

DOE/ER/60533--5

DE93 003658

**Progress Report on DoE Grant 87 ER 60533**

**"The Study of Human Mutation Rates"**

**(J.V. Neel, P.I.)**

**1989 - 1992**

**University of Michigan Medical School**

**Ann Arbor, Michigan 48109**

**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.

April 28, 1992

**MASTER**

**DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED**

This progress report, covering the period November, 1 1989 - April 1992, will be essentially bipartite. First we will describe recent developments with which Neel is associated regarding the question of induced mutations in the survivors of the atomic bombings of Hiroshima and Nagasaki. We will also, as part of that section, describe some interesting developments with respect to the Amerindian blood samples collected under DoE sponsorship between 1964 and 1982. Then we will describe the developments in Ann Arbor, primarily the work of Strahler, Kurachi and Andrews, regarding the application of two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) to the study of genetic variation and mutation affecting protein characteristics. In particular, we will report on the identification and isolation of genes of especial interest as reflected in the behavior of the proteins which they encode.

#### **Section I. The genetic effects of the atomic bombs; some Amerindian studies (J.V. Neel)**

##### **1. Findings in Hiroshima and Nagasaki.**

A major analysis and synthesis of the data on the genetic effects of the atomic bombs collected at ABCC-RERF<sup>1</sup>, involving all the material accumulated over a 40-year period, was projected for the mid-1980s. Unfortunately, in 1982, it developed that there had been major errors in the estimation of the amount of radiation received by the survivors of the atomic bombings. Not until 1986 was a new algorithm available for computing exposures, the Dosage Schedule 1986 (DS86), and even this did not encompass some 20% of the survivors with 'complicated' exposure histories, forcing us to develop interim dosage estimates (cf. Otake *et al.*, 1990).

Within the past three years (which we will equate to the period of this grant support), Neel has been associated with three major analyses of components of the Japanese data, concerned with 1) congenital malformation and/or stillbirth and/or early mortality (collectively termed an untoward pregnancy outcome or UPO) (Otake *et al.*, 1990); 2) malignant tumors during

<sup>1</sup> These genetic studies were originally within the framework of the Atomic Bomb Casualty Commission (ABCC), which organization was superceded in 1975 by the Radiation Effects Research Foundation (RERF).

the first two decades of life (Yoshimoto et al, 1990), and survival among live-born infants (Yoshimoto et al, 1991). These data have been combined with earlier data on mutations involving protein charge or activity (Neel et al, 1988), cytogenetic abnormality (Awa et al, 1987), sex ratio (Schull et al, 1966), and physical development of children (Neel and Schull, 1956; Furusho, 1978 et seq.) in an attempt to develop an estimate of the human doubling dose of acute ionizing radiation (Neel et al, 1989, 1990).

The estimation of a genetic doubling dose of radiation for humans has presented many problems not encountered by the geneticist who estimates a doubling dose for *Drosophila* or mice. First, none of the indicators employed in this study is significantly altered by parental exposure. The experimentalist in this situation can increase the dose and/or enlarge the sample; this option is not open to the genetic epidemiologist. However, when all the various regressions of indicator on dose which have been derived are considered, the net effect of parental exposure on the indicators is positive, i.e., in the direction of a radiation effect. On the thesis that this is the most significant body of such data that exists, and that we can assume that some mutations were induced in the survivors, we have elected to explore what these data imply concerning the human genetic doubling dose of acute ionizing radiation. [The doubling dose is that amount of radiation which will produce the same mutational spectrum as occurs spontaneously each generation; the simplicity of this definition belies some of the complexities in the situation, complexities we have discussed elsewhere (Neel et al, 1990)]. The decision to calculate a doubling dose required that we estimate for such complex indicators as UPOs, early death, and cancer, what fraction of these outcomes in children is due to mutation in the preceding (parental) generation.

We will not reproduce the precise calculations here (cf. Neel et al, 1990). With no statistical significance, it is possible that to some extent we are only manipulating the "noise" in the system in estimating a doubling dose. On the other hand, the data, without equivocation, can and have been used to place minimal (95% probability) estimates on the doubling doses for various indicators. These suggest a value of not less than approximately 1.0 Sv. The value for

the most probable doubling dose is between 1.7 and 2.2 Sv. Because the error term for this estimate should reflect not only the sampling error but also any error introduced by fallacious assumptions, it is difficult to attach an error to the estimate. However, we believe we have been conservative in our assumptions and treatment of the data and that the true value of the doubling dose for the acute radiation exposures of Hiroshima and Nagasaki is unlikely to be less than 1.0 Sv.

For the past 40 years, the principle basis for estimating the genetic risks of radiation for humans have been the experimental results obtained with mice. There is a vast literature on mouse radiation genetics. In the early 1970s, on the basis of experiments with male mice, national and international committees endorsed a figure of approximately 0.4 Sv as the doubling dose for acute ionizing radiation and 1.00 Sv as the doubling dose for chronic or intermittent radiation, a figure extrapolated to humans. The data primarily responsible for that estimate have been summarized by Lüning and Searle (1971). Since that time, a large amount of additional data have been accumulated. The estimate derived from the findings in Hiroshima and Nagasaki was sufficiently disparate from the estimate based on experiments with mice to lead Neel and Lewis (1991) to undertake an analysis of all the mouse data available at the time. Because mice are born at a very immature stage of development (as compared with humans) and in litters (with more intrauterine and immediate postnatal competition for survival than for humans), it is inappropriate to compare mice and humans with respect to many of the reference points in the Hiroshima-Nagasaki study. On the other hand, Neel and Lewis (1991) did find in the literature the results of studies on eight different specific locus or specific phenotype systems in mice. The simple, unweighted average of the doubling dose of acute radiation computed for these eight systems was 1.4 Sv.

We note in passing that the system which more than any other has been employed by murine geneticists to generate data from which risks could be estimated is the very convenient and much used "7-locus" system of Russell (1951). It has yielded several very important insights into mammalian radiation genetics. However, we believe this system is biased toward relatively

high mutation rates. We have observed in somatic cell studies of ENU-induced mutations in a lymphocytoid cell line that loci known to exhibit genetic variation exhibit a higher mutation rate than loci not associated with variation (Chu et al, 1988). The development of the Russell system required that at least two alleles be known for each locus employed. We believe that the selection of such loci inadvertently introduced a bias toward the inclusion of more mutable loci in the system.

The radiation to which humans will usually be exposed is chronic (continuous or small pulses) rather than acute. It has now been abundantly demonstrated that if a given dose of, say, 3.0 Sv is delivered in 10 spaced installments, the genetic yield is substantially less than when the same dose is delivered acutely. The magnitude of the "dosage factor" varies with the endpoint under consideration as well as the acute dose of radiation employed. In mouse radiation genetics, a factor of 3 has in the past been applied to extrapolating from the effects of acute to chronic radiation. When we apply this factor to the results of our reanalysis of the data on the acute radiation of mice, the doubling dose becomes 4.2 Gy. For the lower gonadal doses which the survivors of Hiroshima and Nagasaki sustained, we have suggested that a dose rate factor of 2 is more appropriate. The doubling dose for chronic radiation in humans then becomes 3.4 - 4.4 Sv, a value in excellent agreement with our revision of the estimate derived from the mouse experiments.

Because the reports on the results of the studies in Japan have extended over so many years and are scattered through a diverse literature, Neel and Schull (1991) undertook to edit a book which republished 13 of the most significant studies on the genetic effects of the bombs, with a preface to provide perspective and an epilogue which addressed the need (and techniques) for continuing studies. This epilogue, as well as a recent editorial in *JAMA* (Neel, 1991), in addition addresses a problem which haunts epidemiological studies of radiation effects, namely, the issue of false positives. This issue is especially well illustrated by the epidemiological study of Gardner and associates (1990) on the excess of childhood leukemia and lymphoma in the vicinity of the Sellafield nuclear reprocessing plant in the United Kingdom. In a multifaceted

case control study for environmental associations with these diseases, they reported a significantly increased relative risk (RR) for childhood leukemia if the fathers were exposed to ionizing radiation in the course of their employment at the Sellafield nuclear plant. Four children were found to have fathers exposed to 0.1 - 0.2 Sv total of intermittent radiation, where the expectation was for 0.6 children, leading to an RR of 6.4 (confidence intervals 1.5 - 25.8). These findings are simply incompatible with the data from Japan as well as the murine experimental data, and are also internally inconsistent, in that a radiation dose large enough to result in such an increase in leukemia (which in humans is not a "genetic" tumor), should have resulted in major additional effects. We mention this development to emphasize the continuing need for the application of the results of the hard-won Japanese data to human problems (see Proposed Program).

