Am J Hum Genet 38:461–481, 1986

Computer Prediction of Peptide Maps: Assignment of Polypeptides to Human and Mouse Mitochondrial DNA Genes by Analysis of Two-Dimensional-Proteolytic Digest Gels

DOUGLAS C. WALLACE,[1] JUNHUI YANG,[1] JIANHONG YE,[1] MARIE T. LOTT,[1] NOELYNN A. OLIVER,[2] AND JUDITH MCCARTHY[3]

SUMMARY

We have prepared a computer program that predicts complete and partial peptide maps from amino acid sequences. The program fragments amino acid sequences at designated cleavage sites and calculates the molecular weight and relative labeling of each peptide. These data are graphed as log molecular weight of the original protein (X-axis) vs. log molecular weight of the component peptides (Y-axis). The program is interactive, permitting adjustment of a number of graphic parameters and alteration of the position of proteins in the first dimension to accommodate aberrations in protein mobility.

The program has been used to predict the V8 protease peptide maps of the 13 open reading frames (ORFs) identified in the human and the mouse mitochondrial DNA (mtDNA) sequences. The results were compared to the V8 protease peptide maps obtained for mouse and human mitochondrially synthesized proteins by two-dimensional proteolytic digest gels. A high correlation was observed between the predicted and observed peptide maps. These results suggest the assignment of several proteins to mtDNA genes.

INTRODUCTION

Rapid nucleotide sequencing procedures have resulted in the characterization of a wide variety of viral, plasmid, and organelle genomes. Such nucleotide

Received June 5, 1985.

This work was funded by grants GM33022 and NS21328 from the National Institutes of Health, NP-460 from the American Cancer Society, PCM 83-40190 and BNS 85-08497 from the National Science Foundation, and a clinical research grant from the Muscular Dystrophy Association (to D. C. W.).

[1] Departments of Biochemistry and Pediatrics (Division of Medical Genetics), Woodruff Memorial Building, Emory University School of Medicine, Atlanta, GA 30322.

[2] Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142.

[3] Stanford Medical School, Stanford, CA 94305.

sequences frequently reveal amino acid coding regions (open reading frames or ORFs) for which there is no identified gene product.

Peptide mapping provides one approach by which polypeptides can be assigned to ORFs. In cases where the protein products of a genome can be identified (i.e., virion proteins, differentially labeled organelle proteins, etc.), peptide maps for the proteins can be obtained using proteases of known specificity. Gene assignments can then be made by comparing the experimentally observed peptide maps to those predicted from the amino acid sequences of the ORFs [1].

SDS-polyacrylamide (SDS-PA) gel systems have been used to determine peptide maps of separated proteins [2]. Peptide maps can now be generated for complex mixtures of differentially labeled proteins using the two-dimensional-proteolytic digest (2D-PD) gel system [1, 3–6]. In this system, radioactive polypeptides are separated by molecular weight in a first-dimension cylindrical gel followed by cleavage and resolution of their component peptides on a second dimension slab gel overlaid with a protease.

We have used this 2D-PD gel system to determine the peptide maps of proteins synthesized in the mammalian mitochondrion [1, 3]. These proteins are translated on mitochondrial ribosomes, and they can be differentially labeled by growth of cells in [^{35}S]methionine and emetine, an inhibitor of cytosolic ribosomes [7]. We used the *Staphylococcus aureus* V8 protease for this analysis. This protease cleaves peptide bonds at glutamic acid residues [8, 9] and generates a limited number of discrete peptides from mitochondrial proteins [1, 3].

Previously, we compared the V8 protease peptide maps observed for the smaller human mitochondrially synthesized proteins with those that would be expected from glutamate cleavage of the smaller human mtDNA ORFs. This permitted us to assign a 14–15 kilodalton (Kd) polymorphic polypeptide to mtDNA unidentified reading frame 3 (URF3) and to tentatively assign a 15–16 Kd polymorphic protein to URF6 [1].

The extension of this approach to other mitochondrial proteins has been limited by difficulties both in calculating the expected peptide maps for larger ORFs and in comparing predicted maps with experimentally determined peptide maps. To overcome these limitations, we report here a computer program that permits the rapid and accurate prediction of peptide maps from known amino acid sequences and their two-dimensional graphic representation using the same parameters as the 2D-PD gel. Using this program, we have predicted the V8 peptide maps of the 13 human and mouse mtDNA ORFs and compared these to our 2D-PD gel results. The correlations obtained suggest a number of additional mtDNA gene assignments.

MATERIALS AND METHODS

Computer Model

The computer program "2DPDMODL" was prepared on an IBM-PC with 64 kilobytes random access memory using DOS 2.0. Other necessary peripherals included two disk drives, an IBM-PC color/graphics monitor, and a dot matrix printer (EPSON FX-80).

The program was written in IBM-PC advanced BASIC (Version A 1.10, Copyright, IBM Corp., 1981, 1982). The program is "menu" driven and has a "conversation" mode to facilitate communication with the user during the modeling process. The program can be requested from D. C. W. An instruction manual will be provided with the program diskette.

Cell Cultures and Mitochondrial Labeling

Maintenance of attached cell lines has been described [7, 10]. Human lymphoblast lines were established from venous leukocytes. Thirty milliliters of blood was drawn in ACD (acid citrate dextrose) anticoagulant and the leukocytes purified by Ficoll-Hypaque gradient (Pharmacia, Piscataway, N.J.). Leukocytes were washed in RPMI 1640 without serum and 5×10^6 cells pelleted and resuspended in 1 ml of spent, filtered, medium from a B95-8 EBV-transformed marmoset cell culture (CRL 1612, American Type Culture Collection, Rockville, Md.). After 1 hr at 37°C, 4 ml of RPMI 1640 + 20% fetal calf serum + 2.0 μg/ml cyclosporine A (Sandoz, East Hanover, N.J.) was added [11]. Cells were maintained in this medium until the lymphoblasts started to proliferate.

Mitochondria for 2D-PD gel analysis were isolated by differential centrifugation from cells grown in [^{35}S]methionine and emetine. Approximately 2×10^5 acid precipitable counts per minute were loaded per gel [1, 3]. The mitochondrial proteins of lymphoblast cultures were labeled using a modification of the protocol of Spinner and King [12]. Five million Epstein-Barr virus-transformed lymphoblasts were suspended in 1 ml of methionine-free RPMI 1640 + 5% dialyzed fetal calf serum + 100 μg/ml emetine (Sigma, St. Louis, Mo.). After 20-min preincubation, 400 μCi of L-[^{35}S]methionine (New England Nuclear, Boston, Mass.) was added and the culture incubated at 37°C for 2 hrs. For controls, 100 μg/ml chloramphenicol (a mitochondrial ribosome inhibitor) was added to suppress all mitochondrial protein labeling. After labeling, unlabeled methionine was added to 10 mM and the cells incubated 20 min. Cells were collected, washed thoroughly with buffered saline, and the dry pellets frozen at −80°C. Cell pellets were resuspended in lysis buffer containing 5% beta-mercaptoethanol and 4% SDS [3] and briefly sonicated on ice. Approximately 5×10^4 acid-precipitable counts per sample were loaded per gel channel.

