Lesion Spectra: Radiation Signatures and Biological Gateways

K. Rupnik,*,† L. Klasinc,† M. Varma,‡ J. Battista,§ and S. P. McGlynn*,†

Departments of Chemistry and Microbiology, Louisiana State University, Baton Rouge, Louisiana 70803, and U.S. Department of Energy, Washington, D.C. 20585

Received November 10, 1993®

We describe an integrative approach to the modeling of biophysical radiation effects. The model takes aim at practical applications of the knowledge provided by molecular studies of radiation–matter interactions in DNA. The central proposition is the ideat that the distribution of molecular lesions (i.e., a molecular lesion spectrum, MLS) generated in DNA by exposure to a particular radiation is a characteristic of that causal radiation (i.e., is a radiation signature, RS). We have found that adaptive neural networks (ANN's) provide an efficient way to validate that proposition and that ANN's are also likely to be invaluable in any attempt to correlate cancers with radiation types (i.e., with RS's), to use RS's for evaluating individual carcinogenic susceptibilities, and to develop a low-dose personalized monitoring capability. Although efforts to identify products of radiation that are specific to radiation type and to link those with biological responses are almost a century old, the RS concept has provided the first quantitative confirmation of such causal relations. That is, RS's and radiation markers have been identified for various types of radiation, electromagnetic (EM) and particulate, and these signatures and markers may constitute a new way for fast radiation exposure estimates, risk assessment, and cumulative low-dose evaluation. In this work, while we will present a short review of the concepts and methods related to both RS's and markers, almost the entire effort will relate to the modeling and interpretation of RS's using ANN processing.

1. INTRODUCTION

Much biophysical modeling takes aim at the relations between physical events such as ionization, excitation, etc., and carcinogenesis. All types of radiation are of concern because each is potentially carcinogenic. This concern has grown as exposure to natural and artificial sources of radiation has increased. The sun, for example, was thought to be relatively harmless; for most people, a sun shield and the relatively large thresholds of individual susceptibility were thought to be sufficient protection against everyday visible (VIS) and ultraviolet (UV) radiation. That, however, turns out to be inadequate for those individuals with very susceptible skin or eyes (e.g., persons of Celtic descent), for those who are constantly exposed to solar radiation, or for those who suffer from a variety of diseases (e.g., xeroderma pigmentosum). Regardless of individual susceptibilities, the modern world is now replete with radiation sources (e.g., nuclear arms, medical sources, nuclear power, oil production, aircraft or space travel, etc.), which make radiation risk a common concern. While this concern has been around for more than 50 years, it has produced very little success. True, we have learned that different radiation types, from electromagnetic (EM) waves to elementary particles, are ubiquitous and that they pose a very real threat, and true, we find many communities and working environments in which there is realization of the urgent need for radiation risk assessment technology. Unfortunately, a good risk assessment technology does not exist—a consequence, we believe, of an inability to relate cause (i.e., radiation) to effect (e.g., cancer) in any scientific way.

The situation for other nonspontaneous cancers may well be somewhat better: some partial biological models of development exist for cancers induced by viruses or chemicals. While these models are multistage, and the relationships between one stage and another may not be understood, models do at least exist. Not so, however, with radiation-induced cancers. And the situation gets worse when one considers the more delicate problem of low-dose induced cancer: low-dose exposures afflict by far the most people, and the complexity of the cancers they induce seems to be a consequence of a large number of parallel physico/cheimical and biological processes that take place during or after radiation insults.