Some five years ago, the Scientific Counsellors to RERF recommended that the study of the genetic effects of the atomic bombs progress to the DNA level, and since that time, the appropriate staff at RERF have been involved in establishing the cell lines on which these studies will be based and exploring the appropriate DNA technologies. On November 12-14, 1991, a major workshop was held on this subject at RERF, in which Neel participated; together with Dr. C. Satoh and Dr. R. Myers, he is currently engaged in preparing for publication a report on that workshop. The current strategy involves, as the material for these studies, establishing 500 family constellations of lymphocytoid cell lines derived from an exposed father and/or mother and one or more children, and another 500 derived from unexposed parents and one or more of their children. Because the number of cell lines from children is so relatively small for a study of this type, it will be necessary to screen large amounts of DNA from each child with reference to variation (and possible mutation), if the study is to add significantly to the present data base. The emphasis of the workshop was accordingly on techniques for the efficient detection of variants of DNA. The workshop agreed that at present the most efficient single technique appears to be one in which RERF staff already has a considerable investment, the denaturing gradient gel electrophoresis of Fischer and Lerman (1983), and that it would be appropriate for staff to initiate

a pilot study in which this was the "backbone" technique. At the same time, it was recognized that a number of possibly better techniques are lurking in the wings, and this technique was unlikely to continue to be the best approach over the next ten years (the estimated duration of the study). The preparation of a consensus report is proving to be somewhat difficult, but it is estimated a report will go to press before this grant year is out.

An unexpected product of Neel's continuing collaboration with Japanese investigators at RERF involves "rogue cells." These are cytogenetically extremely abnormal cultured lymphocytes arrested at the first cell division. We (with DoE support) were the first to describe such cells, in the members of two very isolated villages of Yanomama Amerindians (Bloom et al., 1970), but the fact that two years later the cells had almost entirely disappeared cast some doubt on the validity of the findings (Bloom et al., 1973). Subsequently, they were reported in Englishmen by two different groups (Fox et al., 1984; Tawn et al., 1985). In the study of Tawn et al (1985), it was possible to restudy the two affected persons 50 days later, at which time it was observed that the cells had disappeared, an observation which renders less troubling our inability to confirm our original observation on Amerindians in studies two years later. The next major development was the observation of these cells in low frequency in control individuals for the cytogenetic studies being conducted in Hiroshima and Nagasaki under the auspices of RERF (Awa and Neel, 1986). The most recent sighting is now in samples sent to RERF in connection with follow-up studies on the biomedical effects of the Chernobyl disaster. The cells were observed in 8 of 24 persons from the noncontaminated, control village of Krasilovka, 30 km southwest of Chernobyl.

A manuscript describing the findings is nearing completion. The epidemiologic data suggest a viral trigger, but it is not clear whether the damage is directly due to this trigger or whether the trigger somehow activates some of the many retroposon sequences known to be present in human DNA, and they are responsible for the complex damage. The ultimate biomedical significance of these rogue cells is at this time conjectural. The chromosomes of most affected cells are so severely scrambled that the cells would in general self-destruct at a cell

division, in which case we have defined an interesting bit of biological trivia. There may be, however, much more exciting possible implications of the data. We present evidence that in the individuals with rogue cells, "simple" chromosomal damage is also increased, and suggest that in the spectrum of damage presented by rogue cells, some of the least damaged cells may be able to complete mitosis. In lymphocytes, such cells could possess the rearrangements which are sometimes a point of departure for leukemia-lymphoma. If the process occurs in other types of somatic cells, here, too, it could play a role in oncogenesis. Finally, if it occurs in the germ line, it could result in some of the chromosomal abnormalities responsible for congenital defects, and the stable aberrations could conceivably play a role in the chromosomal rearrangements of evolution.

The prevailing view of the origin of the chromosomal abnormalities encountered in studies of "spontaneous" oncogenesis, teratogenesis, and mutagenesis is that they are for the most part stochastic in origin, inherent in the complexity of the DNA molecule. We now hypothesize that some significant fraction of these chromosomal abnormalities result from a previously unrecognized process which is still very poorly defined and understood. We further hypothesize that to the extent the process depends on a viral trigger, it may be susceptible to manipulation, as by immunization procedures. There are thus both theoretical and practical aspects to this development. --

## 2. Developments with respect to stored Amerindian blood samples.

Between the years 1964 and 1982, the DoE (together with several other agencies) funded our extensive study of the Indian populations of Central and South America, representatives of some 16 tribes in all. These activities resulted in the accumulation in liquid nitrogen storage of one or more 1 cc aliquots of serum and of packed erythrocytes from approximately 10,000 different Amerindians. This became a truly unique resource. With the official retirement of Neel in 1985, the field and laboratory work associated with this activity was tapered down, although Neel and collaborators continued to develop publications based on this material. During this period, considerable thought was devoted to the ultimate disposition of the samples.

A question central to any planning was whether DNA could be recovered from the samples in useful quantities (no leukocytes had been put down). With the advent of the polymerase chain reaction (PCR), our potential collaborators conducted the appropriate pilot studies, and it became clear that for multi-copy DNA (such as mitochondrial DNA), recovery was no problem.

Accordingly, in March of 1990, these samples were distributed among three different laboratories, namely: 1) the laboratory of W. Blattner/R.J. Biggar at the National Cancer Institute, Bethesda, MD; 2) the laboratory of D. Wallace at Emory University, Atlanta, GA; and 3) the laboratory of M. Stoneking/K. Weiss at Pennsylvania State University, College Park, PA. Each of these groups has a different focus, the NCI group being primarily interested in evidences of retroviral infection, the Wallace group in mitochondrial DNA (mtDNA), and the Penn State group in single copy DNA, with reference to a variety of genes of anthropological-population interest. The latter group has just generated from the National Science Foundation the funding to enable them to pursue their program and has no data to report at this time. The other two groups were adequately funded at the time of sample distribution, and already we can report some exciting results.

With respect to the retroviral studies, a survey of 13 tribes has revealed foci of high endemicity for the human T-cell lymphotropic virus-type II (HTLV-II) in two quite isolated tribes of Gé-speaking Indians, the Kraho and Cayapo. The virus was detected by serological approaches but the infection confirmed by the recovery of viral DNA from the serum samples with the PCR reaction and the appropriate studies of this DNA. This viral infection, whose medical significance is yet to be defined, has been encountered in scattered ethnic groups over the world but also, and especially, in intravenous drug users. The question of its primary origin is quite controversial. In the present situation, it is very unlikely that the virus has been recently introduced to these tribes, and it must be presumed to have been present in some of the forebearers of the present Amerindians, forebearers who migrated across the Bering land bridge some 20,000 to 40,000 years ago. This finding, now in press (manuscript appended), sets the stage for a number of future studies, to be described in the application for renewed support.