Gel Electrophoresis and Analysis

The 2D-PD gels included a 13% SDS-PA gel for the first dimension and a 16% resolving gel with a 4.75% stacking gel in the second dimension. The second-dimension gel was overlaid with a solution of 1% low-melting-temperature agarose solution containing 1–2 mg of V8 protease (Miles Laboratories, Elkart, Ind.) [1, 3].

Human and mouse mitochondrially synthesized proteins were compared by using a linear 12.5% 1D-SDS-PA gel. A 12%–20% gradient SDS-PA gel was used to resolve the mitochondrial proteins of lymphoblasts. All gels were run in the Tris-glycine buffer of Laemmli [13] and fluorographed using En3Hance (NEN, Boston, Mass.) [7].

The apparent molecular weights of mitochondrially synthesized proteins in 1D-SDS-PA gels were estimated by parallel electrophoresis of known protein size standards. The molecular weights of peptides in 2D-PD gels were estimated by inclusion of known protein standards of mitochondrially synthesized proteins in slots lacking protease along the sides of the gel. Apparent molecular weights determined from 1D-SDS-PA gels were used for the mitochondrially synthesized proteins. Different 2D-PD gels were aligned using the diagonal created by the undigested protein spots [9].

To confirm that V8 protease cleaves at glutamates under the 2D-PD gel conditions used, the peptide map of lysozyme was predicted and then determined experimentally. Three peptide spots were observed that correspond to the expected peptides with molecular weights of about 12.2 Kd, 9.4 Kd, and 3.5 + 2.8 Kd. The smallest expected 0.7-Kd peptide was not observed. It probably migrated off the gel [9].

The relative labeling of the peptide spots in 2D-PD gels and protein bands in 1D-SDS-PA gels was determined by densitometry. Autoradiographs were placed on a light box and viewed with a video camera linked to a digitizing board in a microcomputer. The area encompassed by the spot or band was scanned, the darkening under each pixel summed, and the background subtracted [14].

Computer Modeling

The computer-generated V8 peptide maps of the 13 human and mouse mtDNA ORFs were optimally aligned with the 2D-PD maps by preparing a slide of the 2D-PD gel and showing it on the computer monitor. The positions of the mtDNA ORF proteins were then adjusted to optimize the alignment of the predicted and observed peptide maps of each protein. Open circles were included for the intact proteins and for a few partial digestion products. The circles for intact protein spots serve to identify undigested proteins that migrate along the diagonal.

RESULTS

We have prepared an interactive computer program for predicting complete and partial peptide maps. The program can print out the amino acid sequence and indicate the position of protease cleavage sites and labeled amino acids. Further, it will calculate the peptide maps of a designated polypeptide sequence using any stated protease specificity. The protease specificity is designated by the protease's name, the recognized amino acid, and the position of cleavage (either amino or carboxyl side). For labeled proteins, the program also calculates the peptide's specific activity based on the number of labeled amino acids in the peptide.

The program identifies each amino acid in the sequence that would be cleaved by the designated protease (e.g., V8 protease that cleaves on the carboxyl side of glutamates) and stores these amino acid numbers. Next it identifies and addresses each labeled amino acid (e.g., methionine). Each adjacent pair of cut amino acids (glutamates) defines the ends of the peptide. The molecular weights of the amino acids within the peptide, minus the water of polymerization, are added to yield the molecular weight of the peptide. The number of labeled amino acids (methionines) between the cut sites is then totaled to obtain the specific activity of the peptide. These data are stored in a digest file for that protein and protease that includes the first and last amino acid in each peptide, its molecular weight, and the number of labeled amino acids included. The peptides are rank ordered by molecular weight.

The same logic is used to prepare a separate catalog for all partial digest peptides of the protein. In this case, all possible combinations of cut points are used. These files also tabulate and retain the number of cut points skipped in each partial peptide.

A two-dimensional graphic display can be made using these data. In the first dimension (X-axis), the log molecular weight of the intact protein is graphed. In the second (Y-axis), the log molecular weight of the constituent peptides in the complete digests are plotted. The relative spacing of the proteins and peptides can be changed by adjustable constants. The area of each peptide spot is set proportional to the number of labeled amino acids. The resulting graph is displayed on the screen and can be printed on a dot matrix printer.

The user has a range of options for manipulating the graphic display of the results. These are particularly useful if the user wishes to model the computer relative to an observed 2D-PD autoradiograph. All options for changing the graphic parameters are menu driven. The program has a subroutine that permits the user to modify the position of proteins in the first dimension either left or right. This permits compensation for aberrations in protein migration in the first dimension due to unknown factors such as extreme hydrophobicity. The position of the peptides is moved along with the protein, but the relative position of the peptides is kept fixed. The user also can include open circles in the second dimension to show the position of undigested protein or partial digest spots.

Prediction of Peptide Maps for Human and Mouse Mitochondrial Proteins

We used our computer program to predict the V8 peptide maps of the 13 ORFs in human [15] and mouse [16] mtDNAs (figs. 1 and 2). Consistent with standard labeling protocols, methionine was designated as the labeled amino acid. Using the interactive component of the program, the relative positions of the proteins in the first dimension were adjusted to optimize the correspondence between the peptide spots predicted by the program and those observed in 2D-PD maps of human and mouse mitochondrially synthesized proteins. Each group of peptide spots is designated by a number that corresponds to the name of the ORF in the mtDNA.

Open circles have been included in the computer map to show the expected positions of the undigested proteins (upper diagonal group of spots) and of those few partial digests that were necessary to maximize the correspondence between the computer maps and the observed 2D-PD gels.

The dominant peptide spots are summarized in tables 1 and 2. Each mtDNA gene is designated by the number of the ORF given in figures 1 and 2 but listed in the order that most closely approximates the peptide maps observed in the 2D-PD gels. The relative labeling (specific activity) of the peptides from each protein was calculated from the number of methionines in each peptide normalized to the peptide having the fewest methionines. When a number of peptides would be expected to overlap on a 16% resolving gel, the molecular weights of these peptides were averaged, but the number of methionines was totaled.

Two Dimensional Proteolytic Digest Maps of Mouse and Human Mitochondrially Synthesized Proteins

The 2D-PD V8 protease peptide maps of the mitochondrially synthesized proteins of human HeLa [17] and mouse RAG [18] cells are shown in figures 1 and 2. These proteins were labeled with [^{35}S]methionine in the presence of emetine.