We have recently initiated a research effort in biophysical modeling that is aimed at fast radiation exposure assessment and cumulative low-dose evaluation using radiation signatures (RS's) and radiation markers (RM's).[1-4] We have proposed and, in part, justified that the distribution of molecular lesions (i.e., a molecular lesion spectrum, MLS) generated in DNA by exposure to a particular radiation is a characteristic of that causal radiation (i.e., is a radiation signature, RS). The use of neural networks for pattern recognition, the concept of an MLS as an RS, and the possible correlation of carcinogenesis (and other critical effects) to MLS's were first proposed at the 24th U.S. DOE Contractors Meeting on Radiation Signatures (New York, Columbia University, April 1991).[1] It may be a bit presumptuous to suppose that the concept of an MLS as an RS is totally unique to us. After all, efforts have been underway for decades to identify the products specific to a given radiation type, and to link these with biological responses and health outcomes. Nonetheless, no general ansatz or paradigm seems to have emerged. Our approach is also unique in that adaptive neural networks (ANN's) are used to analyze data from a wide range of dosimetric, chemical/biochemical, and radiobiological studies.

Any attempt to formulate mechanistic models of relationships between signatures and carcinogenesis must await the completion of investigations for a variety of radiation types. In specific, MLS's must be constructed for both *in vivo* and *in vitro* systems and for a variety of radiation types. Quicker assays for a variety of lesions (i.e., **breaks, adducts, multiple locally damaged sites**, etc.) must be developed. And, ANN

† Department of Chemistry, Louisiana State University.
‡ U.S. Department of Energy.
§ Department of Microbiology, Louisiana State University.
® Abstract published in *Advance ACS Abstracts,* September 1, 1994.

0095-2338/94/1634-1054$04.50/0

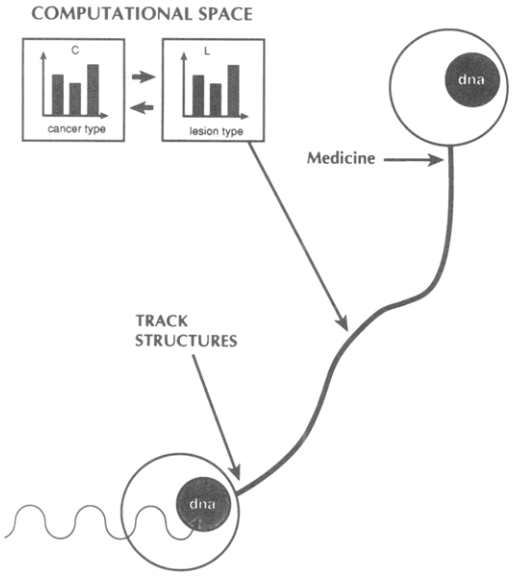

Figure 1. Postradiative spatiotemporal development of a cell from "normal" to "abnormal". The radiative insult occurs at the lower left. Clinical effects occur at the upper right. The MLS's introduced in this work are located about midway along the curve. In the same way as MLS's are mapped onto the radiation types, there exists the possibility that clinical effects might be mapped onto MLS's.

processing of lesion distributions (i.e., patterns) to establish the distinctiveness of lesion spectra must be investigated in depth.

This present work is totally concerned with this last point, the use of ANN's for RS recognition.

2. RADIATION SIGNATURE PARADIGM

We now present the RS paradigm.

A number of mechanisms of radiation–matter interaction have been investigated in the past 50 years. Much attention has been paid to initial track or ion distribution structures. These structures are functions of radiation quality. They can be viewed as "physical signatures" of the radiation. The longest persistence time of ion distributions is certainly less than 10^{-7} s, at which time all of the physical processing of the radiative insult may be assumed to be complete. The "chemical" regime, during which much of the initial chemistry is completed, certainly extends into the domain of seconds. The "biological" regime, where biological response mechanisms dominate, continues thereafter. The "clinical" regime, which is the last phase of development of a health outcome, may last 20 or more years.

The best way to understand the RS paradigm is by inspection of Figure 1, which shows a spatiotemporal processing path from a normal just-insulted cell to a fully transformed carcinogenic (or mutagenic, etc.) cell. A great deal of radiation physics has been concentrated in the short time region of this pathway, track structures to $\sim 10^{-9}$ s, and ion distributions to $\sim 10^{-7}$ s. These tracks/distributions are too far removed, in a spatiotemporal sense, from the clinical manifestation of cancer to be of much use. Clinical manifestations of cancer may take more than 20 years to develop, which places an RS based on malignancy (i.e., a clinical signature) at $\sim 10^{9}$ s. Thus, it becomes necessary to follow the radiation track through a whole host of chemical, biochemical, biological repair, and emergent cellular happenings before its manifestation as cancer. Thus, the time scale alone, which covers an 18 decadic range, is too long to permit easy correlation of the two end points of the path.

