Average locus differences in mutability related to protein “class”: A hypothesis

(spontaneous mutation rates/two-dimensional polyacrylamide gel electrophoresis/protein class differences in mutability)

JAMES V. NEEL

Department of Human Genetics, The University of Michigan Medical School, Ann Arbor, MI 48109

Contributed by James V. Neel, December 26, 1989

ABSTRACT Substantially less genetic variation has been recognized in studies of the proteins of nucleated cells by the technique of two-dimensional polyacrylamide gel electrophoresis than has been encountered in studies of serum transport proteins and erythrocyte enzymes with one-dimensional electrophoresis. Technical factors appear to account for only part of the difference. The two remaining (nonexclusive) explanations are more stringent biological selection against variants of these proteins or lower mutation rates at the loci encoding the proteins visualized with two-dimensional electrophoresis. While the former possibility cannot be rigorously excluded, the evidence suggests the latter hypothesis merits serious consideration. Some consequences of this latter suggestion for genetic monitoring and the interpretation of molecular evolution are mentioned.

The advent of two-dimensional polyacrylamide gel electrophoresis of proteins (2-D PAGE), coupled with an appropriate protein stain, made it possible to inspect many previously inaccessible (or poorly accessible) proteins for genetic variation characterized by changes in electrophoretic mobility based on amino acid substitutions altering protein charge. Prior to the availability of this technology, population surveys establishing norms for charge-change genetic variation in proteins had almost exclusively relied on one-dimensional electrophoresis (1-D E) of serum transport proteins and erythrocyte enzymes. In recent years, various investigators, including ourselves, have reported encountering less electrophoretic variation in the proteins visualized with 2-D PAGE than has previously been reported from 1-D E studies. The present contribution will first review the data on variation in the proteins visualized in 2-D PAGE gels, concluding that although technical factors influence the findings, the reduced variation encountered in these proteins is not entirely artifactual. We will then consider two obvious explanations for this reduced variability—namely, (i) stronger stabilizing selection against genetic variants and (ii) lower mutation rates on the part of the genes encoding the proteins that can be scored on these gels. It will be concluded that the latter explanation, which has not previously received consideration, is at least as likely as the former.

Validation of the 2-D PAGE Technique

The ability of the 2-D PAGE system to detect protein variants first recognized with 1-D E/isoelectric focusing was established by O'Farrell (1) in the original description of 2-D PAGE (the L₇–L₁₂ ribosomal proteins of *Escherichia coli*). This ability has been further documented for variants of transferrin (2), L-glycerol-3-phosphate dehydrogenase (3), α₁-antitrypsin (4), and esterase D (5, 6). In our laboratory, we

could readily detect with 2-D gels 8 out of 12 variants first identified with 1-D E, representing allele products of five loci (7). Failure to detect the remaining four was in three instances due to the variant migrating to a position in the gel where resolution of all proteins was very poor; with an altered pI gradient, the variants could be demonstrated. Finally, for a series of 22 transferrin, carbonic anhydrase, triose-phosphate isomerase, and haptoglobin variants identified in 1-D E studies on Japanese samples, 16 were detected with 2-D PAGE, the haptoglobin variants presenting the greatest difficulty in detection (J. Asakawa and J.V.N., unpublished observations). We conclude that the ampholyte-based 2-D PAGE systems in general use in the past, with a pH range of 4.0–7.5, do not have the resolving power of a 1-D E system which has been especially “honed-up” to optimize the detection of variants of a particular protein. To a first approximation, it appears that 2-D PAGE detects about 70–80% of the variants originally identified with 1-D E (but may detect an additional class of variants; see below).

Heterozygosity Indices of the Proteins Detected in 2-D PAGE Preparations

Given these demonstrations, the early efforts to study genetic variation in a number of cell types with 2-D PAGE yielded surprising results. I shall limit this treatment to studies on human material. The most convenient genetic yardstick for measuring variability is the index of heterozygosity (*H*), defined as the average percent frequency of the heterozygous state for a series of loci presumably unselected with reference to their genetic variability. Table 1 summarizes the studies to date on the *H* values revealed by 2-D PAGE preparations of various tissues, cells, and fluids, and contrasts the findings with those from large 1-D E series. The table also lists the proportion of variable loci encountered in each study, but these proportions are much more sample-size dependent than *H* and, since sample size in these studies is relatively small, these proportions will certainly be revised upwards in the future.