On 13% 1D-SDS-PA gels, comparable to those used in the first dimension of our 2D-PD system, we have observed that human and mouse mitochondrially synthesized proteins fall into four groups (fig. 3). In order of decreasing molecular weight, these include a top group of three bands, a middle group of three bands, a lower group of three bands, and four to five additional lower

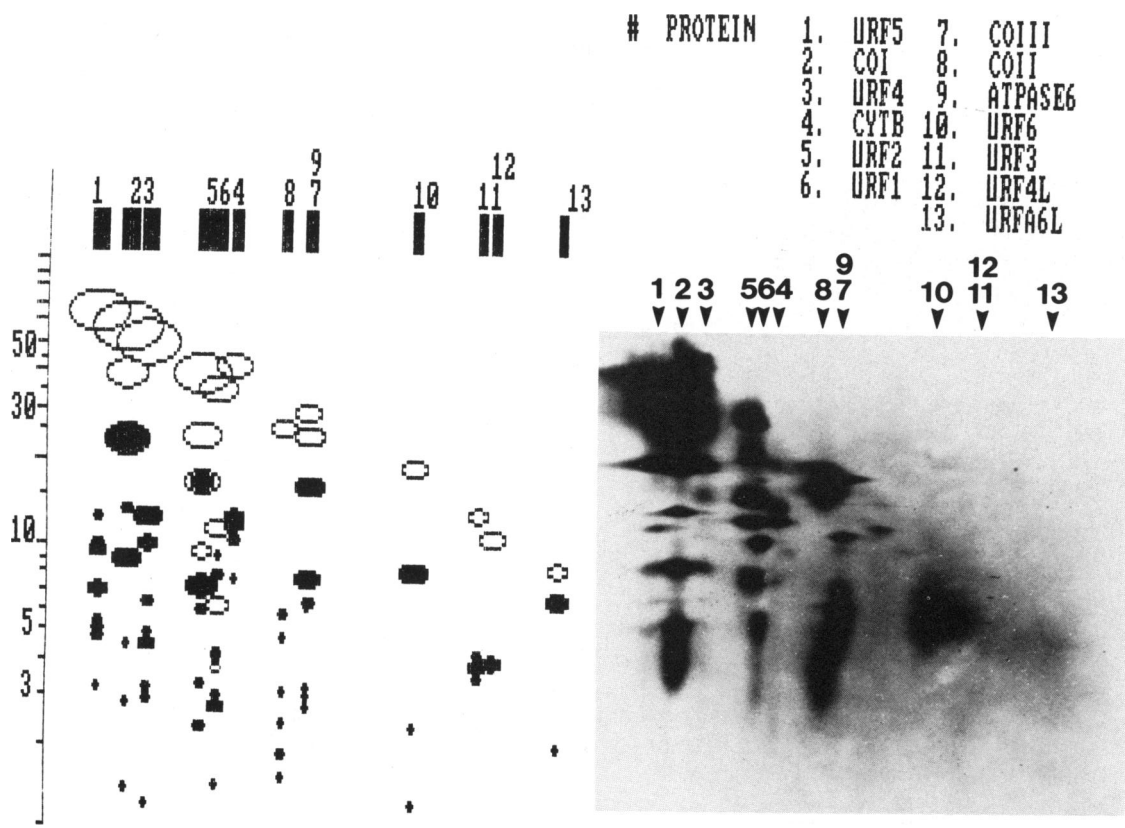

FIG. 1—Computer-generated peptide map and 2D-PD gel of human mtDNA proteins. The ORFs corresponding to the computer-generated peptide maps are indicated by nos. Mitochondrially synthesized proteins thought to have comparable peptide maps are designated by the same no. The ORFs are numbered in descending order of molecular weight.

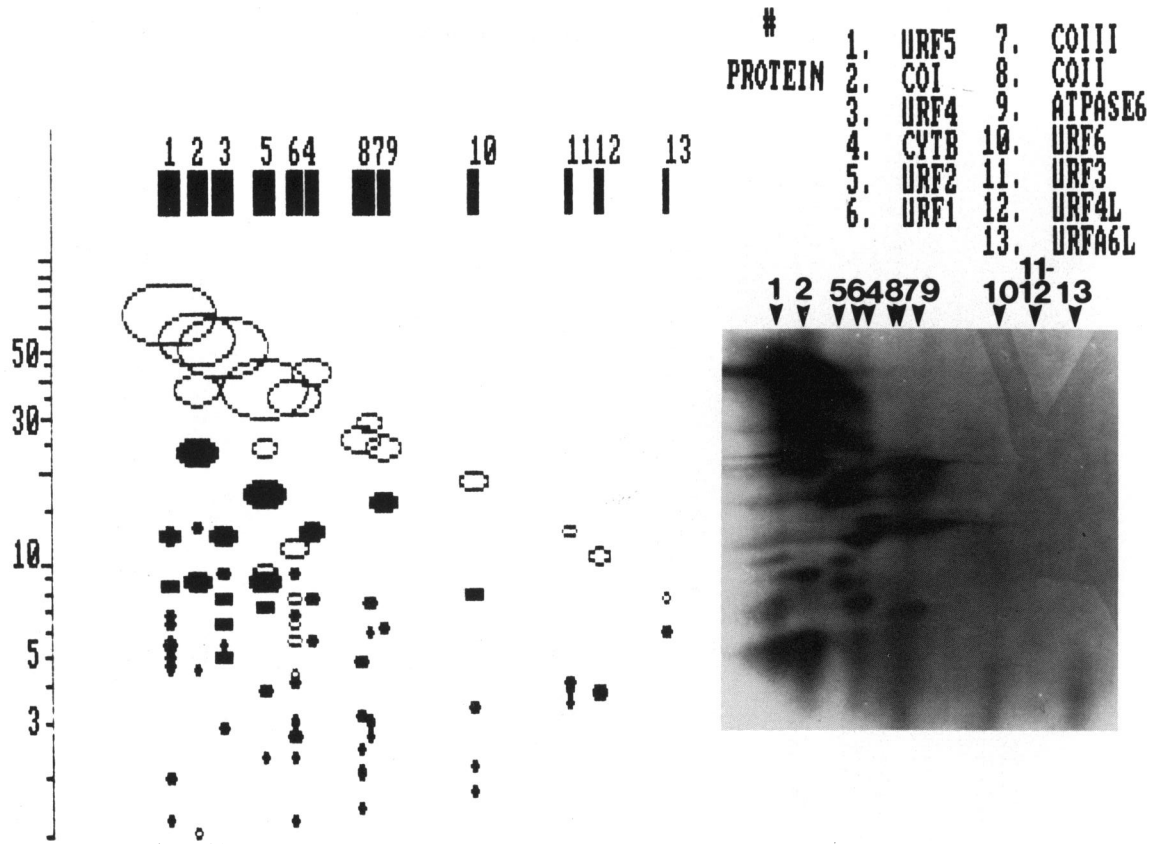

FIG. 2.—Computer-generated peptide map and a 2D-PD gel of mouse mtDNA proteins. The ORFs and proteins are numbered as in figure 1. The relative order of COII and COIII is uncertain.