We suggest that the most appropriate place to seek "signature" is at the junction of the chemical and biological

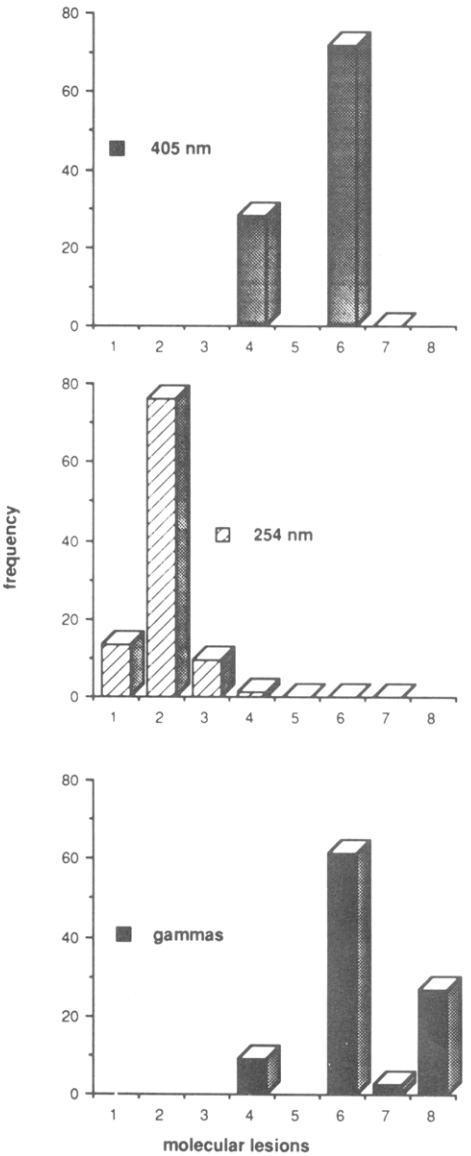

Figure 2. The histogram of molecular lesions, or MLS, for the eight-lesion set for 405-nm, 254-nm, and γ-ray radiations. The MLS is a plot of lesion frequency versus discrete lesion type. The abbreviations on the abscissa correspond to lesion names given in the text: **1 = cytidine hydrate, 2 = thymine dimer, 3 = pyrimidine adduct, 4 = DNA-protein cross-link, 5 = DNA–DNA cross-link, 6 = single strand break, 7 = double strand break, and 8 = locally multiply damaged sites.**

regimes. Such a point lies approximately in the middle of the spatiotemporal path. If such a signature were to exist, it should also serve as a gateway point for the bioprocessing which ultimately ends in malignancy. That is, the correlation of the two ends of the spatiotemporal path might be facilitated by relating both ends of the path to the middle. It is our contention that such a signature exists and that it is the spectrum (or histogram) of DNA molecular lesions (Figure 2).

This choice is also practical: it could provide fast risk assessments. Clinical methods require decades, as witnessed with the continuing confusion about the Hiroshima mutation studies. Since molecular lesions are quantifiable at times of the order of seconds after exposure, MLS's and RS's can be used to monitor both exposures and the efficacies of intervention treatments. MLS's may also be used for radiation exposure estimates in situations where target tissue sampling is either impractical or dangerous: urine, for example, could be used to chronicle exposure. Nevertheless, independent biological dosimeters will always be required, and *Deinococcus radiodurans* would seem particularly apt for this purpose.[5]

In sum, the RS paradigm requires that the MLS be specific to the causal radiation; it contends that such an RS is a particularly pertinent biological gateway and radiation risk assessment device; and it holds that the correlation of the two ends of the spatiotemporal path to the middle, where RS's are positioned, is much easier than any attempt to relate one end to the other.

