Progress Report on Grant DE-FG002-87ER60533
entitled, "Studies of Human Mutation Rates"
for period November, 1992-October, 1993

To: Mr. Peter Waldman, Acquisition and Assistance Unit, Department of
Energy, Chicago Operations Office, 9800 South Cass Avenue, Argonne,
IL 60439

From: Profs. James V. Neel and Samir M. Hanash, Principal Investigators,
Departments of Human Genetics and Pediatrics, University of Michigan
Medical School, Ann Arbor, MI 48109-0618

Date: October 27, 1993

The progress during 1992-93 with respect to ER 60533 (J. V. Neel and S. Hanash, co-PIs) can be summarized under three headings:

1. The development of two-dimensional DNA gels for the detection of mutation.

In recent years several different groups of investigators have advanced techniques for the two-dimensional display of some of the DNA fragments resulting from enzymatic digestion of isotope-labeled genomic DNA. Dr. J. Asakawa of the Radiation Effects Research Foundation saw in this development the potential for studies of new mutations in DNA, and significantly improved the techniques in several respects. Fig. 1 represents an autoradiograph of a two-dimensional DNA gel.

A new mutation in this system would be manifested as a DNA fragment in a child not present in either parent, or the loss of a fragment in a child present in homozygous condition in both parents. Here in Ann Arbor, we have been quite involved in the past 10 years in working with two dimensional polyacrylamide gel preparations of protein suspensions, also with the study of mutation in mind, and have developed rather sophisticated computer programs for the analysis of these gels. In this regard, it should

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

be mentioned that the analysis of these gels by eye would be extremely tedious, so that a computerized approach is highly desirable.

Dr. Asakawa spent 7 weeks in the summer of 1993 here in Ann Arbor, during which time we engaged in an all-out effort to determine whether these gels were suitable for mutation monitoring. He brought with him 18 gels, consisting of 3 mother/father/child trios, in duplicate, and these became the basis for our feasibility study. The first task was to determine a subset of highly reproducible fragments that were free-standing (i.e., did not overlap in position with other fragments). This task was made difficult by the fact that some 10% of the spots exhibited genetic variability, of a variety of types.

Eventually a subset of 482 fragments was selected as suitable for genetic analysis. In particular, this system is seen as particularly suitable for the detection of insertions/deletions/rearrangements (I/D/R) events in the DNA. Such events in a normal (two-corresponding fragments) individual would result in the loss of a fragment and a decrease of spot intensity by 50%. To detect such an event with accuracy, the coefficient of variation (CV) of a spot selected for study should not exceed 50%. All of the 482 spots selected met this criteria. Average spot size is approximately 2000 bp. Thus, at present, in one gel some $2000 \times 482 \times 2$ bp, or approximately 2×10^6 bp can be monitored for mutational events.

The kinds of variants encountered are illustrated in Figs. 2-6. Since the fragments are prepared by digesting genomic DNA with a selection of restriction fragment enzymes (NotI, EcoRV, and Hinf I), any nucleotide variation at these sites which interferes with the action of the enzymes will result in a variant. The clearest manifestation of this type of variant is variation in the vertical axis, such as is illustrated in Fig. 2. On the other hand, variation such as is represented by 1A and 1B in Fig. 6 probably results from an I/D/R-type event.

Forty-three of the 482 spots selected for study exhibited genetic variation. At present we estimate that approximately half of this is associated with I/D/R events. Ultimately, for meaningful studies of mutation, it will be necessary to distinguish with accuracy between these two types of events. To this end, back in Japan, Dr. Asakawa, employing the standard but, in this context, laborious techniques of molecular genetics is in the process of identifying the nucleotide composition of the spots selected for monitoring. This will permit identifying the types of DNA under inspection, and the precise nature of a mutation when one is detected.

A test of the genetic validity of the variants being recognized is possible by the analysis of the segregational behavior of the variants in mother/father/child trios. For this purpose, three types of genetic systems have been recognized, namely, 1) systems where a pair of segregating fragments have been identified (i.e., the genotypes are AA, AA', and A'A'); 2) systems where a null situation exists (i.e., the genotypes are AA, Ao and oo); and 3) systems where only two genotypes have been identified (AA, Ao). Table 1 summarizes the results of the genetic analysis. No exceptions to Mendelian genetics have been encountered, a fact which gives confidence in the present scoring of the variant phenotypes.