TABLE I

HUMAN MITOCHONDRIAL PEPTIDES: OBSERVED VS. EXPECTED

	GENE			PROTEIN		
DESIGNATION†	Open reading frame‡ (MW-Kd)	Peptide§ Kd	Relative‖ no. met.	Apparent# (MW-Kd)	Peptide†† Kd	Relative intensity‡‡
1	URF5 (66.8)	12.6	2.0	(43.5)	13.0	2.1
		10.0*	7.0		9.2*	3.0
		6.8	5.0		7.0	1.9
		5.0*	11.0		5.5	1.7
		3.2	1.0		3.9*	1.0
2	COI (56.9)	39.0 p''	6.0	(39.5)	32.0	4.9
		23.3	4.7		22.0	4.5
		13.2	1.0		15.0	1.0
		8.7	3.0		9.0*	1.9
		4.5	1.0		5.0	
3	URF4 (51.5)	12.2	2.7	(36.5)	17.0	2.5
		9.9	1.7		9.5	1.6
		6.3	1.0		7.0	1.0
		4.5*	2.0		5.7	1.4
5	URF2 (39.0)	23.7 p'	< 2.0	(31.5)	23.0	2.0
		16.6	1.3		16.5	3.9
		9.5 p'	1.6		13.5	2.2
		7.1	1.3		8.1*	1.9
		3.8*	.0		5.5*	1.0
6	URF1 (35.6)	11.8*p'	≤10.3	(29.5)	11.5	4.9
		9.0				
		7.8	1.0		8.0	1.0
		6.1*p'	≤ 3.0		5.6*	3.6
		3.3*			4.0*	

4	cytb (42.6)	22.7 p′ 12.5 11.5 10.2 7.6	 1.7 1.7 1.0	(27.5)	21.5 15.0 13.0 10.5 8.0	 1.3 1.5 1.0
8	COII (25.5)	3.2*		(23.6)	3.8*	
7	COIII (30.0)	7.5 3.7*	1.7 1.0	(22.5)	6.2 5.4*	1.6 1.0
9	ATPase 6 (24.8)	15.7 6.2	2.7 1.0	(21.6)	13.0 6.6*	1.5 1.0
9a	?				12.5	
10	URF6 (18.7)	8.0		(16.7)	6.4	
11	URF4L (10.7)	3.9*		(14.8)	3.5*	
	URF3 (13.4)	3.8*		(14.0)**	3.5*	
12	URFA6L (8.3)	6.3 2.3	6.0 1.0	(9.8)	5.5* 1.5*	3.8 1.0

NOTE: Correlation coefficients for molecular weights of observed vs. expected peptides are 1/URF5 = .990, 2/COI = .991, 3/URF4 = .927, 5/URF2 = .980, 6/URF1 = .991, 4/cytb = .974, 7–12/COII + COIII + ATPase 6 + URF6 + URF4L + URF3 + URFA6L = .964.

† The ORF and protein designations correspond to the nos. in figure 1, listed in order of decreasing apparent molecular weight in the 1D-SDS-PAGE gel.

‡ ORFs from mtDNA sequence. Predicted molecular weights are in parentheses.

§ Values followed by (*) are the arithmetic means of a no. similar or continuously dispursed peptides. p′ indicates a primary partial while p″ indicates a secondary partial.

Relative no. methionines in each peptide normalized to the least-labeled peptide.

The apparent molecular weights of mitochondrially synthesized proteins are given in parentheses. **Indicates the HeLa variant.

†† Values followed by (*) are the arithmetic means of top and bottom value of a continuous streak.

‡‡ The spot intensities on the autoradiograph normalized to the spot in each protein with the lowest value.

TABLE 2

MOUSE MITOCHONDRIAL PEPTIDES: OBSERVED VS. EXPECTED

	GENE			PROTEIN		
DESIGNATION†	Open reading frame‡ (MW-Kd)	Peptide§ Kd	Relative‖ no. met.	Apparent# (MW-Kd)	Peptide†† Kd	Relative intensity‡‡
1	URF5 (68.3)	12.4	1.4	(45.5)	10.0	1.0
		8.3	1.0		7.0	1.1
		5.1*	5.0		5.3*	4.1
2	COI (56.8)	39.0 p''	9.0	(39.0)	33.0*	2.9
		23.2	7.5		21.0	2.7
		13.0	1.5		12.5	1.0
		8.7	5.0		9.0	1.6
		4.5	1.0		5.1*	2.5
3	URF4 (51.8)	12.3		(37.7)	12.5	
		9.2			9.0	
		5.4*			4.7*	
6	URF2 (38.7)	24.8*p'	11.5	(33.7)	25.0	6.4
		16.9	3.7		16.5	18.7
		9.4	1.5		10.0	1.0
		8.7	2.8		8.4	2.3
		7.1	1.3		7.2	1.8
		3.8	1.0		4.9*	4.4
5	URF1 (35.6)	partials		(30.7)	19.5	1.3
		11.5*p'	1.4		12.0	1.0
		9.2				
		7.2*p'	1.0		7.2	1.0
		4.1*	3.3		5.0	1.7

4	cytb (43.0)	18.3 p' 12.7 7.7 5.6	3.7 2.7 1.3 1.0	(27.5)	16.0 13.5 8.3 5.0	2.5 1.1 1.0
8	COII (25.9)	4.7 1.7*		(24.5)	4.8*	
7	COIII (29.9)	7.4 5.9 2.9		(23.3)	6.5	
9	ATPase 6 (25.1)	15.9 6.2	3.3 1.0	(22.3)	13.0 5.5*	3.0 1.0
10	URF6 (18.6)	7.9 3.4	1.7 1.0	(16.7)	10.1* 3.8*	1.4 1.0
11	URF4L (10.5)	3.8		(14.7)	3.5*	
12	URF3 (13.1)	3.8*		(14.7)		
13	URFA6L (7.8)	5.9		(9.5)	4.5*	

NOTE: Correlation coefficients for molecular weights of observed and expected peptides are 1/URF5 = .996, 2/COI = .999, 3/URF4 = not determined, 6/URF2 = .998, 5/URF1 = .993, 4/cytb = .972, 7–13/COII + COIII + ATPase 6 + URF6 + URF4L + URF3 + URFA6L = .932. †, ‡, §, ‖, #, ‡‡ as per table 1.

†† As per table 1. The mouse values for protein 3 are tentative due to the low specific activity of these spots in the mouse 2D-PD gel.

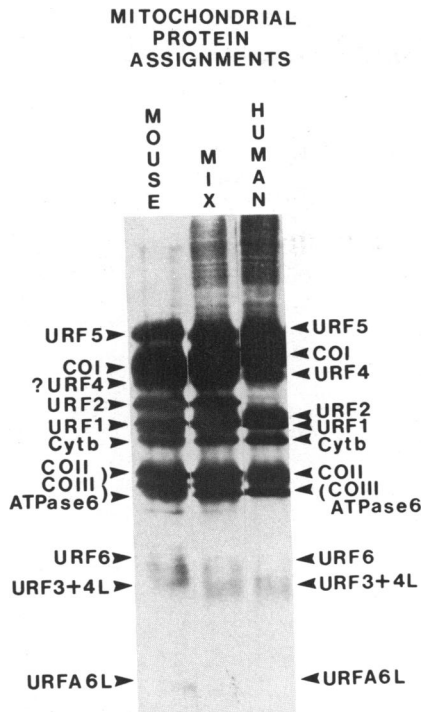

FIG. 3.—Proposed gene assignments of mouse and human mitochondrial translation products and a 1:1 mixture of the two. The protein assignment to mouse URF4 is tentative due to weak labeling of peptides in the 2D-PD gel (fig. 2). The exact correspondence between 1D-SDS-PAGE band and mouse proteins COI, COII, and ATPase is unclear, although ATPase 6 is thought to be at the bottom of this group. Similarly, we believe that the human ATPase 6 and COIII proteins run as a doublet in this gel system. The lower molecular weight proteins assigned to URF6, URF3, and URF4L are not well resolved in this gel.