2.1. Feature Extraction. We now introduce an adaptive parallel distributing processing (APDP) approach to implementation of the RS paradigm.

One wishes to correlate a set of MLS's with both causative radiation and health outcomes. The first question concerns the description of differences between MLS's. We find that artificial intelligence (AI) methods for feature extraction, specifically the ANN apparatus, are excellent tools for such feature extraction.

The concept of a signature (i.e., feature recognition) implies the existence of a set of object characteristics that discriminate between different objects and/or classes of objects. A signature must possess the characteristics of repeatability (i.e., confidence) and ease of recognition (i.e., ease of discrimination from threshold "noise"). Furthermore, since the characteristics of a signature need not be explicitly circumscribed by rules, some element of adaptability in the models used for feature extraction appears to be mandatory. For that reason, the methods chosen for extraction of RS's from diverse data sets are adaptive lesion models which originate in APDP methods and which are related to ANN and genetic algorithms (GA). Using a mathematical/statisical analogy, each MLS can be viewed as a "vector" in a space spanned by the totality of MLS's, and the distinctiveness of patterns may be quantified as "degrees of orthogonality" of these representative RS vectors.

In an adaptive network, the concept of degree of orthogonality refers to the ability to train the network on an orthogonal set of vectors. This orthogonal set of n vectors represents different classes, or features, or patterns in the same way as a set of n basis vectors represents orthogonal directions in an n-dimensional space. The final output of an adaptive network is distributed through a set of output layer processing elements (PE's) which, since they correspond to the orthogonal vectors, provide ready evaluation of the degree of closeness of the input vectors to the trained, orthogonal set. The criteria for degree of orthogonalization or discrimination in an adaptive network can be established quite precisely.[6-14]

Feature recognition techniques become necessary when one deals with data bases that are less than optimal, when one queries the minimum number of lesions adequate for a signature, or when one wishes to pursue the more tenuous and presumptive connection of a radiation and its clinical outcome. At this point, it is pertinent to note that this connection may not be all that tenuous. For example, the recent work of Vähäkangas et al.[12] indicates an ability to connect clinical results with MLS's (i.e., to traverse the upper half of the spatiotemporal path). In specific, these authors report that they were able to distinguish lung cancer of tobacco origin from that attributable to tobacco complemented by radon inhalation using molecular lesions in the p53 tumor suppressor gene. However, although important as illustration, the history of similar attempts indicates that the above example is more exception than rule. Therefore, we advocate a comprehensive ANN approach.

2.2. Some Definitions. We have already introduced[4] definitions for the terms *Molecular Lesion*, *Sub-lesion*, *Molecular Lesion Spectra*, *Radiation Signature*, and *Radiation Marker*. A few more definitions are required here:

Radiation Type: a radiation specified by a set of well-defined physical properties, such as wavelength, frequency, intensity, fluence, etc.

Radiation Class: radiation that produces some well-defined physico/chemical processes, ones that can be used as radiation descriptors, in a material (say DNA) (the radiation, for example, could be said to be ionizing, nonionizing, etc.).

Radiation Signature Class: any class of RS that can be associated with a specific physico/chemical process or a combination of such processes or that can be quantitatively distinguished using ANN processing.

Computational Space: an operational space for ANN processing (for example, a preprocessed set of MLS's).

Risk Category: A specific biological or health outcome associated with risk. It includes cell death, human cancer, or mutation types. Risk categories can be obtained from clinical and biological studies (for example, the conditions which produce $1/e$ or 37% lethalities). An RS may be associated, using ANN processing, with a specific risk category, or a combination of risk categories.

3. ADAPTIVE NEURAL NETWORKS

ANN's are among the most developed APDP systems. Only 10 years ago, investigations such as those described here would be difficult if not impossible. We define "An ANN is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the adaptive processing performed at PE's or nodes."

