Progress Report on Grant DE-FG02-87ER60533
entitled, "Studies of Human Mutation Rates"
for period November, 1990-October, 1991

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

From: Prof. James V. Neel, Principal Investigator, Department of Human
Genetics, University of Michigan Medical School, Ann Arbor, MI
48109-0618

Date: 15 July 1991

The three objectives of the program are:

1) To isolate by the technique of two-dimensional polyacrylamide gel electrophoresis (2-D PAGE), proteins of special interest because of the relative mutability of the corresponding gene, establish the identity of the protein, and, for selected proteins, move to a characterization of the corresponding gene;

2) To develop a more efficient approach, based on 2-D PAGE, for the detection of variants in DNA, with special reference to the identification of a variant in a child not present in either parent of the child (i.e., a mutation); and,

3) To continue an effective interface with the genetic studies on the children of atomic bomb survivors in Japan, with reference to both the planning and implementation of new studies at the molecular level.

For administrative purposes, the program is subdivided into four sections, each under the direction of one of the four co-PIs; the progress during the past year will be summarized in accordance with this sectional structure.

Section I. Isolation of proteins of special genetic interest from two-dimensional polyacrylamide gel electrophoresis (Strahler)

a) Protein recovery from gels. The strategy for the isolation of sequencable amounts of proteins of interest continues to be the use of 2-D electrophoresis as the final purification step. Several proteins of special interest because of an increased frequency of spontaneous or

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

g22

induced (ethylnitrosourea) mutation and/or because of genetic variation encountered in population studies have been selected for analysis. The proteins are in the moderate to low abundance class, thus requiring partial enrichment prior to 2-D electrophoresis. Enrichment protocols include subcellular fractionation, removal of RNAs by RNase A digestion, acetone fractionation, and preparative liquid phase isoelectric focusing under denaturing conditions using Rotophor fractionation. Frequently dramatic increases are seen in the number and relative abundance of polypeptides visualized on analytical 2-D gels prepared from Rotophor enriched fractions. Unambiguous identification of the protein of interest has required the use of preparations containing the structural variant where available or the use of radiolabeled tracer protein recovered from "analytical" 2-D gels under electrophoretic conditions identical to those used for the initial identification of the variants. Thus far, four proteins of special interest because of their genetic variability have been prepared for sequence analysis, as well as an additional six "control" proteins.

The first step in obtaining protein for structural analysis is usually to electroblot protein from 2-D gels onto PVDF membranes for direct N-terminal sequence analysis. Unfortunately, thus far only one polypeptide of special interest has yielded amino acid sequence information (see task II) with this approach. This corresponds to the results of our past experience: of 51 polypeptides on which direct sequencing has been attempted, only 17 have yielded direct N-terminal sequence information. The remainder require a strategy to obtain an internal sequence. This requires recovering peptides in high yield a) from protein electroblotted onto PVDF membranes (discussed in task II) or b) from protein digested *in situ* in gel pieces, an approach which has been successful for moderate to high abundance proteins in several prior studies. In approach b, protein is visualized in 2-D gels by Coomassie-staining. The protein of interest is excised and, following extensive washing of the gel pieces, the protein is digested *in situ* with trypsin. Peptides are eluted from the gel, separated by microbore reverse phase HPLC, and submitted for Edman sequence analysis. A drawback to this approach is that protein is denatured/precipitated in the gel during the staining and is complexed with dye, thus decreasing the efficiency of digestion. In addition, bound Coomassie is extracted along with the peptides and obscures peptide peak identification in the subsequent microbore HPLC chromatogram. In a modification to this procedure, protein is visualized in 2-D gels using a negative-staining technique. Unbound SDS in the gel is complexed with Zn^{++} forming an opaque background. SDS-protein complexes are seen as clear areas in the gel. Sensitivity of detection is equal to or even greater than Coomassie staining. Protein is visualized within 5-10 min after staining, and the protein pattern is stable for several hours, allowing sufficient time to excise proteins of interest from the gel. Gel pieces are washed extensively with 50 mM EGTA and digested as before. In a study aimed at identifying the phosphorylation site of a protein designated as Opl8, this approach has yielded 90% recovery of radioactivity from the *in situ* digest in a single phosphopeptide. In contrast, Coomassie-staining and subsequent phosphopeptide isolation had previously yielded <60% recovery of radioactivity. The "negative staining" approach has been used to obtain peptides for internal sequence analysis for one of our proteins

of interest which previously was shown to be N-terminally blocked. Two other proteins in the special interest class have been purified and are awaiting digestion and peptide isolation.