With respect to the studies of mtDNA, some background is necessary. In the past 30 years, a major genetic effort has centered on classifying populations according to the frequency of various very objective biochemical traits determined by nuclear genes. These studies are now rapidly being expanded to the DNA level. More recently, there has been interest in phylogenies based on mtDNA. Since mtDNA appears to evolve more rapidly than nuclear DNA, there are advantages in working with mtDNA -- but also some problems. The literature is currently quite contentious with reference to the limits to inference with this approach. Previous publications of our collaborators concerning the mtDNA of the indigenous peoples of the Americas have suggested three waves of migration which, in order of their recency, are Eskimo, Na-Dene, and Paleoindian, these waves given dates of 4,000-5,000 years ago, 6,000-10,000 years ago, and 20,000-40,000 years ago respectively. Our collaborators have previously examined mtDNA from three tribes of Paleoindians. All of our tribal samples (16) involve Paleoindians, and what we are proposing to do is type representatives of all these tribes and undertake statistical treatments to work out the evolutionary relationship of the tribes. The fact that we have already studied extensively the genetic attributes of these tribes with respect to biochemical markers will enable us to make a detailed comparison of the phylogeny suggested by the mtDNA with that suggested by nuclear genes. Furthermore, if HTLV-II is ultimately detected in a sufficient number of Amerindian tribes, and a sufficient number of genetic markers studied in HTLV-II, then one can construct phylogenetic relationships based on HTLV-II. The RNA-DNA of this virus appears to evolve even more rapidly than mtDNA. There is then, with time, the prospect of comparing phylogenies based on three types of DNA with differing rates of evolution.

At present, 160 samples from members of 11 of the tribes in our collection have been examined with 14 restriction endonucleases. The task of analyzing these data and integrating the new findings with the results of previous studies is just under way. One of the unexpected earlier findings in the studies of biochemical traits based on nuclear DNA was the relatively high frequency of "private polymorphisms." These are defined as unique alleles which attain polymorphic frequencies (>2% heterozygotes) in single tribes. These are proving to be relatively

common in the mtDNA, more common than for nuclear gene traits, and have the potentiality for establishing hierarchies of descent which should be very useful in the reconstruction of tribal histories.

Finally, this grant period has witnessed the publication acknowledging grant support of two papers which deal with aspects of Amerindian genetics, reflecting the momentum of the previous Amerindian program, and one paper synthesizing the results of past studies on mutation rates. The first (Chakraborty and Neel, 1989) describes a method for the simultaneous estimation of effective population size and mutation rate from Amerindian population data, developed while Chakraborty was a visiting investigator supported by this grant. On the basis of mathematics supplied by Chakraborty and data and a conceptual framework supplied by Neel, a method was developed for utilizing population data on electrophoretic variants of proteins to estimate simultaneously the effective sizes ( $N_e$  values) of the populations in question and the rate of mutation resulting in electromorphs at the loci whose products were surveyed.

The method was applied to data from 12 relatively unacculturated Amerindian tribes for whom census data and independent estimates of the number of different electrophoretic variants at 27 loci are available. Because of tribal demographic structure,  $N_e$  should be less than the current number of reproductive-aged adults. In fact, it is substantially greater for 7 tribes, most likely due to intertribal migration and a recent decrease in tribal size. Estimates of locus mutation rates for the 27 loci vary by more than a factor of 20, with an average of  $1.1 \times 10^{-5}$  per locus per generation. This latter estimate is in satisfactory agreement with the results of other indirect approaches to the estimation of mutation rates in these tribes but about two times higher than the results of direct estimates based on these same loci in studies on civilized populations. This discrepancy could be due to the above-hypothesized migration and to decreases in tribal size.

The second paper (Barrantes et al, 1990) deals with the complex relationships of the Chibcha-speaking tribes of Central America. The summary of that paper reads as follows: "There is evidence that Amerindians have continuously occupied the lower Central American

Isthmus for as long as 10,000 years. There remains some doubt about the relationships of these original colonizers to the resident peoples of this zone at the time of European contact (approximately A.D. 1500). We present new genetic data for up to 48 genetic loci for 570 members of 6 Chibcha-speaking tribes of lower Central American -- the Boruca, Bribri, Cabecar, and Guatuso of Costa Rica and the Kuna and Teribe of Panama -- and delineate the genetic affinities among the various groups (these six tribes and the Guaymi and Bokota) of lower Central American. We convert standard genetic distance metrics into a form that is linear with the effective time since divergence, and we compare the genetic distances with linguistic distances for the same groups ( $r = .74$ ,  $P < .001$ ). Geographic affinity accounts for some of the genetic divergence among groups ( $r = .49$ ,  $P < .084$ ) and for some of the linguistic divergence ( $r = .53$ ,  $P < .037$ , but the correspondence between geographic position and taxonomic affinity is not high. We combine all of the genetic and linguistic data to construct a synthetic overview taxonomy of the lower Central American Chibcha. Both the genetic and linguistic data exhibit hierarchical organization of tribal groups, showing a general east-to-west pattern of groups, with greater affinities between close neighbors. The presence of private genetic variants of some antiquity within the region and their absence outside the zone, coupled with the essential absence of the DI\*A polymorphism of mongoloid origin that is widespread outside the zone, argue for a relatively isolated development of the Central American Chibcha. Our results do not support the old view of lower Central America as a frontier between more advanced cultures to the north and south. Any such explanation would require recent waves of migration from outside the region, migration that is not compatible with either the genetic or linguistic data or with the archaeological history of the region."

The third paper, by Neel (1990), compares much of the data on genetic variation detected with one-dimensional electrophoresis collected under DoE funding, with similar data on a different group of proteins collected with two-dimensional polyacrylamide gel electrophoresis with NCI funding, to argue that there are locus differences in mutability related to protein "class." It appears that genes encoding proteins which, broadly termed, are structural may well

have on average lower mutation rates than the genes encoding proteins which function as enzymes or in the serum transport of ligands. The argument is indirect, based on the greater genetic variation encountered in the latter proteins, and our previous demonstration (Chu et al, 1988) that the genes encoding the variable proteins exhibit higher ENU-induced mutation rates than the genes encoding less variable (i.e., monomorphic) proteins. The practical significance of this finding is to underscore the need for care in the selection of any battery of proteins (or the DNA corresponding to the genes encoding for these proteins) in studies of radiation or chemically-induced mutation.

**Section II. Identification of genes responsible for proteins of genetic/physiologic interest (J. Strahler, K. Kurachi and P. Andrews).**

In the past three years, this portion of the DoE-supported program has been primarily concerned with developing and applying the technologies which enable an investigation to proceed most efficiently from a protein of interest visualized with two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) to the identification and mapping of the gene encoding for this protein. In this program, our primary objective has been to contribute to an understanding of why some genes seem so much more prone to genetic variation and mutation than others, as well as to the identification of post-translational and physiological variations in proteins that could interfere with the study of genetic variation. Although the effort during the current funding was targeted toward proteins of special interest as defined above, clearly this program could be extended to include all the proteins of lymphocytes and other cell types which can be recovered in sufficient quantities for amino acid sequencing (see Proposal).

**1. Obtaining purified proteins.**

The basic approach to obtaining purified protein utilizes the high resolution of 2-D PAGE as the final step in the purification. Two fundamental problems have been addressed in the progress presented below. One is the ability to recover protein from 2-D gels in amounts sufficient to yield reliable sequence information for a sufficient number of amino acid residues to identify unambiguously the polypeptide in a sequence database or, if the protein represents a

novel sequence, to obtain sufficient material that sequence information enabling the construction of an oligonucleotide probe having low redundancy for cloning purposes can be constructed. Since the majority of the proteins of interest visualized on 2-D gels represent moderate to low abundance proteins in human lymphoid cells, enrichment protocols which allow unambiguous identification of the target protein(s) in the 2-D gel "assay" are essential. A second obstacle to sequencing is that of N-terminal blockage, whether artificially produced during purification or naturally occurring *in vivo*. The majority of soluble proteins in mammalian cells are believed to be N-terminally blocked, most frequently as a result of N-acetylation. To meet these problems, we have investigated in some detail several of the steps in the production of preparative 2-D gels, and have also improved our strategy for generating internal amino acid sequences.