The common denominator of all ANN's is the ability to handle nonlinear parallel distributed processing. Neural networks take advantage of large numbers of parallel PE's either as objects in the software or as real electronic or bioprocessing systems. The working of a nonlinear feedforward back-propagation neural network (BPN) is presented in Figure 3. A BPN with enough PE's in the hidden layer(s) can learn to recognize any functional relation between input layer PE's and output layer PE's, regardless of its complexity. However, such an ability (i.e., the successful operation of an ANN) depends heavily on adequate training.

While ANN's have been inspired largely by the architecture and topology of nervous systems, their structures are heavily influenced by parallel processing models and by physical models of large complex systems. Most performance estimates of an ANN are transplants from statistics (regression). Finally, a large variety of applications (e.g., brain and sensor studies, electronic signal analysis, sonar, seismic, and oil and gas studies, chemical processing, adaptive control mechanisms, feature extraction and imaging, explosives detection, jet engine diagnostics, banking and financing, artificial organs, etc.) have imposed sets of specific demands which, in turn, have influenced the architectures, learning mechanisms, and performance estimates of ANN's.

In the framework of the RS paradigm, the ANN's of interest are those aimed at (1) feature extraction or pattern recognition and (2) predictions. A third desideratum, optimization, is also of interest; however, in order to retain focus, we shall not discuss it here. We emphasize BPN's in this work. To be sure, we have also investigated other network types and architectures. Unfortunately, no simple recipe for comparative evaluation of them exists. There is, indeed, only one criterion: is there an ANN that performs better than a first principle approach (if available) that can give better results than a statistical model (if available) or a rule-oriented AI method (if available), or that yields better results than some other ANN? And, there is only one pertinent question: which

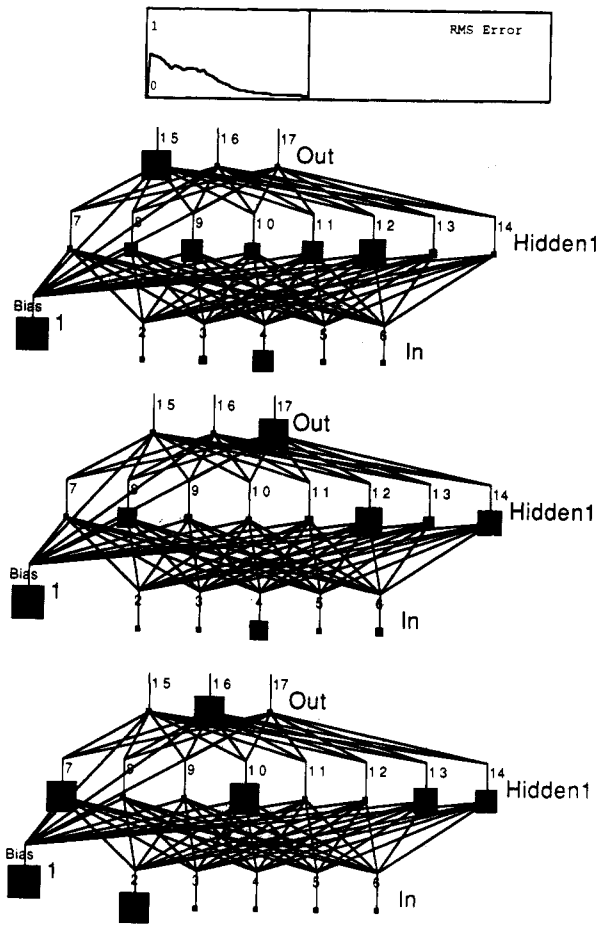

Figure 3. A test of the trained BPN for 405-nm (upper), 254-nm (middle), and γ-ray (lower) radiations using MLS's. The network is discussed in the text. Input occurs in the bottom layer (five PE's, each corresponding to the frequency of a specific lesion). The MLS's (each consisting of five lesion frequencies) are the input to the net (PE's no. 2–6). The output neurons (PE's no. 15–17) are identified as the three components of each of the three orthogonal vectors for the three EM radiation types. The RMS error of the output layer during the training process is shown at the top of the diagram, time being on the horizontal axis.

method, relative to the investigative goals, has the best set of properties? The answer to this question is determined by network performance.