b) Studies of protein structure using mass spectrometry. The results of our analyses indicated that a substantial fraction of polypeptides yielded no direct sequence, not because of insufficient material but rather due to N-terminal blockage. In a collaborative effort with Dr. Jack Watson, Director of the Mass Spectrophotometry Facility at Michigan State University, one polypeptide, designated Op18, previously known to be N-terminally blocked, and for which the full sequence has been deduced from a cDNA clone, was investigated to identify the nature of the block. Ten Coomassie-stained Op18 spots were cut from blots and a tryptic digest prepared. HPLC fractions from the digest were collected and analyzed by fast atom bombardment mass spectrometry (FAB-MS). The FAB mass spectrum of one of the HPLC fractions exhibited an MH⁺ ion at m/z 889, a mass not expected from the predicted amino acid sequence. However, if the predicted N-terminal tryptic fragment, MASSDIQVK, had been post-translationally modified by cleavage of the terminal methionine residue and N-acetylation of the adjacent alanine residue to yield Ac-ASSDIQVK, this modified peptide would produce a protonated molecule of m/z 889 during FAB. This interpretation was subsequently confirmed by subjecting the m/z 889 ion to collisionally activated dissociation tandem mass spectrometry (FAB-CAD-MS/MS). Fragment daughter ions in the FAB-CAD-MS/MS spectrum were consistent with the sequence assignment for this peptide. A complete series of overlapping daughter ions for the entire peptide could not be obtained, in part because the sample was analyzed at low levels (<10 pmole). Nevertheless, adequate sequence information was obtained to deduce that the protein was post-translationally modified *in vivo* via loss of the N-terminal methionine and subsequent N-acetylation of the penultimate alanine (manuscript in press, PNAS).

Phosphorylation is another common post-translational modification of proteins and is usually important in the regulation of activity of these biomolecules. Location of the phosphate group is often difficult using traditional Edman sequencing methods. FAB-CAD-MS/MS can be used to determine the location of the phosphate group. The ethyltri-phenylphosphonium (ethyl-TPP) derivatives of peptides have been shown to lower the limits of detection for hydrophilic peptides and to reliably produce an abundant and complete series of sequence-related fragment ions in FAB-CAD-MS/MS. This approach has been successfully used to identify the presence of phosphate and the site of phosphorylation in a model peptide (manuscript submitted). An isolated phosphopeptide from Op18 (discussed above) is pending analysis using this approach.

In collaboration with Dr. David Lubman (Department of Chemistry, University of Michigan), we have initiated studies to use immobilized exopeptidases coupled on-line via an ion spray interface with a mass spectrometer to analyze protein/peptide digestion products. By using stop flow, digestion time (residence time on the exopeptidase column) is controlled. A stable immobilized carboxypeptidase Y column has been prepared and used to test the kinetics of digestion of model peptides. At this point in the development of the procedure, analysis of digestion

products has been by reverse phase HPLC. In some cases, we have used the technique of "glow discharge" to introduce sample into the mass spectrometer. Refinements in the adaptation of an ion spray interface will lower the sensitivity of detection to the low- to sub-pmole level and will allow the immobilized CPY column to be put on-line to analyze N-terminally blocked peptides and proteins.

Section II. Preparation of partial amino acid sequences of the proteins isolated from gels (Andrews)

Four proteins which in our earlier experiments with the TK-6 cell line were shown to be mutable and one control protein have been subjected to automated protein sequencing. Only one of these proved to have a free amino terminus. Sequences for the four remaining proteins are being pursued as described in the final paragraph of this Section.

The one protein sequence obtained (Table I) was derived from the most mutable protein observed in this series of experiments (three mutations observed). Interestingly, the TK-6 cell line used is heterozygous at this locus and PVDF membrane preparations contain the normal and a mutant spot, so that if desired, the amino acid substitution occasioned by this particular mutation can be determined. A database search indicated sequence identity for the first 29 residues with human protein disulfide isomerase (PDI). This is a particularly interesting protein in terms of its mutability because it is the product of a fairly well-characterized gene which appears to play multiple roles in the cell.