To date we have obtained partial sequence information for some 25 proteins isolated from preparative 2-D gels of human lymphoblastoid cell lines. Of the 25 sequences, 20 were obtained by N-terminal sequencing of proteins electroblotted onto either PVDF (polyvinylidene difluoride) membranes or GlassyBond (siliconized glass fiber sheet, Biometra) filters. With this technique the number of gels needed to generate material for sequencing varied from as few as 2 for a highly abundant protein to as many as 15. The amount of protein sequenced in these studies ranged from 0.8 to 20 picomole. Fifty or 72% of the proteins for which we have attempted direct N-terminal sequencing failed to yield a sequence presumably due to N-terminal blockage. We have obtained internal sequences for 5 of these proteins. Tryptic peptides for these proteins were obtained either by *in situ* digestion of protein in Coomassie Blue stained spots, or more recently by digestion of protein eluted from PVDF membranes (discussed further below). With the former technique, the number of 2-D gels required to obtain 1 to 15 picomole of purified peptide was substantially greater (20 to 80 gels) than with electroblotting.

Of the 69 proteins examined, 26 were of special interest because of demonstrated genetic variation either in population studies (9 proteins) and/or in spontaneous or induced mutation studies (17 proteins). Of these 26 proteins, we have obtained sequence information for 6 proteins. The results are summarized in Table 1. A pair of spots were found to be the products

Table 1. Summary of protein sequence data obtained during current grant period

spot number*	special interest	protein identity/designation	pl	MW kD	Location of fragment	sequence
w2		protein H precursor (R13)	4.03	79.5	N-terminal	VEVKPRGVSLTN
w22		glucose regulated 78K protein precursor	5.00	71.4	N-terminal	EEDKKEGVTVWGLD
w26		RO/SS-A antigen	4.00	68.0	N-terminal	EPAVYKEQ
w28	poly/mutation	prolyl 4-hydroxylase (beta subunit)	4.39	59.9	N-terminal	DAPEEEDHVLVLRKSNFAEALAAHRYLL
np28		B23**	4.61	40.6	Internal**	XXEL(G)V(S)XLXE
w29	poly/mutation	prolyl 4-hydroxylase (beta subunit)	4.44	59.9	N-terminal	DAPEEED
w35	mutation	-	5.81	70.0	N-terminal	ASEAIGGA
w79	poly/mutation	phospholipase C, form 1	6.03	59.9	N-terminal	SDVLELTDFNNXNSKR
w85		HuCha;P1 matrix protein	6.03	61.5	N-terminal	AKDVFQGADARALMLQGVDL
w136		kinase related transforming protein (Brat)	5.93	44.4	Internal	NQSDVYIA
w159	polymorphic	-	6.73	39.0	N-terminal	IIGGGTG
w174		PCNA	4.20	36.0	N-terminal	MFEARLYQGSILK(K)V
w176		alternative splicing factor-assoc. 32K chain	4.07	32.7	N-terminal	LHTDGDKAFLVDFLSDEIKE(E)(R)(E)KQIH(K)TXL
w179	polymorphic	elongation factor 1-beta	4.14	30.4	N-terminal	GFGDLKSPAGLQVNLND
w208		triosephosphate isomerase	6.76	27.1	N-terminal	APSRRFFVGGNMWAWR
w210		triosephosphate isomerase	6.90	27.3	N-terminal	APSRRFFV
w228		nm23; NDP kinase	6.20	19.8	Internal	TATPQQAQEVHEK
w229		-	6.03	20.4	Internal	AHALEGLIGV
w230		Op18a. phosphorylated	-	-	Internal	VTNGAFTGEIS
w232		-	-	-	Internal	FMQASEDLITFIAKPDGVCRWMLGET
w235		glutathione S-transferase pi	6.00	18.6	N-terminal	(T)FIAIKPDGVVEQR(R)
w248		tumor protein (R15)	5.80	17.7	Internal	WMLGET
w255	EIF4D	-	-	-	Internal	DASEHATAPT(R)
w256	EIF4D	-	-	-	Internal	AFGYDLXGAXDYTP
w235		-	-	-	Internal	XXGYDLYXAYDYNTPXXE
w248		-	-	-	Internal	TDQIALPSGKYG
w255	EIF4D	-	-	-	Internal	XXDVGXGIVDDEMYR
w256	EIF4D	-	-	-	Internal	FYPEEEVQALDDTER
w230		-	-	-	Internal	MWVGILKII
w232		-	-	-	Internal	XKENVSVGPPEFP(LS)
w235		-	-	-	Internal	XXVFX(F)PL(S)P
w248		-	-	-	Internal	XXQAFELXL
w255	EIF4D	-	-	-	Internal	PPYTWWFP
w256	EIF4D	-	-	-	Internal	MNFRDLSHDEMFSDDM
w235		-	-	-	Internal	ATFPQMSAFLRK
w248		-	-	-	Internal	ATF

\* Spot designations with the prefix w were scored in mutagenesis experiments (Chu et al, 1988). np28 was not scored.

\*\* An internal peptide sequence was obtained for a lower molecular weight protein which led to our identification of B23 by immunological methods.

of alleles of the prolyl 4-hydroxylase  $\beta$  subunit; the responsible locus has been found to be unusually mutable in one of the lymphoid cell lines used in these studies (discussed below). Three of the remaining proteins are associated with genetic variation while the locus for one protein has been observed to have mutated in induced mutation studies. Two of these latter four proteins yielded novel sequences.

A more detailed description of the methodologies utilized and developed follows:

As a first approach to obtaining sequence information on polypeptides of interest, we examined our ability to purify polypeptides on 2-D gels in sufficient quantities to obtain direct N-terminal sequence information and to determine the amount of material necessary with existing "state of the art" microsequencing capability. Two approaches to the generation of preparative 2-D gels were followed which differed in the methodology used for the first dimension isoelectric focusing separation.

In the first approach, conventional carrier ampholyte (CA)-based isoelectric focusing was employed (250  $\mu$ g total protein per gel), essentially as we have done in prior sequencing efforts. Protein spots visualized with Coomassie Blue at this relatively high "load" retained overall resolution comparable to that obtained for "analytical" gels. About 100 spots are visualized in preparations of lymphocytoid cells, of which <50 are of sufficient intensity to warrant attempts at isolation for sequencing. Increasing the amount of protein per gel (400-650  $\mu$ g) results in drastic deterioration of the first dimension separation. Some pI regions of the gel, however, do maintain the required resolution. We have obtained material for sequencing using a protein load of 250  $\mu$ g soluble cytosolic or nuclear protein per gel. Proteins were electroblotted onto GlassyBond membranes using a semi-dry electroblotting apparatus.

In the first year of the present grant period, prior to the arrival at Michigan of Dr. Andrews, the protein sequencing aspects of this project have been contracted with the Max Planck Institute of Biochemistry in Martinsried, Germany. Employing the above mentioned methodology, we have made two different sets of polypeptides available to them. In one group, 28 major polypeptides were isolated for direct N-terminal sequence analysis, 8 of which

represented the products of loci which had mutated in our prior ENU mutagenesis studies (Chu et al., 1988) or which represented loci at which spontaneous genetic variation had been observed. Sequence information was obtained for 7 of the 28 polypeptides. Six of them were known proteins while one represented a novel sequence. Twenty-one, or 75%, of the polypeptides yielded no sequence information presumably due to N-terminal blockage. In another group of polypeptides isolated from CA-based 2-D gels of soluble nuclear polypeptides, 18 were selected for analysis. Of these, 8 were the products of loci at which protein variants had been recognized. Eleven polypeptides yielded sequence information. Ten of the 11 were known proteins, but one yielded a novel sequence. Three of the 11 were in the special interest category, i.e., 3 had exhibited known genetic variation in population surveys, and, of these 3, 2 had also exhibited induced variation in the ENU mutagenesis studies (sequence information was obtained for the common allele in all cases). These three proteins were subsequently found to be known human proteins [ $\beta$ -subunit of prolyl 4-hydroxylase, 60 kD chaperonin (HuCha), and elongation factor 1-beta]. Two polypeptides were common to both the cytosolic and nuclear sets [proliferating cell nuclear antigen and 60 kD chaperonin (HuCha)], and both yielded sequence information.