Comparison of a BPN with the "flow diagram" of operations for a statistical method (e.g., a regression analysis) emphasizes the "topological" difference between the two "architectures": when the hidden layers are removed, the remaining "substructure" of that BPN is easily transformed to a statistical regression program. The manner in which data are processed is also different: statistical regression algorithms fit the data to a preconceived model, whereas neural networks, with no preconceived model, use adapting or learning procedures based on persistent presentation of the data set (i.e., the training set) to the network. Since an ANN assumes no a priori model and no corresponding algorithms, the operation of an ANN consists of two phases: (1) training (or adapting) and (2) recall of the trained network (or testing). In the first phase, an ANN reorganizes or adapts according to the input data and the desired output. In the second phase, an ANN merely generates a calculated output, no further adaptation of the network being allowed.

Performance evaluations of an ANN may be made using standard deviations, RMS's, Pearson's R coefficients, rates of incorrect predictions, etc. It is expected that ANN's will be more successful for large complex systems because of the nonlinear parallel distributed learning capabilities which simultaneously activate different connectivities and, consequently, involve parts of the computational space that are eliminated by the rigidity of conventional statistical techniques. The statistical regression approach calculates or "trains" fast, often in one pass. It has fewer calculation procedures and equations than an ANN but usually finds approximate, rigid models with a minimal number of variables. It interprets the model in terms only of these variables. When presented with other situations, it often fails. It is more difficult to train a BPN. Training of a BPN often requires more than 10 000 presentations of the training data set. However, BPN's take all variables into account, not merely those that dominate a first fitting. Consequently, the explanation of the BPN also depends on all variables. The number of variables is proportional to the complexity of the problem. This property of a BPN is of paramount importance when dealing with RS's.

No "first principle" model has been proposed for RS's. Consequently the use of statistical models for prediction and classification is inappropriate. A mechanistic model for the RS paradigm could be constructed from the set of physicochemical equations, were they known, that describe the temporal and spatial behavior of the molecular systems. However, such a standard modeling program would be cost prohibitive and surely unsuccessful. The use of a BPN avoids these difficulties. Since MLS's contain information about physico/chemical processes and since ANN's can extract that information, it is not inconceivable that mechanistic models will evolve from RS studies. Indeed, this exploratory part of BPN processing is a major goal of our work.

3.1. Back-Propagation Networks for Radiation Signatures. We have found BPN's to be most advantageous for RS studies. The advantages are at least 2-fold: (1) they have a well-defined mathematical basis, and (2) they have been successful in feature recognition applications.

The goal of the adaptive system of BPN's is minimization of error, that is, minimization of the difference between desired (D_i) and predicted (x_i) values at each output PE_i. There are two steps in the BPN processing: (1) feedforward, where input signals are processed to the output layer, and (2) back-propagation, where error propagation and weight adjustments occur. It is often assumed for this type of network that all PE's "share" the responsibility for error, that is, that the calculated error, e, is redistributed back to all PE's by adjusting the values of the weights entering each PE. The term "back-propagation" refers to the backward redistribution of error combined with these weight adjustments. This redistribution may be accomplished in various ways and will depend on the learning mechanism for the network. Minimization of the error function can also be done in different ways; some of these ways are as follows: (a) the use of a derivative of the error function (e.g., gradient descent in the interconnection weight space) to adjust weights so as to minimize overall error; (b) making a random search; or (c) genetic algorithms, GA's, which are slow but are a method of choice for complex systems with large data sets.