It is convenient to divide the studies employing 2-D PAGE into an early group examining fibroblasts, kidney, brain, and hair follicles, and a later group employing the constituents of blood. The early group, with on average in excess of 200 stained protein moieties (spots) scored per gel, yielded values of *H* averaging <<0.5%. The later group of studies, with the exception of those by Goldman *et al.* (19, 20), all emanate from our laboratory or from a collaborating laboratory (The Radiation Effects Research Foundation, Hiroshima) whose personnel acquired the technique in our laboratory and employed nearly identical techniques in Japan. In general, we have scored fewer spots than other groups of investigators, restricting our attention to polypeptides (spots) which are relatively isolated on the gel, in an effort to decrease the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: 2-D PAGE, two-dimensional polyacrylamide gel electrophoresis; 1-D E, one-dimensional electrophoresis.

Table 1. Heterozygosity indices for series of proteins studied by 1-D E and 2-D PAGE

Material	Variable loci/ loci observed		<i>H</i> , %	Ref.
	Nos.	%		
Results with 1-D E				
Enzymes, predominantly erythrocyte				
Caucasian	24/104	23.1	6.3	8
Erythrocyte enzymes and serum proteins*				
Caucasian	13/19	68.4	6.9	Calc. from 9
Am. Black	13/19	68.4	7.3	Calc. from 9
Japanese	18/19	94.7	7.6	Calc. from 10
Results with 2-D PAGE				
Fibroblasts				
Caucasian	<1.0%	?	<0.5	11
Caucasian	≈2/240	0.8	0.6	12
Caucasian	5/500	1.0	<0.5	13
Caucasian	3/659	0.5	<0.01	14
Kidney				
Caucasian	0/83	0.0	0.0	15
Brain				
Caucasian	1/176	0.6	<0.01	16
Hair follicles				
Caucasian	1/289	0.3	<0.01	14
Plasma				
Caucasian	5/20	25.0	6.2	17
Japanese	6/23	26.1	4.8	18
Caucasian	7/40	17.5	5.6	19
Lymphocytes				
Caucasian	19/186	10.2	2.4	20
Caucasian	14/106	13.2	3.1	21
Platelets				
Caucasian	3/33	9.1	2.6	22
Japanese	11/62	17.7	2.4	6
Erythrocytes				
Caucasian	4/46	8.7	3.1	23
Japanese	14/55	25.5	4.0	24

For the 2-D PAGE data the indices are based on actual counts of heterozygotes but for the 1-D E data [except that of Harris (8)] the indices are calculated from allele frequency data by the formula $\Sigma(1 - \Sigma p_i^2)/n$, where p_i = allele frequencies and n = number of loci surveyed.

*These results, based on identical loci in the three ethnic groups, will be presented in detail elsewhere (J.V.N., R. Chakraborty, and P. Smouse, unpublished). The high frequency of variable loci is in part due to large sample size.

likelihood that a one- or two-charge-change variant will be superimposed on another spot and so not easily recognized. These spots were selected before the investigators had detailed familiarity with the pattern and so with no regard to the variability of the particular spots selected for scoring. In addition, each time we ran a test gel, we ran gels from the individual's parents. This practice, not observed by the earlier investigators, greatly facilitated decisions concerning the validity of a putative variant.

The findings with respect to the serum proteins are especially instructive. Whereas only a small fraction of the proteins visualized in other 2-D PAGE preparations have been identified, the proteins of human blood plasma are relatively well known. At last count, the identity of some 46 of the polypeptides visualized on a 2-D PAGE preparation of plasma had been established; many of these have also been examined for genetic variability with 1-D E. The batteries of 20 and 23 spots which we have scored for variability in 2-D