molecular weight protein bands. These proteins are best resolved in gradient 1D-SDS-PA gels, as shown for human proteins in figure 4. In humans, these groups fall into the apparent molecular weight ranges of about 35 to 45 Kd, 25 to 35 Kd, 20 to 24 Kd, and 9 to 18 Kd, respectively [7]. On 13% SDS-PA gels we see a reasonably good correspondence between the mobilities for mouse and human mitochondrial proteins (fig. 3). This suggests that the peptide maps of the mouse and human mitochondrial proteins may be similarly arranged in the 2D-PD fluorographs (figs. 1 and 2).

The apparent molecular weights of these mitochondrially synthesized proteins (estimated from 12.5% gels) and their component peptides (estimated from the 2D-PD gels in figs. 1 and 2) are listed in tables 1 and 2. The proteins are listed in order of decreasing apparent molecular weight but numbered according to their designation in figures 1 and 2 for easy comparison with the computer-generated peptide maps. The relative specific activities of the peptides for each protein were determined by densitometry of the entire spot. These

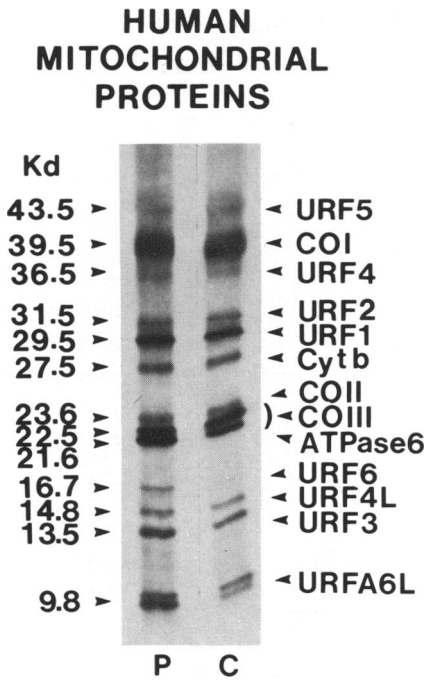

FIG. 4.—Human lymphoblast mitochondrial translation products resolved on a 12%–20% SDS-PAGE gradient gel. *C* = control lymphoblasts, and *P* = patient with mitochondrial myopathy and myoclonus [33]. The ambiguity in relative mobility of ATPase 6 and the lower CO subunit is indicated by a parenthesis. Apparent molecular weights relative to known protein standards are given [7].

values were normalized to the least-labeled peptide of each protein and are listed in tables 1 and 2.

DISCUSSION

We present here a program for predicting and interpreting complete and partial peptide maps. This program also permits a graphic representation of the maps and the adjustment of a number of parameters to permit comparison of the computer output with experimental results. The program will be of use to researchers who wish to analyze the peptide maps of a group of related proteins whose gene sequences are known. We have used this program to predict the V8 protease peptide maps of the 13 ORFs of the human and the mouse mtDNAs and have compared these results with the observed peptide maps from 2D-PD gels.

The organization and genes of the mammalian mtDNAs are highly conserved. Nucleotide sequence analysis has revealed that each mtDNA contains a large and small rRNA gene, a set of tRNAs, and 13 ORFs [15, 16, 19, 20]. Of the 13 ORFs, six have been assigned functions through homology with known amino acid sequences. These are three cytochrome c oxidase genes (COI, COII, and COIII), cytochrome b (cytb), and two ATPase genes (ATPase 6 and

tentatively ATPase 8, formerly URFA6L [21]). The remaining URFs 1, 2, 3, 4L, 4, 5, and 6 are not homologous to any known protein.

The proteins encoded by several of the human mtDNA ORFs have been identified by preparing antibodies either to purified enzyme complexes or to chemically synthesized oligopeptides identified in the nucleotide sequences of the genes. These antibodies have, in turn, been used to precipitate differentially labeled mitochondrial translation products. Protein assignments have been made for COI, COII, COIII, ATPase 6, URF1, URF3, and URFA6L [22–25] and have been suggested for cytb, URF2, URF4, and URF4L [26].

Here, we have compared the V8 peptide maps predicted for human and mouse mtDNA genes with those we have observed on our 2D-PD gels (figs. 1 and 2). For this analysis, we have made three assumptions. First, we assumed that V8 protease digests proteins to near completion at glutamate residues. Second, we assumed that the linear displacement of the peptides throughout the central region of the gel was proportional to the log of the molecular weight of the peptides. Third, we assumed that the migration of the V8 peptides provided a reasonable approximation of their molecular weight. These assumptions were supported by our analysis of lysozyme.

The first computer maps we generated were based on two hypotheses. The first hypothesis was that the 13–14 mitochondrially synthesized polypeptides observed on our 1D-SDS-PA gels [7] correspond to the 13 ORFs identified in the mtDNA sequence. The second was that the mtDNA gene products migrate in approximately the same order as their predicted molecular weights. Comparison of the two-dimensional peptide maps predicted by these hypotheses with the 2D-PD maps obtained revealed that minor adjustments in the order of the mtDNA gene products were required to maximize the correlation between predicted and observed peptides. Once these adjustments were made, however, the two peptide maps corresponded quite well. The side-by-side comparison of the peptide maps is shown in figures 1 and 2 and tabulated in tables 1 and 2. The similarities in the molecular weights of the observed and expected peptide maps are expressed as correlation coefficients. These range from .927 to .999 for the various proteins (tables 1 and 2). Since the peptide maps predicted from these straightforward hypotheses closely approximated the gel data, we concluded that testing more complex hypotheses (i.e., extensive protein processing) was unjustified.

Based on our optimized computer model, the largest three human mitochondrial translation products (from slowest to fastest mobility) are provisionally assigned to mtDNA ORFs URF5 (66.6 Kd), COI (57.0 Kd), and URF4 (51.4 Kd), respectively. We think that the diffuse radioactivity at the top of this group results from undigested protein and a secondary partial (two sites skipped) for COI yielding a 38 Kd peptide. For all three proteins, the correlation coefficient between observed and predicted peptides is greater than .927 (table 1). The relative labeling indices are also comparable.

The assignment of mouse proteins 1 and 2 to URF5 and COI is the same as human. However, the 2D-PD for mouse RAG cells does not have strong spots for protein 3. A protein in this region is seen in the 1D-SDS-PA gel (fig. 3), so we have provisionally assigned this protein to URF4 in mouse as well.