Thus, of the 46 different polypeptides recovered from CA-based 2-D gels and which were in the moderate to high abundance class of polypeptides, sequence information was obtained for 16 of them by direct N-terminal sequence analysis. The vast majority of polypeptides visualized on 2-D analytical gels are below the abundance level necessary for this approach. To obtain sufficient protein to provide internal sequence information for those proteins yielding non-useful or no sequence data, large numbers of gels (50-100 or more) would be required.

As an alternative approach, we have used immobilized pH gradient (IPG) gels for polypeptide separation in the isoelectric focusing dimension. The pH range chosen for these studies (pH 4-7) is comparable to the pH range obtained in CA-based 2-D gels used in our previous sequencing efforts and in prior screening programs. In such gels about 1-1.2 mg total protein can be loaded on such pH gradient gels, resulting in 4-5 fold more protein per gel. About 200 proteins are visualized with Coomassie Blue staining, a substantial increase in the number of

"sequencable" proteins from the same number of gels as contrasted with CA-based separations. While the majority of spots identified on CA-based 2-D gels can be identified also on IPG-based gels, occasionally a spot of interest is not readily identifiable on the IPG-based pattern. We have utilized five known structural variants (identified on CA-based 2-D gels) to verify the location of these proteins on IPG-based gels. To assess the number of gels required to obtain sequence information, IPG-based 2-D gels were prepared from whole cell extracts of a lymphoid cell line (Jurkat). Polypeptides were isolated by electroblotting onto PVDF membranes (Immobilon P) using a semi-dry blotting apparatus. Fourteen polypeptide spots were excised from 6 to 10 blots for N-terminal sequence analysis. Five polypeptides yielded sequence information with initial yields of 2-11 pmoles. Four corresponded to known proteins and one yielded a novel sequence. Nine polypeptides did not yield sequence information, presumably because of N-terminal blockage since they were present in equivalent amounts to the other five. Twenty or more blots derived from CA-based 2-D gels would have been needed to harvest the amount of protein resulting in these sequencing results. The increased protein capacity and maintenance of high resolution separations on IPG-based 2-D preparative gels significantly reduces the number of gels needed for a given separation. A description of this method has been published (Hanash et al, 1991) as well as a solicited methods paper (Strahler and Hanash, 1991).

Polypeptide spots electroblotted onto either GlassyBond or PVDF membranes have yielded comparable results when equivalent amounts of polypeptide were studied. For high abundance proteins such as proliferating cell nuclear antigen (PCNA), triose phosphate isomerase, and HuCha, as few as two electroblots from either type of membrane provided the protein from which sequence information could be obtained, whereas 6 blots (IPG-based 2-D gels blotted onto PVDF) or 20 blots (CA-based gels blotted onto GlassyBond) were required for the moderately abundant protein glutathione S-transferase. Conversely several polypeptides which did not yield any sequence from PVDF blots also did not produce sequence information from GlassyBond.

## 2. Obtaining internal amino acid sequences.

Three approaches have been used to obtain internal amino acid sequences. In one approach, protein is excised from Coomassie Blue-stained gel pieces. Following extensive washing and drying of the gel pieces, protein is digested *in situ* with trypsin (~1:10; enzyme:protein) and the peptides recovered by sequential elution with 75% TFA in water followed by 50% TFA in acetonitrile. Tryptic peptides are separated by reverse phase HPLC (Vydac 204 TP column). In one example using this method, we have explored the use of liquid phase isoelectric focusing on a large scale as the first step to purify a very low abundance protein from a complex protein mixture. The target protein in this study was a low abundance 28 kD nuclear protein (np28) which is induced in the Jurkat T cell line following activation with phorbol esters and calcium ionophor or is constitutitively expressed in tat III transfected Jurkat cells. The protein did not yield a sequence by N-terminal sequencing. To obtain sufficient material for generation of internal peptide sequences, a nuclear extract derived from  $4.5 \times 10^9$  transfected cells was fractionated under denaturing conditions (urea, non-ionic detergent and dithioerithritol) similar to the condition of the first dimension of 2-D gel separations, using the Rotophor apparatus (Biorad) operated at 12 w constant power for 6 h. The preparation was fractionated according to the manufacturers instructions into 20 fractions. Np28 was enriched about 5 fold in two acidic Rotophor fractions, which were used to purify the protein on preparative 2-D gels. Gels were stained with Coomassie Blue and np28 excised from 40 gels. Following *in situ* digestion, peptide recovery and separation, reliable sequence information was obtained for two of the isolated peptides, allowing identification with a known highly conserved nucleolar phosphoprotein B23 (manuscript in preparation).

A second example utilizing this approach is the identification of nucleoside diphosphate kinase, also known as nm23. In this instance, the protein was isolated from a neuroblastoma cell line for sequence analysis. Three tryptic peptides were isolated yielding 28 total amino acid residues. A protein database search revealed a complete match of the three sequences to nm23-H1, which has been shown to be identical to the A subunit of NDP kinase. Subsequent studies

using a rabbit polyclonal antibody established the identity of nm23 in 2-D gels of lymphoid cell proteins. Aspects of this work were funded by an NCI grant to this laboratory.

The second approach involves electroblotting onto PVDF membranes, visualization of the proteins with Coomassie Blue and elution of the protein from the membrane prior to digestion. We have evaluated several methods to elute protein from Coomassie Blue-stained PVDF membranes preliminary to proteolytic digestion in solution. DMSO has been described as an efficient extraction agent; however, in our hands considerable polymeric material in the PVDF membrane is also extracted; this material interferes with subsequent efforts to remove DMSO prior to digestion. SDS is an efficient elution agent; however, it is difficult to remove and interferes with subsequent hplc separation of peptide fragments. We have also evaluated the use of Tween 20 to extract phosphorylated polypeptides from PVDF membranes. Proteins were detected by Coomassie Blue staining and phosphoproteins were detected after  $^{32}\text{P}$  labeling using a Phosphoimager. The amount of protein extracted from isotopically labeled spots was determined by Cerenkov counting. Recoveries of radioactivity for 20 spots ranged from 48% to 98% and was moderately correlated with the amount of protein as determined by intensity of Coomassie Blue staining. Fourteen phosphopeptides had recoveries greater than 70%, and, of these, two were undetectable by Coomassie Blue staining. Recovery did not correlate with pI; however, there was moderate correlation of decreased recovery with increased molecular weight. Tween 20 can be removed from the sample either by TCA precipitation and subsequent washing of the precipitate with acetone or by reverse phase hplc. Tween 20 is eluted in the unretained fraction. Using the method of TCA precipitation, we have obtained internal sequences for several "standard" proteins run on the 1-D gels used to develop the method as well as for three proteins (triose phosphate isomerase, kinase-related transforming protein Braf, and one unidentified protein) obtained from 2-D gels of total cytosolic proteins from Jurkat cells. As a test of the sensitivity of the method, we demonstrated that sufficient quantities for sequencing of one protein known to us from prior N-terminal sequencing efforts to yield a sequence (triose phosphate isomerase) could be isolated from only four blots. The expected sequences were

obtained for three tryptic fragments derived from this protein. Similarly, one of the two previously uncharacterized proteins mentioned above was isolated from only four blots, and two tryptic peptides of this protein provided sufficient sequence at the 3 pmole level to identify the protein in a sequence database. The third of these proteins, a low abundance protein, was isolated from 80 2-D gel PVDF spots. Sequence information was obtained for 68 residues in 6 tryptic peptides which permitted identification of 79.4% homology (67.6% identity) with a viral protease-like protein. Homology of our human eukaryotic protein with the viral protein was considerably higher than the homology of this protein with its supposed protease homologues. In parallel work funded by the NCI, the homologue to the viral protein has been found to be a polypeptide whose expression is greatly upregulated during G1 in mitotically stimulated T lymphocytes. A manuscript is in preparation describing these findings. A second manuscript describing the methodological aspects of this approach is nearing completion and will be submitted shortly.