PAGE preparations of serum samples from U.S. Caucasians and Japanese, respectively, include hemopexin, albumin, plasminogen, transferrin, α_1 -antitrypsin, Gc globulin, fibrinogen α chain, acid glycoprotein, apolipoprotein E, apolipoprotein A, fibrinogen γ chain, fibrinogen β chain, and prealbumin. Our observed *H* values for serum proteins were $6.2 \pm 0.7\%$ for Caucasians and $4.8 \pm 0.4\%$ for Japanese (18, 23). Goldman *et al.* (20) reported an *H* for the 40 polypeptides they scored in Caucasian samples of $5.6 \pm 0.7\%$; 7 of 40 loci were characterized by variants. It is noteworthy that most of the variable proteins of plasma on which these *H* values are based had already been identified with 1-D E, with similar heterozygosities; thus, unlike the earlier studies summarized above or the results to follow, the 2-D PAGE studies on plasma are confirmatory in nature rather than breaking new ground.

For the four types of blood constituents listed in Table 1, the unweighted average *H* is 3.6%. This result is heavily influenced by the (extracellular) serum transport proteins, which also enter into the 1-D E *H* values; a cleaner separation between the results of 2-D PAGE with various cell types and 1-D E with erythrocyte enzymes and serum transport proteins will be achieved by eliminating the results with plasma. Now the average *H* in 2-D PAGE preparations of lymphocyte, erythrocyte, and platelet proteins becomes 3.0%; 16.2% of the proteins were characterized by variants. To a first approximation, the unweighted average *H* resulting from all the studies to date on cellular proteins is certainly <2%. Less than 4% of proteins exhibit variants, but this figure should increase substantially as the observations on the various types of preparations are expanded. (It is given in all the arguments to follow that electrophoresis detects only a minority of the genetic variation at a locus.)

Possible Explanations for the Generally Lower Heterozygosity Indices Encountered in the Proteins Visualized by 2-D PAGE

There are three principal, overlapping, explanations to be pursued regarding the apparently lower heterozygosity indices being revealed by 2-D PAGE.

Technical. It seems clear that for any particular protein visualized with 2-D PAGE, variant detection is not yet as effective as with a 1-D E system specifically defined to detect variants of that protein. The data reviewed earlier suggest it would be appropriate to allow for 20–30% underdetection. What, then, is the interpretation of the early studies summarized in Table 1? To some extent, the results may reflect the poorer gel techniques of the earlier days and poor spot selection. In addition, the investigators also did not have recourse to family studies. These investigations need to be repeated, but unfortunately, for such tissues as kidney (15) and brain (16) it is impossible to obtain the family studies so helpful in defining variation. On the other hand, the essential agreement between four different groups of investigators regarding fibroblasts makes the findings regarding this cell type especially plausible. Although we would be ill advised to compare in detail such heterogeneous material, the more recent 2-D results are almost surely yielding “significantly” higher *H* values than the earlier reports, but inasmuch as different cell types are involved, there is no essential conflict between the two sets of findings, and even the more recent studies, with allowance for underdetection, are not revealing the amount of variation identified with 1-D E.

A rigorous comparison of the results from 1-D E with those from 2-D PAGE must be approached with caution for many reasons.

First, the departure from normal of about 20% of the variants detected by ourselves with 2-D PAGE is predominantly in the molecular weight axis, presumably for the most part due to electrophoretically silent amino acid substitutions

which, even under the denaturing conditions of these gels, alter molecular migration. Such variants should only very rarely be detected by 1-D E.

Second, 2-D PAGE should detect some variants of the isoelectric type not detected by nonequilibrium 1-D E (and variants whose frequency is not usually included in the H values quoted for 1-D E). Thus, variation of the type observed with 1-D E is even less common in the proteins visualized with 2-D PAGE than the data would seem to indicate.

A third complication in any comparison of the results of 1-D and 2-D gels is that the number of independent gene products being scored on a 2-D gel is thus less than the total number of spots. In the genetic studies thus far, about 10% of the variants have been recognized as characterized by absolutely correlated variation in two or more spots. Since, however, the same phenomenon presumably exists for non-variable as well as variable polypeptides, the appropriate statistical correction would leave the H unchanged but, since the loci involved are fewer than the polypeptides scored, would increase the error of the estimate.