Based on expected molecular weight alone, the middle group of three proteins should be the translation products of cytb (42.7 Kd), URF2 (38.9 Kd), and URF1 (35.6 Kd). However, in this order, the computer-generated peptide maps did not correlate with those seen in the 2D-PD gels. A major improvement was obtained when the relative position of the mtDNA ORF for cytb (product 4) was moved from the upper to the lower side of this group. This change was supported by the computer prediction that the human cytb protein should yield two prominent V8 peptides close together at 11.5 and 12.5 Kd (five methionines each) while the mouse cytb protein should yield one prominent fragment at 12.8 Kd (eight methionines). The next mouse peptide should be at 7.7 Kd. In our 2D-PD gels, only the lowest protein in the second group of three (protein 4) meets this criteria. Hence, from our data, protein 4 is the best candidate for the cytb gene product.

The peptides of the remaining two proteins of the second group of three proteins [5, 6] are distinguished by a human-mouse interspecific polymorphism in protein 5 [3]. In both species, the peptide maps of protein 5 most closely resembled those predicted for URF2. However, to fully explain the 2D-PD peptide maps, certain primary partials (one site skipped) had to be included. In both species, the largest spot is on the diagonal and represents undigested protein. The next peptide at 23–25 Kd is best explained in both species as a primary partial. Additional primary partials predict the human 10–13 Kd peptide and the "hat-like" appearance of the 16.5 Kd peptide. Although inclusion of these partial digest products weakens this assignment, the overall correlation between the observed protein 5 peptides and the expected human and mouse URF2 peptides was excellent (.98 and .998, respectively).

The peptide maps of human and mouse mitochondrial protein 6 do not correspond with those predicted for the complete V8 digestion of cytb, URF1, or URF2. However, if it is assumed that several of the prominent peptides result from primary partials, then this protein can tentatively be assigned to URF1. Both human and mouse 2D-PDs have prominent peptides at 11.5–12 Kd, which are not predicted from the computer digest. The complete digest would predict a peptide at about 9 Kd, which is absent in both species. However, these two anomalies might be explained by assuming primary partials at the V8 sites flanking the 9 Kd peptide. Both of these partials would create 11.5–12 Kd peptides with two to six methionines. Hence, the loss of the 9 Kd peptide would predict the coordinate appearance of an 11–12 Kd peptide, which was observed. A similar 5–6 Kd peptide spot is also explained by a discrete set of partials. Although the necessity of hypothesizing extensive incomplete digestion of this protein severely weakens the assignment, still, the resulting peptides observed correlate well with the proposed partial digestion products (.99, tables 1 and 2).

According to predicted molecular weight, the third group of three proteins should include COIII (30.0 Kd), COII (25.5 Kd), and ATPase 6 (24.8 Kd). Because of the poor resolution of these proteins in 13% 1D-SDS-PA gels (fig. 3), comparison of the computer-generated peptide maps with the 2D-PD maps of these proteins [7–9] proved difficult. In both mouse and human mtDNAs, the ATPase 6 protein would be predicted to have two V8 peptides, one prominent

15.7–15.9 Kd peptide (eight to 10 methionines) and a second less prominent 6.2 Kd peptide (three methionines). In the mouse 2D-PD, these two peptides are dramatically displayed (see protein 9 of fig. 2). The larger peptide lies below a spot of undigested protein. This peptide migrated at 13 Kd, slightly faster than the predicted 16 Kd. The smaller peptide migrated at 5.5 Kd. These results imply that the mouse mitochondrial ATPase 6 probably migrates in this region and at the bottom of this group of three proteins. Identification of the human ATPase 6 subunit is less clear. However, using the mouse data as a guide, we believe that the 13 Kd spot at the lower end of this group is the larger ATPase 6 peptide. The smaller peptide is thought to fall in the diffuse spot below it. This interpretation suggests that the human ATPase 6 protein is also found in the lower region of this group of three proteins, overlapping with one of the cytochrome c oxidase subunits.

Computer analysis of the V8 peptide maps of human COII and COIII indicates that COII would be digested by V8 protease into a number of small peptides of 5.5 Kb and less of similar specific activity. COIII, by contrast, should yield a high-specific activity 7.5-Kd peptide and a number of smaller low-specific activity peptides below it. The best candidate for COII is protein 8, having a single continuous lens-shaped spot in the region of the gel expected for the COII peptides. This would imply that the COIII peptides are mixed with those of the ATPase 6 protein.

Computer analysis of the mouse COII and COIII genes predicts that the COIII gene should have a 7.5-Kd peptide (four methionines) and COII should have a smaller (4.7-Kd) peptide. Both should have additional smaller peptides. In our 2D-PD gels, the mouse peptides of both COII and COIII have relatively low-specific activity and are not well resolved. We feel they represent the streak slightly to the left of protein 9 (ATPase 6) in figure 2. In our computer model, COII was placed slightly above COIII. However, the opposite arrangement would be equally likely. Thus, our best interpretation is that mouse COII and COIII migrate close together and above ATPase 6 on 13% SDS-PA gels.

Based on predicted molecular weight, the protein product of URF6 (18.6 Kd) should migrate faster than COII, COIII, and ATPase 6 on SDS-PA gels. Computer analysis of the URF6 peptides predict that the human protein should yield an approximately 8 Kd spot while the mouse should yield 7.9 and 3.4 Kd spots. In the 16–20 Kd region of the 2D-PD gel, protein 10 from human cells has one dominant peptide at 5–7.8 Kd, while protein 10 from mouse cells has two weaker spots at 10.1 and 3.8 Kd. This interspecific peptide polymorphism provides strong support for our previous proposal [1] that protein 10 is the gene product of URF6.

The next largest gene products are URF3 (13.1 Kd) and URF4L (10.7 Kd). Computer predictions of the human URF3 and URF4L peptide maps indicate that V8 protease should cleave them into peptides in the 3–4 Kd range. In 13% 1D-SDS-PAGE, we believe that these proteins may overlap at the location designated proteins 11 and 12. The peptides of these proteins are better visualized with prolonged exposure [1]. We previously showed that the HeLa cell mtDNAs have a mutation in the URF3 gene that changes the apparent mobility

of its protein product from 15 to 14 Kd. In HeLa, the displacement of this band reveals a fainter band remaining at 15 Kd [1, 7]. These two proteins are resolved in higher-density 1D-SDS-PA gels (fig. 4). In HeLa cells, the lower band is displaced even farther and thus must be URF3 (data not shown). Hence, the upper band (fig. 4) could be the product of URF4L. However, since URF4L is expected to give only one V8 peptide and we have no independent means of confirming this assignment, this proposal must be considered tentative.

The smallest mtDNA gene is URFA6L (9.8 or 9.5 Kd). Computer analysis of the URFA6L gene predicts that the human protein should have two V8 peptides (6.3 Kd and 2.3 Kd), while the mouse should have only one (5.9 Kd). This interspecific peptide polymorphism is seen in the peptide maps of the lowest protein in mouse and human 2D-PDs (figs. 1 and 2). Hence, we feel this is the URFA6L gene product.