In another study,  $^{32}\text{P}$ -labeled proteins were eluted as described above, yielding >98% recovery of radioactivity from the PVDF membrane for each of two 18 kD phosphorylated forms of the Op18 protein, designated Op18a and Op18b. Tween 20 was removed by hplc using a C4 reverse phase column. Recovery of radioactivity was 95% and 90% respectively for Op 18a and Op18b. We have isolated tryptic phosphopeptide fragments of Op18a and are using this material to identify the sites of phosphorylation using mass spectrometric approaches. Partial Edman sequences were obtained for three peptides derived from Op18a to establish its identity as a phosphorylated form of the previously described non-phosphorylated form of Op18.

A drawback to the techniques of *in situ* digestion of Coomassie Blue stained proteins 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  $\text{Zn}^{++}$ , resulting in 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 Op18, this approach has yielded 90% recovery of radioactivity from the *in situ* digest. 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 currently being processed.

### 3. Characterization of nuclear DNA binding proteins.

Calcitrol induced terminal differentiation of HL-60 cells is accompanied by a number of changes in phosphorylation of nuclear phosphoproteins. *In vitro* phosphorylation of calcitrol-treated HL-60 nuclei revealed phosphorylation changes in five proteins which could be detected on 2-D gels. One of these, a 66 kD polypeptide, p66, remained associated with the nuclear pellet following high salt extraction of isolated nuclei. Partially purified phosphorylated p66 was identified in the analytical 2-D gel pattern of total cellular proteins and Western blot analysis of 2-D gels was used to identify the protein as lamin B. This work, only partially supported by this grant, is currently in press (Mertell et al, in press).

Mutations in the enhancer region of the  $\gamma$ -globin genes in individuals with hereditary persistence of fetal hemoglobin point to the importance of this region in the regulation of  $\gamma$ -globin gene expression. In prior studies using a combination of gel retardation and footprinting as assays, we have isolated two proteins from K562 nuclear extracts which recognize the sequence spanning +2352 to +2455 bp from the  $\gamma$ A-globin gene CAP site. This sequence contains a copy of TATACT, recently identified in the  $\gamma$ -globin gene as a protein (GF-1) binding site. In a competition assay, one of the two proteins we have identified was competed by the

fragment of the  $\gamma$ -globin gene promoter spanning -160 to -195. Further characterization of the native form of this protein by gel filtration HPLC, indicated that it was a part of an ca. 500 kD complex. A band corresponding to a 100 kD polypeptide was detected by SDS electrophoresis, following DNA affinity chromatography and southwestern blotting with either DNA fragment. By southwestern blotting, a similar binding activity was detected in T cells (Jurkat) and in HL-60 cells. However, the activity corresponded to 100 kD and 65 kD polypeptides, respectively. We purified the protein responsible for the binding activity in K562 nuclear extracts using a combination of ion exchange HPLC, DNA affinity chromatography and SDS-PAGE. The observed binding to both  $\gamma$ -globin gene enhancer and promoter and the apparent erythroid-specificity of the 100 kD polypeptide suggested that this protein may play an important role in  $\gamma$ -globin expression. In an extension of this work funded by the current grant, internal amino acid sequence analysis of the 100 kD band indicated that it was composed of two proteins, nucleolin and the ribosomal assembly protein poly(ADP-ribose) polymerase. Identification of these proteins indicated that the above described DNA binding properties of the 100 kD band were not erythroid specific.

#### 4. Studies of protein structure using mass spectrometry.

As we have seen, a substantial fraction of polypeptides were not amenable to direct sequencing, 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 (Hanash et al. 1991).

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 use of the ethyltri-phenylphosphonium (ethyl-TPP) derivatives of peptides has been shown to lower the limits of detection for hydrophilic peptides and to produce an abundant, reliable 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 (Watson et al., 1991). An isolated phosphopeptide from Op18 (discussed above) is pending analysis using this approach.

We have also initiated, in collaboration with Dr. David Lubman (Department of Chemistry, University of Michigan), pilot studies on the use of 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.

##### **5. Improvement of the methodology for detecting mutations in DNA: some results.**

During the current grant period, the DNA-oriented aspects of this proposal have focused on two projects, one to develop a two-dimensional denaturing gradient gel electrophoresis technique (2-D DGGE) to detect mutations in DNA, and one other to isolate and characterize new genes of interest in relation to our mutation studies.

a) Development of methods to efficiently detect mutations. In order to establish an improved, efficient method to identify all types of mutation in a gene (single base mutations, subtle and gross deletions, as well as insertions), a multiplex polymerase chain reaction (PCR) was first established for the human factor IX gene as a model. In this approach, all exons are amplified in a single reaction. Because of the differences in activity of Taq polymerase used in amplifying sequences with different sizes, the amounts of primers used to amplify each exon (or exons in a combined PCR amplification) were varied. In this reaction, exons 2 and 3 were amplified as one fragment because of the small size of intron 2 (188 bp), while all other exons were amplified separately. Primers used in the PCR reaction are 50 ng for exons 7 and 8, 400 ng for exon 4, 100 ng for exon 1, 150 ng for exon 2/intron2/exon 3 combined, and 200 ng for exons 5 and 6. These conditions were determined to be optimal to obtain approximately equimolar amounts of 7 PCR fragments which contain all the exon sequences and their splicing junction and the 5' flanking sequence (up to -490). The improvements of the PCR conditions included the modifications of the PCR primers used (designing primers minimizing the imbalanced competition among them and the possible primer dimer formation), PCR buffer, and reaction time conditions. The PCR fragments amplified ranged from 431 to 783 bp in size. Aliquots of the PCR product were used for subcloning the amplified fragments and sequence analysis of all exons. The rest of the product was separated by electrophoresis on 5% polyacrylamide gel (the

first dimension to separate fragments by fragment sizes). In this study we have identified a unique mutation in exon 8 (base 31008 C to T) which alters amino acid Thr 296 to Met in the factor IX gene of an affected family [nucleotide numbering system as in our previous paper (Yoshitake et al, 1985)]. No mutations were found in all other exons or the 5' end promoter region. This amino acid residue is in the catalytic domain of factor IX (residue 116). The PCR fragment (783 bp in size) of exon 8 which contains the unique mutation at base 31008 was originally observed to be non-informative in the DGGE analysis. However, this was found to be due to its larger size, in consequence of which this fragment takes a much longer time to reach its denaturing zone on the gel compared to other, smaller fragments in a single 2-D DGGE analysis. We think that slowly migrating fragments have been a common problem of DGGE analysis for other investigators, since a significant number of reports describing DGGE analyses as not being informative have appeared under circumstances in which the technique should work. Realizing and solving this problem, we now should be able to establish an efficient method combining multiplex PCR and two-dimensional DNA DGGE for mutation analysis.

In the DGGE analysis of the factor IX gene, a known heterozygosity at base 20422 (exon IV) serves as a convenient control of the efficiency of the method. This heterozygosity consists of the occurrence of either G or A, with, consequently, either a Ala or Thr at the corresponding position in the factor IX gene. The mutant factor IX gene with T at base 31008 was found to have an A at the position (base 20422) mentioned earlier, with the attendant substitution of Thr. DGGE clearly separated the PCR-amplified exon VI fragment with G at base 20422 from the fragment with A at this position. These PCR-amplified DNA fragments were prepared by using the genomic DNA of the affected male patient and normal genomic DNA with A at base site 20422. The testing of the methods with a number of abnormal factor IX genes is in progress. We have also tested the usefulness of the single-strand conformation polymorphism method (SSCP) in identifying DNA variation (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 potential of SSCP for our purpose. A manuscript describing the combination of multiplex PCR and 2-D DGGE is in preparation.

b) Isolation and characterization of genes responsible for proteins identified in protein 2-D gel electrophoresis. A 33 kD protein, designated W176 in our human lymphoblastoid cell line TK6 protein 2-D gel numbering scheme, has been subjected to characterization of its DNA sequences. This protein was one of the monomorphic proteins for which we have amino-terminal sequence information. A synthetic oligonucleotide probe sequence designed from the amino terminal sequence was prepared and was used to isolate cDNA clones for the protein. A second probe has also been designed based on the amino terminal region adjacent to the previously mentioned probe and has been used for cross screening the positive clones we have identified by screening a leukemic T-cell (HSB2) cDNA library with the first (N-terminal) probe. Attempts to identify positive clones by screening the library (about  $5 \times 10^5$  clones in pVc) with these two probes radiolabeled with  $^{32}P$  are in progress. If this approach is not successful, screening an expression cDNA library with monoclonal antibodies raised against the protein will be carried out. For raising antibody, the protein spots will be collected from multiple preparative gel electrophoretic runs, as will be described in the Proposal for Renewal.