Fourth, the same variable protein may be scored in several different types of preparations. Thus, the variable platelet polypeptides we have designated PL-3 and PL-5 appear also to be present (and scored) in preparations from lymphocytes and neutrophils (21, 22). Since, however, the same monomorphic protein may also appear in preparations of different cells or tissues, this fact will not artificially inflate H .

Fifth, as a generalization, the proteins visualized with 2-D PAGE are more abundant than the enzymatic proteins toward which most of the 1-D E studies have been directed. It is possible that many of the poorly visualized spots on 2-D PAGE gels are enzymatic proteins that are just as variable as the enzymes studied with 1-D E but are not being scored. Were this true, it would decrease the apparent difference in the results of the two approaches but not negate the principle.

Finally, it is possible that when investigators were developing 1-D E systems primarily to study genetic variation, there was a bias toward publishing on systems for which genetic variation was detected. This notwithstanding, we believe the series reported by Harris (8) to be relatively unbiased, and in our own studies (9, 10), oriented toward the detection of mutation, the criterion for selecting a system for study was simply clarity of resolution of the (stained) protein bands.

These various technicalities render a rigorous comparison of the results of 1-D and 2-D E impossible. In the past, because of the profound biological implications of a difference as great as appeared to be the case, we have been reluctant to accept it as real until technical factors were convincingly excluded (6). We now believe that the cellular proteins visualized with 2-D PAGE exhibit well under half the variation previously encountered with 1-D E, and we turn to a consideration of possible explanations for this finding.

Biological Selection. On the basis of their early studies, McConkey *et al.* (11) (see also refs. 25 and 26) suggested that the paucity of electrophoretic variants in the more abundant and presumably often "structural" proteins visualized by 2-D PAGE resulted from greater selective constraints on these proteins than on the proteins previously visualized with 1-D E. These and other authors have emphasized the many differences in physical characteristics and function between the enzyme and serum proteins visualized in 1-D E studies and the more abundant proteins visualized with 2-D PAGE. In view of the low average H values encountered for these latter proteins, individuals homozygous for an electromorphic variant of the proteins visualized in 2-D gels will in general be extremely rare, and the postulated negative selection must be directed against the heterozygous individual. Under these circumstances, the deleterious effects of these

variants should follow the pattern of classical dominant inheritance. This would be in strong contrast to the situation with respect to electrophoretic variants of the erythrocyte enzymes and serum proteins, where with respect to *overt* phenotypic expression, the variants in their expression very predominantly follow the pattern of classical recessive inheritance, with even homozygotes for the variant usually exhibiting no obvious handicap.

The haploid number of different genes in humans is commonly set at approximately 50,000. Let us assume that half of these (25,000) code for the kinds of less variable proteins visualized with 2-D PAGE, proteins with respect to which charged amino acid substitutions lead, as suggested by McConkey *et al.* (11), to dominantly inherited disease. The number of dominantly inherited disorders listed in McKusick's compendium (27) which might be related to structural proteins appears to be less than 1000. There is thus a relative paucity of dominantly inherited traits under the selection hypothesis, a paucity which in principle can be met by the postulate that these mutations are early prenatal lethals. A simple calculation illustrates the consequences of this postulate. From studies of serum proteins and erythrocyte enzymes, we have suggested that the rate with which mutation results in amino acid substitutions in proteins is approximately 1×10^{-5} per locus per generation (28). The probability that a zygote will *not* have received one or more such mutations will be $[1 - (1 \times 10^{-5})]^{25000} = 0.778$. If these mutations tend to function as unconditional dominant lethals, this implies the early loss of as many as 22.2% of all zygotes from this single type of mutation. This seems an excessive loss to be attributed to this one class of mutations, especially since these changes constitute only a portion of all the possible (many presumably equally deleterious) mutations at these loci. Although the calculation is only approximate, it illustrates clearly the type of problem raised by the assumption of dominant lethality of electromorphs.