The trios employed in this analysis have been digitized and the images processed in duplicate. Surprisingly, we find by the appropriate analysis that more than half of the variation in spot intensity (but not position) is associated with the execution of the analysis, i.e., is an artifact produced by the equipment and algorithm. For the present this is discouraging but in the long run, very encouraging. With the improvements in image analysis which the future holds, a significantly larger number of spots should meet the criterion of a CV less than 0.12, and so the system should hold an even greater potential for monitoring for mutation.

Table 1. Segregation analysis with respect to the three classes into which the data on genetic variation have been subdivided.

a. Spots with respect to which a segregating variant has been recognized

Mating type	Child			Total
	AA	AA'	A'A'	
AA x AA	32			32
AA x AA'	14	14		28
AA' x AA'	1	1	2	4
AA x A'A'		5		5
	<u>47</u>	<u>20</u>	<u>2</u>	<u>69</u>

b. Spots which vary quantitatively but for which no segregating variant has been recognized but complete spot absence has been observed

Mating type	Child			Total
	AA	Ao	oo	
AA x AA	2			2
AA x Ao	2			2
AA x oo		2		2
Ao x Ao	1	1		2
Ao x oo			1	1
	<u>5</u>	<u>3</u>	<u>1</u>	<u>9</u>

c. Spots which vary quantitatively but for which no segregating variant has been observed nor has complete spot absence been observed

Mating type	Child			Total
	AA	Ao	oo	
AA x AA	21			21
AA x Ao	11	5		16
Ao x Ao		<u>2</u>		<u>2</u>
	<u>32</u>	<u>7</u>		<u>39</u>

2. The mitochondrial DNA of American Indians. Over the period of 1962-1976, one of the PIs (JVN), supported by DoE and NSF funding, was engaged in extensive studies of some of the surviving culturally intact tribes of American Indians. A legacy of that activity was a collection of some 10,000 blood samples of various types, which three years ago were divided among several laboratories for further study. An important product of this distribution the past year has been studies on the mitochondrial DNA of these Indians. A paper just published, acknowledging DoE support, can be summarized as follows:

"The mtDNA variation of 321 individuals from 17 Native American populations was examined by high-resolution restriction endonuclease analysis. All mtDNAs were amplified from a variety of sources by using PCR. The mtDNA of a subset of 38 of these individuals was also analyzed by D-loop sequencing. The resulting data were combined with previous mtDNA data from five other Native American tribes, as well as with data from a variety of Asian populations, and were used to deduce the phylogenetic relationships between mtDNAs and to estimate sequence divergences. This analysis revealed the presence of four haplotype groups (haplogroups A, B, C and D) in the Amerind, but only one haplogroup (A) in the Na-Dene, and confirmed the independent origins of the Amerinds and the Na-Dene. Further, each haplogroup appeared to have been founded by a single mtDNA haplotype, a result which is consistent with a hypothesized founder effect. Most of the variation within haplogroups was tribal specific, that is, occurred as tribal private polymorphisms. These observations suggest that the process of tribalization began early in the history of the Amerinds, with relatively little intertribal genetic exchange occurring subsequently. The sequencing of 341 nucleotides in the mtDNA D-loop revealed that the D-loop sequence variation correlated strongly with the four haplogroups defined by restriction analysis, and it indicated that the D-loop variation, like the haplotype variation, arose predominantly after the migration of the ancestral Amerinds across the Bering land bridge."

3. Molecular verification of a suggested phylogeny for the eight most common phosphoglucosyltransferase-1 (PGM1) alleles. In 1982, with Japanese colleagues, one of the PIs (JVN) suggested a phylogeny for eight of the common alleles of phosphoglucosyltransferase-1, which related them through three mutations and four intra-genic cross-over events. Now, ten years later, with the determination of the nucleotide sequence at this locus by the Hopkinson group at the Galton Laboratory, London, it became possible to test the validity of that hypothesis. The abstract of a paper now in press in the *Proceedings of the National Academy of Sciences USA* reads as follows:

"In 1982, we advanced a phylogeny which attributed eight alleles of the phosphoglucosyltransferase-1 locus (PGM1) to three independent mutations in a primal allele, followed by four intragenic recombination events involving these mutants [Takahashi, N., Neel, J. V., Satoh, C., Nishizaki, J., & Masunari, N. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6636-6640.] The recent description of a cDNA probe for this locus [Whitehouse, D. B., Putt, W., Lovegrove, J. U., Morrison, K., Hollyoake, M., Fox, M. F., Hopkinson, D. A., & Edwards, Y. H. (1992) *Proc. Natl. Acad. Sci. USA* 89, 411-415] now renders it possible to test the validity of this phylogeny. cDNAs of PGM1 reverse-transcribed from mRNAs obtained from Japanese individuals possessing eight different electrophoretically defined alleles (PGM1* 1+, PGM1* 1-, PGM1* 2+, PGM1* 2-, PGM1* 3+, PGM1* 3-, PGM1* 7+, PGM1* 7-) were amplified by the polymerase chain reaction (PCR) and the sequences were determined. Only three different base substitutions were identified when PGM1* 1+ was taken as the reference allele, as follows: an A to T transversion at residue 265, a C to T transition as residue 723, and a T to C transition at residue 1320. The second of these substitutions creates a BglII restriction enzyme site and the third a NlaIII site. At the amino acid level, these substitutions respectively alter the 67th amino acid from Lys to Met, the 220th amino acid from Arg to Cys, and the 419th amino acid from Tyr to His. These mutations resulted in the electrophoretic properties defining PGM1* 7+, the PGM1* 2+, and the PGM1* 1- alleles respectively. Subsequent intragenic recombinational events resulted in the remaining four alleles. For two of these

latter alleles (PGM1* 7- and PGM1* 3-), more than one type of intragenic cross-over can produce the allele. These findings verify the predicted phylogeny and provide a case study in the evolution of complexity at a genetic locus."

This paper reflects the collaboration of JVN with the Radiation Effects Research Foundation made possible by the DoE funding.

4. Protein identification in the face of N-terminal blocking. In the course of systematic studies of qualitative and quantitative variation in protein patterns of expression in lymphoid cells undertaken in part to identify normal physiological variation distinct from protein expression due to mutation, we previously described a 19-kDa phosphopeptide which was induced with proliferation in mature T cells and constitutively expressed in immature thymocytes. This low abundance protein is N-terminally blocked. To obtain internal amino acid sequence information, the protein was isolated from preparative 2-D gels by electroblotting resolved protein onto PVDF membranes. Protein was recovered from PVDF membranes using methodologies developed with DoE support, as described in our previous progress report. Sequence information was obtained for 68 residues in 6 tryptic fragments which permitted identification of the polypeptide as a phosphorylated form of dUTPase, which is thought to be the enzymatically active form. We also identified the non-phosphorylated form of dUTPase. Studies of the expression and phosphorylation of dUTPase in human T cells indicate that accumulation and phosphorylation of dUTPase in mature T cells occurs in a cell cycle dependent manner. Interestingly, non-cycling immature thymocytes express constitutively high levels of phosphorylated and unphosphorylated dUTPase. These results suggest an important role for dUTPase in immature thymocytes that is independent of proliferation. It is thought that dUTPase functions in part to reduce intracellular levels of dUTP thereby maintaining a low dUTP/dTTP ratio and minimizing misincorporation of uracil into newly replicated DNA. Incorporation of uracil into DNA elicits an excision repair mechanism involving uracil N-glycosidase, DNA polymerase I

and DNA ligase. The presence of high dUTP such as would exist when dUTPase levels are low results in continued misincorporation of uracil during the polymerase/ligase phase, potentially overwhelming the repair process.