Our most noteworthy progress results from analysis of a genetically variable protein identified from the protein 2-D gel electrophoresis analysis of TK6 cells. 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 the neighboring fragments. A heterozygous parental type TK6 clone and a mutant clone (ENU4-10D20) expressing only one of the alleles as determined by 2-D gel analysis were then used to isolate poly (A)<sup>+</sup> RNAs. Single stranded cDNAs prepared from these RNA preparations were then used as template for PCR amplifications with the primers. Nucleotide sequences obtained from the

PCR amplified fragments indicate that the cDNA prepared from the cells is identical to the sequence of prolyl 4-hydroxylase ( $\beta$ -subunit), but not to that of the above mentioned thyroid hormone binding protein. Interestingly, a heterozygous sequence (either G or C at base 1066), coding for glutamic acid or glutamine respectively, was found 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. Whether or not the G or C change at base 1066 is the only base difference found in heterozygous cell cDNA will be determined when the complete sequences are obtained. Thus far only C has been found in PCR amplified fragment clones of the mutant ENU4-10D20 TK6 cells. These results clearly indicate that the approach we are pursuing (identification of protein spots, microsequencing, probe preparation, and cDNA isolation and sequencing) will serve to identify and characterize new functional genes and mutation in these genes.

### Bibliography for Progress Report, DoE 87 ER 60533

Awa, A. A., T. Honda, S. Neriishi, T. Sufuni, H. Shimba, K. Ohtaki, M. Nakano, Y. Kodama, M. Itoh and H. B. Hamilton. 1987. Cytogenetic study of the offspring of atomic bomb survivors, Hiroshima and Nagasaki. In *Cytogenetics*. G. Obe and A. Basler (eds). Springer-Verlag, Berlin-Heidelberg. 166-183.

Awa, A. A. and J. V. Neel. 1986. Cytogenetic "rogue" cells: What is their frequency? *Proc. Natl. Acad. Sci. USA*. 83: 1021-1025.

Barrantes, R., P. E. Smouse, H. W. Mohrenweiser, H. Gershowitz, J. Azofeifa, T. D. Arias II and J. V. Neel. 1990. Microevolution in Lower Central America: Genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* 46: 63-84.

Bloom, A. D., J. V. Neel, K. W. Choi, S. Iida and N. Chagnon. 1970. Chromosome aberrations among the Yanomama Indians. *Proc. Natl. Acad. Sci. USA*. 66: 920-927.

Bloom, A. D., J. V. Neel, T. Tsuchimoto and K. Meilinger. 1973. Chromosomal breakage in leukocytes of South American Indians. *Cytogenet. Cell Genet.* 12: 175-186.

Chakraborty, R. and J. V. Neel. 1989. Description and validation of a method for simultaneous estimation of effective population size and mutation rate from human population data. *Proc. Natl. Acad. Sci. USA*. 86: 9407-9411.

Chu, E. H. Y., M. Boehnke, S. M. Hanash, R. D. Kuick, B. J. Lamb, J. V. Neel, W. Niezgoda, S. Piviroto and G. Sundling. 1988. Estimation of mutation rates based on the analysis of polypeptide constituents of cultured human lymphoblastoid cells. *Genetics*. 119: 693-703.

Fischer, S. G. and L. S. Lerman. 1983. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory. *Proc. Natl. Acad. Sci. USA*. 80: 1579-1583.

Fox, D. P., F. W. Robertson, T. Brown, A. R. Whitehead and J. D. M. Douglas. 1984. Chromosome aberrations in divers. *Undersea Biomed. Res.* 11: 193-204.

Furusho, T. and M. Otake. 1979. A search for genetic effects of atomic bomb radiation on the growth and development of the F<sub>1</sub> generation. III. Stature of 12- to 14-year-old junior high school students in Hiroshima. *RERF-TR*. Hiroshima. Radiation Effects Research Foundation: 14-79.

Furusho, T. and M. Otake. 1978a. A search for genetic effects of atomic bomb radiation on the growth and development of the F<sub>1</sub> generation. I. Stature of 15- and 17-year old senior high school students in Hiroshima. *RERF-TR*. Hiroshima. Radiation Effects Research Foundation: 4-78.

Furusho, T. and M. Otake. 1978b. A search for genetic effects of atomic bomb radiation on the growth and development of the F<sub>1</sub> generation. II. Body weight, sitting height, and chest circumference of 15- to 17-year-old senior high school students in Hiroshima. *RERF-TR*. Hiroshima. Radiation Effects Research Foundation: 5-78.

Furusho, T. and M. Otake. 1980. A search for genetic effects of atomic bomb radiation on the growth and development of the F<sub>1</sub> generation. IX. Stature of 12- to 14-year-old junior high

school students in Hiroshima. *RERF-TR*. Hiroshima. Radiation Effects Research Foundation: 1-80.

Furusho, T. and M. Otake. 1985. A search for genetic effects of atomic bomb radiation on the growth and development of the F<sub>1</sub> generation. V. Stature of 6- to 11-year-old elementary school pupils in Hiroshima. *RERF-TR*. Hiroshima. Radiation Effects Research Foundation: 9-85.

Gardner, M. J., M. P. Snee, A. J. Hall, C. A. Powell, S. Downes and J. D. Terrell. 1990. Results of case-control study of leukemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *Brit. Med. J.* 300: 423-429.

Hailat, N., D. R. Keim, R. F. Melhem, X. X. Zhu, C. Eckerskorn, G. M. Brodeur, C. P. Reynolds, R. C. Seeger, F. Lottspeich, J. R. Strahler and S. M. Hanash. 1991. High levels of P19/nm23 protein in neuroblastoma are associated with advanced stage disease and with N-myc gene amplification. *J. Clin. Invest.* 88: 341-345.

Hanash, S. M., J. R. Strahler, J. V. Neel, N. Hailat, R. Melhem, D. Keim, X. X. Zhu, D. Wagner, D. A. Gage and J. T. Watson. 1991. Highly resolving two-dimensional gels for protein sequencing. *Proc. Natl. Acad. Sci. USA*. 88: 5709-5713.

Kato, H., W. J. Schull and J. V. Neel. 1966. A cohort-type study of survival in the children of parents exposed to atomic bombings. *Am. J. Hum. Genet.* 18: 339-373.

Kusumoto, H., S. Hirosawa, J. P. Salier, F. S. Hagen and K. Kurachi. 1988. Human genes for complement components C1r and C1s in a close tail-to-tail arrangement. *Proc. Natl. Acad. Sci. USA*. 85: 7307-7311.

Luning, K. G. and A. G. Searle. 1971. Estimates of the genetic risks from ionizing radiation. *Mutat. Res.* 12: 291-304.

Martell, R. E., J. R. Strahler and R. U. Simpson. Identification of lamin B and histones as 1,25-dihydroxyvitamin D<sub>3</sub> regulated nuclear phosphoproteins in HL-60 cells. *J. Biol. Chem.* In press.

Neel, J. V. 1990. Average locus differences in mutability related to protein "class": A hypothesis. *Proc. Natl. Acad. Sci. USA*. 87: 2062-2066.

Neel, J. V. 1990. Unfolding perspectives on the genetic effects of human exposures to radiation. In DNA Damage and Repair in Human Tissues. B. M. Sutherland and A. D. Woodhead (eds). Woodheads. Plenum Press, New York. 337-349.

Neel, J. V. 1991. Update on the genetic effects of ionizing radiation. *JAMA*. 266: 698-701.

Neel, J. V. and S. E. Lewis. 1990. The comparative radiation genetics of humans and mice. *Annu. Rev. Genet.* 24: 327-362.