With respect to the actual evidence concerning the viability of heterozygotes for mutations involving structural proteins, we consider the still-unfolding story of type I collagen, perhaps the quintessential example of a ubiquitous structural protein studied for genetic variation. The molecule is a triple helix of two identical $\alpha 1(I)$ chains and one similar $\alpha 2(I)$ chain, each about 1100 amino acids long, the two chains encoded by different genes. The molecule has a complex structure, and any alteration of the predominant glycine moiety in the chains—or a small deletion or insertion—appears to have the potential to alter chain configuration and hence its functioning as a member of a trimer. Because of this intricate structure, as well as the large amount of collagen present, the molecule may be an atypical example of a structural protein. The consequences of genetic changes in the molecule have been particularly well studied in the complex of dominantly inherited diseases known as osteogenesis imperfecta (29–31). Many of the changes which have been detected in the collagen of individuals with these diseases are single amino acid substitutions. Paradoxically, total absence of the $\alpha 2$ chain seems better tolerated than some glycine substitutions in an otherwise normal $\alpha 2$ chain. While the possibility of amino acid substitutions which are prenatal lethals cannot be excluded, it is clear that a range of glycine substitutions resulting in disease is compatible with survival, sometimes to an advanced age, and, sometimes, reproduction, resulting in the pattern of dominant inheritance. Unfortunately, the kinds of surveys that would detect the frequency of the presumably less harmful substitutions in the non-glycine moieties of the chains have not yet been carried out, so that amino acid substitutions in the molecule compatible with extended postnatal survival may be even more frequent than appears to be the case at present. Additional molecular defects in collagen are being identified in studies of Ehlers–Danlos syndrome,

Marfan syndrome, and epidermolysis bullosa (discussion in refs. 32 and 33).

Although many so-called structural polypeptides must enter into complex relationships with other polypeptides, the polypeptide-polypeptide interactions of the collagen $\alpha 1$ and $\alpha 2$ chains or of the fibrin chains are probably especially complex. Where there is no such interaction, and the polypeptide functions as a monomer, 50% of the molecules in a heterozygote will be normal; there is little precedent for an amino acid substitution in the remaining 50% regularly conferring the high selective disadvantage required to account for a paucity of variants. Therefore, although the possibility that the stabilizing selection operating on these molecules as a class is stronger than the selection on enzymes and transport proteins, extreme levels of selection, including dominant lethality of newly arisen mutations, seems unlikely to provide the total explanation for the paucity of variants in 2-D PAGE preparations.

Relatively Low Mutation Rates. I suggest that in view of the problems explaining the low frequency of variants by the selection hypothesis, we must seriously begin to consider an alternative, that the findings reflect, to an extent yet to be determined, relatively low average mutation rates at the loci encoding the proteins scored on these 2-D PAGE preparations. That mutation rates were under genetic control has long been suggested by the apparent similarity in spontaneous rates per generation in such diverse organisms as *Drosophila*, mice, and humans, despite the great differences in life span, mean body temperature, and average number of cell divisions in the germ cell line (34). On a more specific note, and with respect to higher eukaryotes, for five specific coat-color loci in the mouse, Schlager and Dickie (35) list rates varying from 3.9×10^{-6} to 29.7×10^{-6} per generation. For the seven loci employed in the well-known Russell test system, the spontaneous rate in males varied on the basis of 195,289 locus tests from 0.0 to 2.0×10^{-5} (36), but the differences were not statistically significant. The range in the x-ray-induced rate (spermatogonia) for these seven loci is more than 30-fold (37). Of particular interest in light of the findings to be discussed below, the three loci characterized by the higher spontaneous rates were also among the four loci characterized by higher induced rates (data in ref. 38). Finally, among 19 loci encoding serum proteins and erythrocyte enzymes, with respect to which at least one variant was encountered in surveys of Amerindian tribes, the application of equilibrium theory resulted in estimates of locus mutation rates for electromorphs varying from 0.24×10^{-5} to 3.48×10^{-5} per locus per generation, with an average across all loci studied, variable or not, of 1.1×10^{-5} per locus per generation (39). The basis for these locus differences in higher eukaryotes is essentially unknown. There is a (variable) positive correlation between molecular size and mutability as measured by direct studies (23) and inferred from locus heterozygosity studies (references in ref. 40), but this is not adequate to explain by itself locus differences in mutability. If, however, individual loci can differ to this extent, it would seem not unreasonable that "classes" of loci could exhibit differences in average mutation rate, although with much heterogeneity within the somewhat arbitrarily defined classes.