Step 1: Feedforward of Input Data. The calculations of a BPN begin upon acceptance of input data by the input layer, which is designated by an upper index "$s-1$". The input data may have been preprocessed and normalized. The input layer sends the data to the first hidden layer, upper index "s", to which it is connected. Only one hidden layer is used. Each PE_j of the hidden layer takes a sum of inputs to the $s-1$ layer multiplied by the weights that connect them to PE_j, to generate $\sum_i(w_{ij}^{[s]}x_i^{[s-1]}) = I_j$. The "internal activation" I_j is then transformed by a nonlinear function F to yield

$$x_j^{[s]} = F(I_j) = F(\sum_i (w_{ij}^{[s]} x_i^{[s-1]})) \quad (1)$$

where $x_j^{[s]}$ is now the output of PE_j in layer s and w_{ij} is the connection weight between PE_j of s and PE_i of $s-1$. Feedforward operation also takes place from the hidden to output layer. The output for PE_k of the output layer is $x_k^{[s+1]}$, which may be compared with the desired value for the same PE_k, namely, $D_k^{[s+1]} \equiv D_k$.

Step 2: Back-Propagation of Error. Successful operation of a back-propagation learning mechanism requires that F satisfy two requirements: (1) it has to be differentiable, and (2) it must increase monotonically. Such a function will have a derivative that must go to zero as the function goes to infinity. The most efficient of such functions are the hyperbolic tangent $T = \tanh(I)$ and the sigmoid function $S = 1/(1 + e^{(-I)})$ whose derivaties are

$$T' = (1 + T)(1 - T) \quad (2)$$

$$S' = S(1 - S) \quad (3)$$

If back-propagation of error is based on gradient descents, an analytically well-defined quadratic global error function E may be taken as the measure of error:

$$E = 0.5\sum_i (D_i - x_i)^2 \quad (4)$$

where $(D_i - x_i)$ is the local error at PE_i. The desired values D_i are often cited as vector components. For example, (1,0,0) may be the set of desired values for the situation in which there are three PE's in the output layer and the transfer function is S; or $(1,-1,-1,)$ may be an equivalent set when the transfer function is T. The global error function is minimized by minimizing $(D_i - x_i^{[s+1]})$ for each PE of the output layer. The only process available for doing this is the adaptation of all weights w_{ij} of the weight space in which the gradient descent takes place.

The parameter that is actually back-propagated is the change of the local error function, which is represented by

$$e_j^{[s]} = -dE/dI_j^{[s]} \quad (5)$$

for PE_j of level s. For the output layer, eqs 4 and 5 lead to

$$e_k^{[s+1]} \equiv e_k = -dE/dI_k = -(dE/dx_k)(dx_k/dI_k) = (D_k - x_k)F'(I_k) \quad (6)$$

where $F'(I_k)$ is the derivative of the transfer function in the output layer. The error function in the output layer now becomes the "input" for back-propagation. The error in the hidden layer PE's is evaluated in the same way, and, akin to forward propagation of an input signal, the output layer error is back-propagated to yield the error for PE_j of the hidden layer $[s]$

$$e_j^{[s]} = F'(I_j^{[s]})\sum_k (e_k^{[s+1]} w_{kj}^{[s+1]}) \quad (7)$$

where $F'(I_j^{[s]})$ is the derivative of the transfer function at the "output" layer and $e_j^{[s]}$ is the error value for the hidden layer.

It is clear from eqs 6 and 7 that F' controls error behavior. If the value of the transfer function (or the activation value) is large, the derivative is small (i.e., for large input values, both S' and T' tend toward zero); the error function is also small and, at that point, learning for that particular PE can be stopped. Substitution of F' by either T' or S' respectively, provides explicit forms for the error calculation.

Step 3: Adaptation. The errors for each PE in the output and hidden layer are given by eqs 6 and 7, respectively. The weights connecting the two levels can now be adjusted using gradient descent

$$dw_{ji}^{[s]} = -K(dE/dw_{ji}^{[s]}) \quad (8)$$

where K is a learning coefficient that speeds-up or slows-down network adaptation in the weight space. The changes occurring in that space are proportional to negative gradients in the error surface E. The local values for weight corrections may be written

$$dE/dw_{ji}^{[s]} = (dE/dI_j^{[s]})(dI_j^{[s]}/dw_{ji}^{[s]}) = -e_j^{[s]} x_i^{[s-1]} \quad (9)$$

so that the change of weight function for the ij connection becomes

$$dw_{ji}^{[s]} = Ke_j^{[s]} x_i^{[s-1]} \quad (10)$$

Starting coefficients for w_{ij} typically lie between -0.1 and $+0.1$. Thus, the process of adaptation consists of weight adjustments which are proportional to the error at a particular PE multiplied by the input to that PE on the connection to be adjusted.