Protein disulfide isomerase is a 491-residue protein which is synthesized as a precursor having a 17-residue signal peptide used for translocation into the endoplasmic reticulum (ER). PDI is not normally a secretory product of the cells but is retained in the ER by binding to the "salvage receptor" which is an intrinsic membrane protein. The recognition site on PDI is the carboxyl-terminal tetrapeptide which has the sequence KDEL. It has been suggested that PDI plays a role in catalyzing the rearrangement of disulfide bonds in proteins undergoing the folding process in the Golgi apparatus. It has also been identified as the beta subunit of prolyl 4-hydroxylase, as the cellular thyroid hormone binding protein (P55), as one of the subunits of the triglyceride transfer complex of the ER, and as a component of the glycosylation complex. As the cellular T3 receptor, it may be the major route of entry into cells for T3. This multitude of roles suggests that it is a particularly important protein in humans.

The four remaining proteins now under study are being subjected to a new method for obtaining internal sequence information. This method involves *in situ* cyanogen bromide cleavage of the protein transferred to a membrane from a two-dimensional gel followed by elution using a series of organic solvents. Past efforts have been limited by inadequate methods for eluting the protein fragments from the membrane. These problems now appear to be resolved by reversed-phase microbore HPLC. It is anticipated that this technique should provide useful internal sequence for most of the major mutable proteins during the next funding period of this grant.

TABLE I. The correspondence between the amino acid sequence of one of our "mutable proteins" and the sequence reported for protein disulfide isomerase.

Residue #	Residue Found	Protein Disulfide Isomerase
1	Asp	Asp
2	Ala	Ala
3	Pro	Pro
4	Glu	Glu
5	Glu	Glu
6	Glu	Glu
7	Asp	Asp
8	His	His
9	Val	Val
10	Leu	Leu
11	Val	Val
12	Leu	Leu
13	Arg	Arg
14	Lys	Lys
15	Ser	Ser
16	Asn	Asn
17	Phe	Phe
18	Ala	Ala
19	Glu	Glu
20	Ala	Ala
21	Leu	Leu
22	Ala	Ala
23	Ala	Ala
24	His	His
25	Lys	Lys
26	Tyr	Tyr
27	Leu	Leu
28	Leu	Leu
29	Val	Val

Section III. Preparation of DNA probes on the basis of sequenced proteins, and development of an improved method to identify mutations in the human genome (Kurachi)

In the past year, this laboratory continued its efforts on two projects, one to improve the two-dimensional denaturing gradient gel electrophoresis (2-D DGGE) method to identify mutations in DNA, using the factor IX gene as a model, and the other to isolate new genes of interest in relation to our mutation rate studies.

a) Improvement of methods to detect new mutations. In order to establish an improved, efficient method to identify all types of new mutations in genes (single base mutations/deletions/insertions), we have first established a multiplex polymerase chain reaction (PCR) using the human factor IX gene as a model. The problems described in last year's annual report have been solved and now all eight exons of factor IX can be efficiently amplified in a single reaction. The improvements included modification of the PCR primers used (designing primers to minimize "imbalanced" competition among them and to exclude primer dimer formation as much as possible) and changes in PCR buffer and reaction time conditions. The PCR fragments amplified ranged from 431 to 783 bp in size. An aliquot of the PCR product was used for subcloning and sequence analysis of all exons; the product was separated by electrophoresis on 5% polyacrylamide gel. The sequence analysis of a new mutant showed that a unique mutation in exon 8 (base 31008 C to T) changes amino acid Thr to Met (nucleotide numbering system in our previous paper, *Biochemistry* 24:3736, 1985). This amino acid residue is in the catalytic domain of factor IX (residue 116). This mutant gene was also found to have a G substitution at base 20422. DGGE clearly separated the G containing exon VI from the A-containing exon VI when the genomic DNA of the affected boy was mixed with a normal gene with A at the site and tested. The PCR fragment (783 bp in size) of exon 8 which contains the unique mutation (C to T at base 31008) did not originally appear abnormal in the DGGE analysis. However, this result was found to be due to its larger size, in consequence of which a longer-than-usual gel-run time is required to reach its denaturing zone on the gel compared to other, smaller fragments. We think that this failure to allow sufficient gel-run time has been a common problem of DGGE analysis as used by other researchers, as there are a number of reports describing DGGE analyses as not detecting a significant number of abnormal genes. With the solution of this problem, we now should be able to establish an efficient method combining multiplex PCR and two-dimensional DNA DGGE for mutation analysis. The testing of the methods with a number of abnormal factor IX genes is in progress. We have also tested the feasibility of the "single-strand conformation polymorphism" (SSCP) method for detecting nucleotide substitutions (Orita et al, *PNAS* 86:2766, 1989). This approach was not informative for the exon 8 fragment in our experiment. This may be due to the size of the fragment. Although we believe 2-D DGGE is superior to SSCP in detecting mutations (mostly point mutations), further tests are required to determine the possible usefulness of SSCP for our purpose.