A recent experimental finding in our laboratory has rather forcibly directed attention to the possibility that the explanation for the lower frequency of variants in 2-D gels rests with lower average mutation rates on the part of the loci encoding the proteins envisioned with 2-D PAGE. Cultured human lymphoblastoid cells were heavily treated with ethylnitrosourea (41, 42). A total of 267 polypeptides encoded by 263 loci were scored on 2-D PAGE preparations for mutation in a series of 1143 lymphoblastoid clones. Sixty-five mutations characterized by the appearance of an additional spot were

identified. The mutants involved 49 proteins: three mutants were observed for each of 5 proteins, two mutants for each of 6, and one mutant for each of 38. No mutants were observed for any of the remaining 214 proteins. Thus, there were 16 "repeats." The random probability of this occurrence is <0.0005 . This nonrandomness was to some extent correlated with molecular weight, the proteins on the upper half of the gel showing a significantly higher rate than the proteins on the lower half. Even more interesting genetically, eight of these mutants involved the 10 proteins being scored in this experiment which were known from our own previously quoted studies in lymphocytes to be associated with genetic polymorphisms. The mutation rate at these loci was 3.6 times greater than at the monomorphic loci, an observation whose probability was <0.004 . An obvious inference, given the correlation between spontaneous and induced rates documented for the mouse (see above), is that the array of genetic variability detected in 2-D gels is to a considerable extent driven by mutation pressure, rather than by selection operating on equimutable loci, with the further inference that the paucity of genetic variation visualized in 2-D gels results from relatively low mutation rates (however achieved) at the loci encoding these proteins rather than intense selection against variants. Two caveats are necessary at present: (i) these are results from cell culture, not intact individuals, and (ii) the functions (enzymatic or structural) of the more mutable proteins are still unknown. The suggestion of average differences in mutability between the serum proteins and the erythrocyte enzyme proteins, on the one hand, and the proteins visualized on 2-D gels, on the other hand, does not preclude a considerable locus range in mutability within each class.

Discussion

The proper explanation of the cause(s) of the lesser genetic variation being recognized in 2-D PAGE gels is a major issue in contemporary biomedicine. The demonstration with 1-D E beginning in the 1950s of the extent of the previously concealed biochemical variability in enzymatic and transport proteins was surely one of the most significant advances in population genetics in this century; the further demonstration that this variability is significantly less in another equally important group of proteins would be a similar major development. If the explanation is primarily technical, this failure to detect variants for the present seriously limits the usefulness of 2-D PAGE. If it could be shown that this finding is primarily the result of relatively heavy biological selection against variants, this would constitute some of the most important evidence in this decade for the role of determinism in the array of genetic variation encountered in populations. If it is largely or entirely the result of lower mutation rates, this is a major contribution to our concept of the stability of the human genome. In this connection, we note that, given the apparently high efficiency of the DNA repair mechanisms, it requires only a small improvement in DNA repair with respect to any gene to have a major impact on the mutation rate of that gene. The final interpretation of the findings summarized in this paper will probably require a blend of these explanations. With continuing improvement in the 2-D PAGE technology, such as the newly introduced immobilized gradient gels (43-45), the question of underdetection of variants is susceptible to attack in the immediate future. Apportioning the relative roles of rigorous selection and relatively low mutation rates will be much more difficult.

Space permits mentioning only three of the many specific implications of this hypothesis, two theoretical, one practical. At this juncture I note that these consequences apply only to exons, on whose products the current observations are based.

(i) *Selective constraints on molecules.* Certain molecules have been evolutionarily much more conservative than oth-

ers [e.g., the histones, ubiquitin, tubulin, and actin (conserved) vs. fibronectin, the interferons, and relaxin (nonconserved)]. Such conservatism is customarily attributed to stringent selection. If the conservatism is predominantly the result of a down-regulated mutation rate for genes whose products operate within narrow tolerance limits, the cost of stabilizing or purifying selection—a major factor in evolutionary thinking—becomes substantially less. Likewise, one is led to consider the question of the extent to which, and the mechanism by which, a long-lived species with a relatively low reproductive capacity, such as humans, has in the course of evolution down-regulated mutation rates.

