, C., Cai, A., Knowlton, K. U., and Rhoads, R. E. (1999) *J. Virol.* **73**, 709–717
60. Joachims, M., Van Breugel, P. C., and Lloyd, R. E. (1999) *J. Virol.* **73**, 718–727
61. Urzainqui, A., and Carrasco, L. (1989) *J. Virol.* **63**, 4729–4735
62. Joshi, B., Yan, R., and Rhoads, R. E. (1994) *J. Biol. Chem.* **269**, 2048–2055
63. Kozak, M. (1986) *Cell* **44**, 283–292
64. Gan, W., and Rhoads, R. E. (1996) *J. Biol. Chem.* **271**, 623–626
65. Gan, W. N., LaCelle, M., and Rhoads, R. E. (1998) *J. Biol. Chem.* **273**, 5006–5012
66. Hayashi, S., Nishimura, K., Fukuchi-Shimogori, T., Kashiwagi, K., and Igarashi, K. (2000) *Biochem. Biophys. Res. Commun.* **277**, 117–123
67. De Gregorio, E., Preiss, T., and Hentze, M. W. (1998) *RNA (N. Y.)* **4**, 828–836
68. Pappin, D. J. C., Hojrup, P., and Bleasby, A. J. (1993) *Curr. Biol.* **3**, 327–332
69. Beavis, R. C., and Chait, B. T. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6873–6877