This analysis does not account for two additional gel observations. In our HeLa 2D-PD gel, we resolved a single peptide between proteins 9 and 10 at approximately 18 Kd. We occasionally see faint polypeptides in this region but do not know their origin. Also, high-density 1D-SDS-PA gels resolve the lowest polypeptide band into two bands. These could be modifications of the URFA6L polypeptide or indicate the presence of an additional unassigned polypeptide.

The reliability of our gene mapping system can, in part, be assessed by comparing our results with those of others. A number of reports have appeared proposing the assignment of mitochondrial translation products to human mtDNA genes [22–26]. No comparable data are available for the mouse mitochondrial proteins.

Comparison of our human assignments with those of others reveals both similarities and differences. It is unclear whether these differences represent errors in one or the other assignments or simply result from differences in the SDS-PA gel systems used to separate the mitochondrial proteins. For example, extensive studies have been conducted on assigning human mitochondrial translation products to mtDNA genes using antibodies prepared against enzyme complexes or oligopeptides. These studies have employed phosphate-buffered SDS-PA-urea gels or Tris-buffered gels to resolve the mitochondrial proteins [22–26]. By contrast, in all of our studies, we have used Tris-glycine-buffered gels [13].

We are the first to provide evidence that the largest mitochondrially synthesized protein is the product of the URF5 gene. However, others have proposed protein assignments for COI and URF3. These assignments are fully consistent with our results [24–26].

Both our analysis and the reports of others [22, 26] assign the proteins in the second group of three bands to cytb, URF1, and URF2. However, the order proposed for these gene products differs between the two studies. We have assigned these proteins, from slowest to fastest mobility, to the genes URF2, URF1, and cytb, respectively. By contrast, other researchers have assigned these proteins in the order cytb [26], URF2 [26], and URF1 [22]. Although our assignments of URF1 and URF2 are weakened by the necessity of including partial peptides, our assignment of the lowest protein to cytb is supported by a

clear-cut interspecies peptide polymorphism. The discrepancy in the relative position of cytb between the two studies might be explained by the observation that cytb migration can vary substantially in different SDS-PA gel systems [27].

For the third group of three bands, our placement of the human ATPase 6 subunit in the lower doublet of this group of three proteins (fig. 3; resolved in gradient gels, fig. 4) is consistent with previous reports [22]. However, our placement of COII above COIII differs from that reported by others for Tris-buffered SDS-PA gels, although not for phosphate-buffered SDS-PA-urea gels [22–24]. If, in contrast to our proposal, COIII migrates slower than COII in Tris-glycine-buffered gels, then the protein at position 8 should have a prominent 7.5-Kd peptide. Such a peptide is seen at position 7/9 but not at 8 (fig. 1). To explain this discrepancy, we would have to hypothesize that the 7.5-Kd peptide either remains near the origin as part of the high molecular weight undigested material, is modified by additional proteolytic cleavage, or migrates significantly faster than its observed molecular weight. Further, if protein 7 is COII, then it must be subject to extensive partial digestion to yield the observed peptides. Unfortunately, the mouse 2D-PD fluorograph provides little additional data to clarify our COII and COIII assignments.

For the lowest molecular weight bands, we are the first to assign the largest protein to URF6. Our previous placement of the URF3 protein below that of URF6 [1] has been supported by a similar placement of this gene product by others [22]. Our protein assignment to URFA6L is also consistent with the work of others [23]. By contrast, our positioning of the URF4L protein differs significantly from other reports. Other researchers have proposed that this protein migrates ahead of the URFA6L protein in phosphate-buffered-SDS-PA-urea gels [26] while we have tentatively placed it above the URF3 protein. Since our proposed assignment for URF4L is based on limited data, and we have detected a polypeptide below URFA6L, we feel that additional data will be required to support or deny our assignment.

At present, we have no additional means of testing our assignments. If additional enzymes become available for these gel systems, their expected and predicted digestion patterns could also be examined. Alternatively, cell lines with polymorphisms in the various mitochondrial proteins could be identified and the implicated genes cloned and sequenced. We have already used this approach to confirm our URF3 protein assignment [1]. Additional polymorphisms in mitochondrially synthesized proteins have been reported [12].

Our results strongly indicate that all of the mtDNA ORFs are expressed and that the proteins are not extensively modified. On the basis of this conclusion, we were able to calculate the proportion of protein molecules in each band in figure 4 by dividing the relative autoradiograph density of the bands by the number of methionines predicted for that protein from the gene assignment. We normalized all values to the labeling index of COI.

Comparison of the mitochondrially synthesized proteins from Epstein-Barr virus transformed lymphocytes of two different human cell lines (a normal individual [C] and a patient with mitochondrial myopathy [P]) revealed the same protein patterns but with a marked difference in the labeling of protein 10

(URF6) (fig. 4). Variation in protein 10 is a common finding in our studies. If protein 10 is the product of URF6, this variable expression might be explained by the fact that URF6 is the only gene located on the mtDNA light strand and thus under the control of an independent promoter [28–29].

These two cell lines and others also differ in the proportion of expression of the two proteins we have assigned to ATPase 6 and one of the cytochrome c oxidase proteins. If, as we have proposed, this protein is COIII, then such variation might be related to the close proximity of the ATPase 6 and COIII genes. Unlike most mtDNA genes, these two genes are contiguous and not separated by a tRNA [15]. Even so, they are translated from separate mRNAs that are generated by the precise cleavage of a common precursor between the ATPase 6 termination codon and the COIII initiation codon (5′-ATPase 6-UA*AUG*-COIII-3′) [30, 31]. Cleavage between the two As followed by polyadenylation would yield two active transcripts. By contrast, absence of cleavage or cleavage between A and U would destroy the COIII mRNA while cleavage 5′ to the UAA would destroy the ATPase 6 mRNA. Since ATPase 6 and COIII are essential components of the ATP synthetase and the electron transport chain, respectively, such differential cleavage could provide an effective regulatory mechanism for oxidative phorphorylation.

The results of our 2D-PD gel analysis suggest mtDNA coding sites for the various mitochondrially synthesized proteins. These assignments provide, for the first time, a direct link between the mitochondrial translation products and the mtDNA genes. Variation in mtDNA genes can now be identified simply by looking for variation in the mitochondrially synthesized proteins. As a consequence, the proposed assignments may prove of considerable value in searching for the molecular basis of maternally inherited diseases [32] such as the mitochondrial myopathies [33, 34] and Leber's optic atrophy [33, 35, 36].

ACKNOWLEDGMENTS

We thank N. B. Spinner and M. C. King for providing their lymphocyte labeling protocol in advance of publication.