(ii) *Molecular clocks*. The rate in evolutionary time with which amino acid substitutions occur in a protein, or nucleotide substitutions occur in a gene, constitutes a molecular clock. The many discussions comparing such clocks within and across protein lineages usually assume constancy of mutation rate, either per generation or per unit time (summary in refs. 40 and 46). If in fact mutation rates differ by a factor of 10 across loci, and perhaps by a lesser amount for a given locus across lineages, and also by protein "class," the combination of variable mutation rates plus variable selective pressures through time renders the interpretation of either regularities or irregularities in the clock quite problematic.

(iii) *Genetic monitoring*. Effects to monitor human populations for genetic damage which utilize a battery of specific genes, or extrapolations to humans from experimental data which rely on a specific array of genes, now must meet the challenge of the applicability of these results to the genome as a whole. The necessity of basing studies on an array of genes representing the true spectrum of spontaneous and induced mutability will exist whether the end-point is a gross phenotype, a protein, or a segment of DNA. Unfortunately, with reference to the use of gross phenotypes in future studies, one cannot obtain mutation rates for loci whose existence has not been identified by at least one mutation. This source of bias shifts the emphasis in the studies of the future to proteins and DNA segments. The best empirical guide in selecting a battery of either of these latter for study would appear to be that the battery reflect the range of locus variability encountered in a natural (i.e., nonexperimental) population. The requirement necessitates that a large body of normative data be available before an appropriate battery of mutational indicators can be selected.

The financial support of the U.S. Public Health Service (CA-26803) and the Department of Energy (ER-60089) in the original investigations upon which this paper is based is gratefully acknowledged.