Neel, J. V., C. Satoh, K. Goriki, J.-I. Asakawa, M. Fujita, N. Takahashi, T. Kageoka and R. Hazama. 1988. Search for mutations altering protein charge and/or function in children of atomic bomb survivors: Final Report. *Am. J. Hum. Genet.* 42: 663-676.

Neel, J. V. and W. J. Schull. 1991. The Children of Atomic Bomb Survivors. National Academy Press, Washington, D.C. vi and 518.

Neel, J. V. and W. J. Schull. 1956. The Effect of Exposure to the Atomic Bombs on Pregnancy Termination in Hiroshima and Nagasaki. National Academy of Sciences-National Research Council, Washington, D.C.

Neel, J. V., W. J. Schull, A. A. Awa, C. Satoh, H. Kato, M. Otake and Y. Yoshimoto. 1990. The children of parents exposed to atomic bombs: Estimates of the genetic doubling dose of radiation for humans. *Am. J. Hum. Genet.* 46: 1053-1072.

Neel, J. V., W. J. Schull, A. A. Awa, C. Satoh, M. Otake, H. Kato and Y. Yoshimoto. 1989. Implications of the Hiroshima-Nagasaki genetic studies for the estimation of the human "doubling dose" of radiation. *Genome*. 31: 853-859.

Orita, M., H. Iwahana, H. Kanazawa, K. Hayash and T. Sekiya. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA*. 86: 2766-2770.

Otake, M., W. J. Schull and J. V. Neel. 1990. Congenital malformations, stillbirths, and early mortality among the children of atomic bomb survivors: A reanalysis. *Radiation Research*. 122: 1-11.

Russell, W. J. 1951. X-ray induced mutations in mice. *Cold Spring Harbor Symp. Quant. Biol.* 16: 327-336.

Strahler, J. R. and S. M. Hanash. 1991. Immobilized pH gradients: Analytical and preparative use. *Methods: A Companion to Methods in Enzymology*. 3: 109-114.

Tawn, E. J., C. L. Cartmel and E. M. T. Pyta. 1985. Cells with multiple chromosome aberrations in control individuals. *Mutat. Res.* 14: 247-250.

Torres-Rosado, A., K. S. O'Shea, A. Tsuji, S.-H. Chou and K. Kurachi. Hepsin, A novel membrane protease, is essential for mammalian cell growth. *Science*. Submitted.

Watson, J. T., D. S. Wagner, Y.-S. Chang, J. R. Strahler, S. M. Hanash and D. Gage. 1991. Characterization of the ethyl-triphenylphosphonium derivative of model peptides by fast atom bombardment collisionally-activated dissociation tandem mass spectrometry using B/E linked scans. *Int. J. Mass Spec. Ion Proc.* 111: 191-209.

Yoshimoto, Y., J. V. Neel, W. J. Schull, H. Kato, M. Soda, R. Eto and K. Mabuchi. 1990. Malignant tumors during the first 2 decades of life in the offspring of atomic bomb survivors. *Am. J. Hum. Genet.* 46: 1041-1052.

Yoshimoto, Y., W. J. Schull, H. Kato and J. V. Neel. 1991. Mortality among the offspring ( $F_1$ ) of atomic bomb survivors, 1946-85. *RERF-TR*. 1-91. Hiroshima: Radiation Effects Research Foundation: 1-27.

Yoshitake, S., B. G. Schach, D. C. Foster, E. W. Davie and K. Kurachi. 1985. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). *Biochem.* 24: 3736-3750.

List of publications resulting from DoE 87-ER-60533, 1989 - 1992

Barrantes, R., P. E. Smouse, H. W. Mohrenweiser, H. Gershowitz, J. Azofeifa, T. D. Arias II and J. V. Neel. 1990. Microevolution in Lower Central America: Genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* 46: 63-84.

Chakraborty, R. and J. V. Neel. 1989. Description and validation of a method for simultaneous estimation of effective population size and mutation rate from human population data. *Proc. Natl. Acad. Sci. USA.* 86: 9407-9411.

Hanash, S. M., J. R. Strahler, J. V. Neel, N. Hailat, R. Melhem, D. Keim, X. X. Zhu, D. Wagner, D. A. Gage and J. T. Watson. 1991. Highly resolving two-dimensional gels for protein sequencing. *Proc. Natl. Acad. Sci. USA.* 88: 5709-5713.

Maloney, E. M., R. J. Biggar, J. V. Neel, M. E. Taylor, B. H. Hahn, G. M. Shaw and W. A. Blattner. 1992. Endemic human T cell lymphotropic virus type II infection among isolated Brazilian Amerindians. *J. Infect. Dis.* 166: in press.

Martell, R. E., J. R. Strahler and R. U. Simpson. Identification of lamin B and histones as 1,25-dihydroxyvitamin D<sub>3</sub> regulated nuclear phosphoproteins in HL-60 cells. *J. Biol. Chem.* In press.

Neel, J. V. 1990. Average locus differences in mutability related to protein "class": A hypothesis. *Proc. Natl. Acad. Sci. USA.* 87: 2062-2066.

Neel, J. V. 1990. Unfolding perspectives on the genetic effects of human exposures to radiation. In DNA Damage and Repair in Human Tissues. B. M. Sutherland and A. D. Woodhead, B. M. Sutherland and A. D. Woodheads. Plenum Press, New York. 337-349.

Neel, J. V. 1991. Update on the genetic effects of ionizing radiation. *JAMA.* 266: 698-701.

Neel, J. V. and S. E. Lewis. 1990. The comparative radiation genetics of humans and mice. *Annu. Rev. Genet.* 24: 327-362.

Neel, J. V., W. J. Schull. 1991. The Children of Atomic Bomb Survivors. A Genetic Study. National Academy Press, Washington, D.C. pp. vi and 518.

Neel, J. V., W. J. Schull, A. A. Awa, C. Satoh, H. Kato, M. Otake and Y. Yoshimoto. 1990. The children of parents exposed to atomic bombs: Estimates of the genetic doubling dose of radiation for humans. *Am. J. Hum. Genet.* 46: 1053-1072.

Neel, J. V., W. J. Schull, A. A. Awa, C. Satoh, M. Otake, H. Kato and Y. Yoshimoto. 1989. Implications of the Hiroshima-Nagasaki genetic studies for the estimation of the human "doubling dose" of radiation. *Genome.* 31: 853-859.

Otake, M., W. J. Schull and J. V. Neel. 1990. Congenital malformations, stillbirths, and early mortality among the children of atomic bomb survivors: A reanalysis. *Radiation Research.* 122: 1-11.

Strahler, J. R. and S. M. Hanash. 1991. Immobilized pH gradients: Analytical and preparative use. *Methods: A Companion to Methods in Enzymology.* 3: 109-114.

Torres-Rosado, A., K. S. O'Shea, A. Tsuji, S.-H. Chou and K. Kurachi. Hepsin, A novel membrane protease, is essential for mammalian cell growth. *Science*. Submitted.

Watson, J. T., D. S. Wagner, Y.-S. Chang, J. R. Strahler, S. M. Hanash and D. Gage. 1991. Characterization of the ethyl-triphenylphosphonium derivative of model peptides by fast atom bombardment collisionally-activated dissociation tandem mass spectrometry using B/E linked scans. *Int. J. Mass Spec. Ion Proc.* 111: 191-209.

Yoshimoto, Y., J. V. Neel, W. J. Schull, H. Kato, M. Soda, R. Eto and K. Mabuchi. 1990. Malignant tumors during the first 2 decades of life in the offspring of atomic bomb survivors. *Am. J. Hum. Genet.* 46: 1041-1052.

Yoshimoto, Y., W.J. Schull, H. Kato and J.V. Neel. 1991. Mortality among the offspring (F1) of atomic bomb survivors, 1946-85. RERF TR 1-91. Hiroshima: Radiation Effects Research Foundation. pp. 1-27.

END

DATE  
FILMED

12 / 22 / 92