REFERENCES

1. Oliver NA, Greenberg B-D, Wallace DC: Assignment of a polymorphic polypeptide to the human mitochondrial DNA unidentified reading frame 3 gene by a new peptide mapping strategy. *J Biol Chem* 258:5834–5839, 1983
2. Cleveland DW, Fischer SG, Kirschner, MW, Laemmli UK: Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J Biol Chem* 252:1102–1106, 1977
3. Oliver N, McCarthy J, Wallace DC: Comparison of mitochondrially synthesized polypeptides of human, mouse, and monkey cell lines by a two-dimensional protease digest gel system. *Somat Cell Mol Genet* 10:639–643, 1984
4. Bordier C, Crettol-Jarvinen A: Peptide mapping of heterogeneous protein samples. *J Biol Chem* 254:2565–2567, 1979
5. Lam KS, Kasper CB: Sequence homology analysis of a heterogeneous protein population by chemical and enzymatic digestion using a two-dimensional sodium dodecyl-polyacrylamide gel system. *Anal Biochem* 108:220–226, 1980
6. Tijssen P, Kurstak E: An efficient two-dimensional sodium dodecyl sulfate-

polyacrylamide gel electrophoresis method for simultaneous peptide mapping of protein contained in a mixture. *Anal Biochem* 128:26–35, 1983

7. OLIVER NA, WALLACE DC: Assignment of two mitochondrially synthesized polypeptides to human mitochondrial DNA and their use in the study of intracellular mitochondrial interaction. *Mol Cell Biol* 2:30–41, 1982
8. HOUMARD J, DRAPEAU GR: Staphylococcal protease: a proteolytic enzyme specific for glutamyl bonds. *Proc Natl Acad Sci USA* 69:3506–3509, 1972
9. OLIVER NA: *Genetic and Biochemical Analysis of Human Mitochondrial Variants*. Ph.D. dissertation, Stanford Univ., Stanford, Calif. (Univ. of Michigan, Microfilm), 1982
10. WALLACE DC: Assignment of the chloramphenicol resistance gene to mitochondrial deoxyribonucleic acid and analysis of its expression in cultured human cells. *Mol Cell Biol* 1:697–710, 1981
11. BIRD AG, MCLACHLAN SM, BRITTON S: Cyclosporin A promotes spontaneous outgrowth *in vitro* Epstein-Barr virus-induced B-cell lines. *Nature* 289:300–301, 1981
12. SPINNER NB, KING MC: Mitochondrially encoded protein polymorphisms as an assay for maternal inheritance. *Am J Hum Genet* 35:208A, 1983
13. LAEMMLI UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685, 1970
14. LOTT TJ, YANG J, YE J, WALLACE DC: The use of microcomputers for the quantitation of light intensity patterns using digitized video signals. *Comp Appl Biosci*. In press, 1986
15. ANDERSON S, BANKIER AT, BARRELL BG, ET AL.: Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465, 1981
16. BIBB MJ, VAN ETTEN RA, WRIGHT CT, WALBERG MW, CLAYTON DA: Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26:167–180, 1981
17. GEY GO, COFFMAN WD, KUBICEK MT: Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res* 12:264–265, 1952
18. KLEBE RJ, CHEN T-R, RUDDLE FH: Controlled production of proliferating somatic cell hybrids. *J Cell Biol* 45:74–82, 1970
19. ANDERSON S, DE BRUIJN MHL, COULSON AR, EPERON IC, SANGER F, YOUNG IG: Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. *J Mol Biol* 156:683–717, 1982
20. PEPE G, HOLTROP M, GADALETA G, ET AL.: Nonrandom patterns of nucleotide substitutions and codon strategy in the mammalian mitochondrial genes coding for identified and unidentified reading frames. *Biochem Int* 6:553–563, 1983
21. MACREADIE IG, NOVITSKI CE, MAXWELL RJ, ET AL.: Biogenesis of mitochondria: the mitochondrial gene (aapl) coding for mitochondrial ATPase subunit 8 in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 11:4435–4451, 1983
22. CHOMYN A, MARIOTTINI P, GONZALEZ-CADAVID N, ET AL.: Identification of the polypeptides encoded in the ATPase 6 gene and in the unassigned reading frames 1 and 3 of human mtDNA. *Proc Natl Acad Sci USA* 80:5535–5539, 1983
23. MARIOTTINI P, CHOMYN A, ATTARDI G, TROVATO D, STRONG DD, DOOLITTLE RF: Antibodies against synthetic peptides reveal that the unidentified reading frame A6L, overlapping the ATPase 6 gene, is expressed in human mitochondria. *Cell* 32:1269–1277, 1983
24. CHING E, ATTARDI G:High-resolution electrophoretic fractionation and partial characterization of the mitochondrial translation products from HeLa cells. *Biochemistry* 21:3188–3195, 1982
25. HARE JF, CHING E, ATTARDI G: Isolation subunit composition, and site of synthesis of human cytochrome c oxidase. *Biochemistry* 19:2023–2030, 1980
26. CHOMYN A, MARIOTTINI P, CLEETER MWJ, ET AL.: Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. *Nature* 314:592–597, 1985
27. BEATTIE DS, CHEN Y-S, CLEJAN L, LIN L-FH: Electrophoretic behavior of cyto-

chrome *b* in a partially purified preparation and evidence for high molecular weight associated mitochondrial translation products. *Biochemistry* 18:2400–2406, 1979
28. MONTOYA J, CHRISTIANSON T, LEVENS D, RABINOWITZ M, ATTARDI G: Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc Natl Acad Sci USA* 79:7195–7199, 1982
29. WALBERG MW, CLAYTON DA: *In vitro* transcription of human mitochondrial DNA. *J Biol Chem* 258:1268–1275, 1983
30. MONTOYA, J, OJALA D, ATTARDI G: Distinctive features of the 5′-terminal sequences of the human mitochondrial mRNAs. *Nature* 290:465–470, 1981
31. OJALA D, MONTOYA J, ATTARDI G: tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290:470–474, 1981
32. GILES, RE, BLANC H, CANN HM, WALLACE DC: Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77:6715–6719, 1980
33. EGGER J, WILSON J: Mitochondrial inheritance in a mitochondrially mediated disease. *N Engl J Med* 309:142–146, 1983
34. ROSING HS, HOPKINS LC, WALLACE DC, EPSTEIN CM, WEIDENHEIM K: Maternally inherited mitochondrial myopathy and myoclonic epilepsy. *Ann Neurol* 17:228–237, 1985
35. ERICKSON, RP: Leber's optic atrophy, a possible example of maternal inheritance. *Am J Hum Genet* 24:348–349, 1972
36. NOVOTNY EJ JR, SINGH G, WALLACE DC, ET AL.: A neurodegenerative disorder with generalized dystonia: A new mitochondropathy? *Neurology*. In press, 1986

THE VTH INTERNATIONAL CONGRESS ON TWIN STUDIES will be held September 15–19, 1986, in Amsterdam, The Netherlands, at the Grand Hotel Krasnapolsky. The following and related topics will be dealt with: biology of twinning, multiple maternities (perinatology, etc.), psychological aspects of twins and twinning (including behavior and psychiatric genetics), aging, cancer, diabetes, obesity, coronary heart disease, drug abuse, etc. For further information, write to the: Congress Secretariate, Professor Aldur W. Eriksson, Institute of Human Genetics, Medical Faculty, Free University, PB 7161, 1007 MC Amsterdam, The Netherlands; or to the: Congress Bureau (Congrex, Keizersgracht 610, 1017 EP Amsterdam, The Netherlands.