- O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021.
- Anderson, L. & Anderson, N. G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5421–5425.
- Lee, C.-H., Niesel, D. & Bewley, G. C. (1980) *Biochem. Genet.* **18**, 1003–1016.
- Brown, W. T. (1982) *Am. J. Hum. Genet.* **34**, 195–208.
- Kondo, I., Yamamoto, T. & Yamakawa, K. (1984) *Hum. Genet.* **66**, 248–251.
- Asakawa, J., Neel, J. V., Takahashi, N., Satoh, C., Kaneoka, S., Nishikori, E. & Fujita, M. (1988) *Hum. Genet.* **78**, 1–8.
- Wanner, L. A., Neel, J. V. & Meisler, M. H. (1982) *Am. J. Hum. Genet.* **34**, 209–215.
- Harris, H. (1980) *The Principles of Human Biochemical Genetics* (Elsevier/North-Holland Biomed., Amsterdam), 3rd Ed.
- Mohrenweiser, H. W., Wurzing, K. H. & Neel, J. V. (1987) *Ann. Hum. Genet.* **51**, 303–316.
- Neel, J. V., Satoh, C., Smouse, P., Asakawa, J.-I., Takahashi, N., Goriki, K., Fujita, M., Kageoka, T. & Hazama, R. (1988) *Am. J. Hum. Genet.* **43**, 870–893.
- McConkey, E. H., Taylor, B. J. & Phan, D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6500–6504.
- Walton, K. E., Styer, D. & Gruenstein, E. I. (1979) *J. Biol. Chem.* **254**, 7951–7960.
- Giometti, C. S. & Anderson, N. L. (1981) *J. Biol. Chem.* **256**, 11840–11846.
- Klose, J., Willers, I., Singh, S. & Goedde, H. W. (1983) *Hum. Genet.* **63**, 262–267.
- Smith, C. R., Racine, R. R. & Langley, C. H. (1981) *Genetics* **96**, 967–974.
- Comings, D. E. (1982) *Clin. Chem.* **28**, 798–804.
- Rosenblum, B. B., Neel, J. & Hanash, S. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5002–5006.
- Asakawa, J., Takahashi, N., Rosenblum, B. B. & Neel, J. V. (1985) *Hum. Genet.* **70**, 222–230.
- Goldman, D., Goldin, L. R., Rathnagiri, P., O'Brien, S. J., Egeland, J. A. & Merrill, C. R. (1985) *Am. J. Hum. Genet.* **37**, 898–911.
- Goldman, D. & Merrill, C. R. (1983) *Am. J. Hum. Genet.* **35**, 827–837.
- Hanash, S. M., Baier, L. J., Welch, D. & Galteau, M. (1986) *Am. J. Hum. Genet.* **39**, 317–328.
- Hanash, S. M., Rosenblum, B. B., Neel, J. V., Baier, L. J. & Markel, D. (1986) *Am. J. Hum. Genet.* **38**, 352–360.
- Rosenblum, B. B., Neel, J. V., Hanash, S. M., Joseph, J. L. & Yew, N. (1984) *Am. J. Hum. Genet.* **36**, 601–612.
- Takahashi, N., Neel, J. V., Nagahata-Shimoichi, Y., Asakawa, J., Tanaka, Y. & Satoh, C. (1986) *Ann. Hum. Genet.* **50**, 313–325.
- Edwards, Y. & Hopkinson, D. A. (1980) *Nature (London)* **284**, 511–512.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge).
- McKusick, V. A. (1983) *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore), 8th Ed.
- Neel, J. V., Satoh, C., Goriki, K., Fujita, M., Takahashi, N., Asakawa, J. & Hazama, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 389–393.
- Byers, P. H. & Bonadio, J. F. (1984) in *Birth Defects: Original Article Series*, ed. Bergsma, D. (Liss, New York), Vol. 20, No. 3, pp. 65–77.
- Byers, P. H., Bonadio, J. F. & Steinmann, B. (1984) *Am. J. Med. Genet.* **17**, 429–435.
- Byers, P. H. & Bonadio, J. (1989) in *Collagen: Biochemistry, Biotechnology and Molecular Biology*, eds. Olsen, B. R. & Nimui, E. (CRC, Boca Raton, FL), Vol. 4, pp. 125–140.
- Tsipouras, P. & Ramirez, F. (1987) *J. Med. Genet.* **24**, 2–8.
- Weil, D., Bernard, M., Combates, N., Wirtz, M. K., Hollister, D. W., Steinmann, B. & Ramirez, F. (1988) *J. Biol. Chem.* **263**, 8561–8564.
- Neel, J. V. (1983) *J. Hered.* **74**, 2–15.
- Schlager, G. & Dickie, M. M. (1967) *Genetics* **57**, 319–330.
- Searle, A. G. (1974) *Adv. Radiat. Biol.* **4**, 131–207.
- Russell, W. L., Russell, L. B. & Oakberg, E. F. (1958) in *Radiation Biology and Medicine*, ed. Claus, W. O. (Addison-Wesley, Reading, MA), pp. 189–205.
- Russell, W. L. (1964) *Genetics Today*, ed. Geerts, S. J. (Pergamon, Oxford), Vol. 2, pp. 259–264.
- Chakraborty, R. & Neel, J. V. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9407–9411.
- Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
- Hanash, S. M., Boehnke, M., Chu, E. H. Y., Neel, J. V. & Kuick, R. D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 165–169.
- Chu, E. H. Y., Boehnke, M., Hanash, S. M., Kuick, R. D., Lamb, B. J., Neel, J. V., Niezgoda, W., Pivrotto, S. & Sundling, G. (1988) *Genetics* **119**, 693–703.
- Righetti, P. G. (1989) in *Protein Structure: A Practical Approach*, ed. Creighton, T. E. (IRL, Oxford), pp. 23–63.
- Strahler, J. R., Kuick, R. & Hanash, S. M. (1989) in *Protein Structure: A Practical Approach*, ed. Creighton, T. E. (IRL, Oxford), pp. 65–92.
- Strahler, J. R., Kuick, R., Pivrotto, S., Somerlot, L., Chu, E. H. Y. & Hanash, S. M. (1988) in *Two-Dimensional Electrophoresis*, eds. Ender, T. & Hanash, S. M. (VCH, Weinheim), pp. 256–262.
- Li, W.-H., Luo, C.-C. & Wu, C.-I. (1985) in *Molecular Evolutionary Genetics*, ed. MacIntyre, R. J. (Plenum, New York), pp. 1–94.