5. Studies of protein structure using mass spectrometry. In a previous report we described the use of mass spectrometry to investigate the nature of the N-terminal blockage of Op18 (PNAS 88:5709, 1991). Op18 is a widely expressed, cell cycle regulated, phosphoprotein involved in signal transduction of a variety of stimuli. In actively proliferating Jurkat T cells which express Op18 at high level, phorbol 12-myristate 13-acetate (PMA) treatment induces a rapid increase in the level of several Op18 phosphorylated forms. To determine phosphorylation sites involved in the PMA effect, the major Op18 phosphorylated forms were resolved in Jurkat T cells, before and after treatment with PMA, using preparative immobilized pH gradient-based two-dimensional polyacrylamide gel electrophoresis. Tryptic fragments of phosphorylated Op18 were analyzed by two-dimensional thin layer peptide mapping and were resolved by reverse phase high performance liquid chromatography prior to analysis by matrix-assisted laser desorption ionization mass spectrometry. Phosphorylation sites were identified by further treatment of the proteolytic fragments with different enzymes and determination of the mass shifts by matrix-assisted laser desorption ionization mass spectrometry. Two major phosphorylation sites were identified. Low constitutive levels of phosphorylation at Ser²⁵ and Ser³⁸ in Op18a and Op18b was demonstrated. Treatment with PMA resulted in enhanced phosphorylation of Ser²⁵ in Op18a and of both Ser²⁵ and Ser³⁸ in Op18b. Taken together with prior studies of Op18 phosphorylation, the data suggest that Op18 phosphorylation occurs at identical sites in different tissues and organisms. These studies have been recently published (J Biol Chem 268: 14269, 1993).

6. Characterization of genes responsible for proteins identified in protein 2-D gel electrophoresis. We have identified by N-terminal microsequence analysis a 59.9 kD protein which had exhibited genetic variation in population surveys and which in ethyl nitrosourea mutagenesis experiments had exhibited an unusually high frequency of mutation. A direct N-terminal analysis of the first 29 residues of this protein indicated that the sequence is similar to both the beta subunit of prolyl 4-hydroxylase (also known as protein disulfide isomerase) and to thyroid hormone binding protein. Based on the reported cDNA sequences of these proteins, 7 PCR primers were synthesized corresponding to regions of identity within both cDNAs so that each PCR amplified fragment is about 250 bp in size with overlapping ends with neighboring fragments. A heterozygous parental type TK6 clone and a mutant clone expressing only one of the alleles as determined by 2-D gel analysis were then used to isolate poly (A)⁺ RNAs. Single stranded cDNAs prepared from these RNA preparations were then used as template for PCR amplifications with the primers. About 93% of the nucleotide sequences have been obtained from PCR fragments. Whereas the cDNAs for prolyl 4-hydroxylase (beta-subunit) and THBP differ at 28 bases, our sequence is identical to THBP at 20 of these 28 bases and identical to prolyl 4-hydroxylase at the other 8 bases. The remaining bases are identical in the three proteins. Of interest is the finding of a heterozygous sequence (either G or C at base 1066), coding for glutamic acid or glutamine respectively, in the PCR-amplified sequences of the parental type clone. This finding could account for the heterozygous phenotype (i.e. pI difference) observed on 2-D gels. Thus far only the G and C at base 1066 has been found in PCR amplified fragment clones of the mutant cells. About 7% of the mutant sequence remains to be determined. Single strand conformational polymorphism (SSCP) studies of six PCR-amplified clones spanning this region has revealed SSCP band differences among the clones, suggesting an additional sequence difference in the mutant clone. Sequencing of these clones is expected to be completed shortly. These results clearly indicate that the approach we are pursuing (identification of protein spots, microsequencing, probe preparation, and cDNA

Mr. Peter Waldman
October 22, 1993
Page 10

isolation and sequencing) will serve to identify and characterize new functional genes and mutations in these genes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

*Reprints removed
and cycled separately.*

Fig. 1

Portion of a two-dimensional DNA gel preparation.

Fig. 2

Fig. 3

Fig. 4

Fig. 5

1A and 1B are a polymorphic pair which probably differ due to *hinf*I variability. The other two pairs of spots show variability that is most likely due to small deletions between the *Not*I and *Hinf*I sites, so that the change in terms of number of base pairs in the same for both dimensions. There are two other pairs of variable spots in this section.

Fig. 6

1A/1B and 2A/2B are polymorphic pairs whose likely explanation is that the A allele in each case has a deletion of the *Hinf*I site of the B allele. This leads to a smaller *Not*I-*Eco*RV fragment for "A" in each case but in 1A/1B the A allele's *Not*I-*Hinf*I fragment is smaller than B's while for 2A/2B it is larger. 3A/3B are so close it is hard to say if *Hinf*I variability or deletion is the cause.

END

DATE
FILMED

3 / 18 / 1944