b) Cloning of new genes. The isolation of a number of functional genes by utilizing protein 2-D gel electrophoresis is in progress. A

protein, identified as spot 176 with an apparent size of 33 kDa in the protein 2-D gel electrophoresis of human lymphoblasts, has been subjected to such study. This protein was found to be one of the invariable proteins in the protein 2-D gel analysis. A synthetic oligonucleotide probe designed from the amino terminal sequence obtained by microsequencing was prepared and was used in our early phase study. We have also prepared a second probe designed from the region adjacent to the amino-terminus for cross-screening tests of the positive clones we have identified and will identify in the screening of the cDNA library of leukemic T-cells (HSB2) with the first probe.

Section IV. The Japanese Interface (Neel)

Neel was in Japan from 15 November 1990 until 8 December 1990. There were three principal accomplishments during the stay. 1) A manuscript on the effect of parental exposure to the atomic bombs on mortality in their liveborn children, written in collaboration with W. J. Schull and Y. Yoshimoto, was finalized for publication as an RERF Technical Report. 2) In collaboration with Schull, Neel finalized introductory and final chapters for an anthology of the most significant of the studies of the genetic effects of the atomic bombs, to be published in September, 1991, by the National Academy Press. 3) The proposed program of the RERF for the study of mutations in the DNA of the children of survivors was reviewed and calculations made of the program's current efficiency. In addition, in collaboration with Dr. James Trosko, Chief of Research, RERF, plans were made for a workshop on the study of mutations in DNA, to be held in November, 1991.

An unexpected development while Neel was in Japan was the discovery by Dr. Awa's cytogenetic group of highly abnormal ("rogue") cells among cultured lymphocytes from Soviets living in the Ukraine. Such cells were first observed by Neel's group among cultured lymphocytes from isolated Amerindians (study under DoE-sponsorship), and subsequently also observed by Awa's group in presumably normal Japanese. With this latest discovery, there should be sufficient cells available for a statistical study of their attributes. Subsequent to his return to the U.S., Neel ascertained that the old Amerindian preparations were still in good condition (but required restaining). It is proposed that on the occasion of Neel's next trip to Japan, he and Awa will collaborate on an analysis of all the available "rogue cell" material.

In March of 1990, Neel had arranged for the extensive collection of Amerindian serum and red blood cell samples assembled under DoE sponsorship to be distributed among several laboratories, for a variety of studies. The first fruits of this distribution are now apparent, in the discovery by collaborators at the National Cancer Institute of a high frequency of individuals seropositive for the Human T-Lymphotropic Virus-type 2 (HTLV-2) in two very isolated Amerindian tribes. This discovery establishes the antiquity of this virus in Amerindians, suggesting it was present at the time of their arrival in North America.

A visit to this Department by Dr. Rem Sukernik of the Institute of Cytology and Genetics in Novosibirsk, in Siberia, was partially defrayed by this grant. Dr. Sukernik brought with him 250 serum samples from

various ethnic groups in Eastern Siberia. These sample will be studied for evidences of HTLV-2 infection, in an effort to trace the origins of the virus present in Amerindians.

Progress was made on a number of publications during the year. The anthology referred to above is now in press, has been indexed, and is expected out in September of 1991. The manuscript on mortality in the children of exposed, referred to above, has been prepared for journal publication and is in press. Neel has written a commentary on the genetic effects of radiation which will appear in the special August issue of the *JAMA* related to radiation effect. These three publications continue our theme that the genetic effects of ionizing radiation are less than have been projected for many years. Finally, a multi-authored manuscript describing improvements in the two-dimensional polyacrylamide gel technique, important to the objectives of this program, is in press in *PNAS*.

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.

END

**DATE
FILMED**

12/06/91