In order to operate, a BPN must be trained for a desired purpose, for example, extraction of RS's. The input data, the MLS's, are separated into training and testing sets. The training set is used for adaptation of the BPN; the testing set, for the evaluation of performance. The weights in the BPN may be updated after each step, or after a certain number of steps called an "epoch". The introduction of epoch adjustments often improves convergence and may simplify training. The use of desired output values is known as "supervised learning".

3.2. Parameters of a Back-Propagation Network. Adaptation demands large numbers of computational steps, and network operations must be implemented by software. A simple BASIC, FORTRAN, or C language program can operate on a modest hardware configuration (e.g., MacPlus) with considerable success. Most of our activity centers on 486 EISA/66 and Sun workstations, but we also use parallel processing computers (e.g., MasPar). However, the use of parallel machines for neural networks is not well developed.

The market for neural network programs is growing rapidly, and there are many programs available (see, for example, ref 7) that can be used to explore RS capability. We have found the Neural Works Professional II/PLUS program package[16] to be convenient. Its advantages are as follows:

(1) It is an open system, and a user can construct all network elements and completely control all program development and processing. The advantage lies in the modular incorporation of program units written in C language and interactive "visual" presentation of network objects (connections, PE's, histograms, and error diagrams) at all levels of processing. Thus, one can develop new networks and test them with existing data sets. The primary function of the software package is fast construction and reliability, rather than the provision of ready-made products. The "open" approach differs from many "first principle" programs, where users have no control of key decisions in the software development.

(2) It provides fast "testing" of standard networks for different architectures because modular packages for a variety of operations are available.

(3) The ease with which the parameters are controlled and the transferability of modules permits an accurate description

of the details of different network configurations. Results obtained with different networks are readily compared because mathematical rigor and standardization are easily preserved when network architectures are reconstructed.

In comparison with a slow hardware system (e.g., MacPlus), the almost unlimited capacity of a 486 workstation, as measured by the number of PE's that may be used, as a major improvement, reducing calculation times by >100. No significant limitations on PE numbers should occur for Sun workstations. In a typical BPN for RS purposes, it takes <30 s on a 486 station to perform 10 000 learning presentations or cycles, a process that consumes hours on slower hardware configurations.

A considerable amount of time is often spent on data preprocessing. We use histograms and normalizations to preprocess the MLS's. And, in order to take advantage of the nonlinear transfer functions, we also use rescaling procedures. Rescaling refers to a (linear) transformation of input data into a specific range of values. For example, for an input data set that spans the range 0–1000, preprocessing may be used to rescale it into the domain 0–1 where nonlinearity of sigmoid functions is most distinct. For a $\tanh(T)$ function, this domain is -1 to $+1$. Preprocessing usually speeds up learning, but has little effect on training outcomes. Preprocessing is also helpful in avoiding local minima: This property of gradient-based systems refers to the "inertia" which causes processing to "stack" at the first available local minimal error position on the weight surface. The system may then refuse any further learning for that particular PE.

The BPN's used here permit one to monitor local behavior at each PE, the weight connections, and the global parameters for individual network layers. Windows representing sets of computational and evaluational parameters can be opened for each PE and each weight connection. These windows show histograms of weight activity, RMS errors, and scatter diagrams (confusion matrices) of predicted versus desired values for that PE or that layer. Pearson's R correlation coefficients, which lie in the range -1 to $+1$, are also shown and are particularly useful. The value $+1$ indicates "complete positive correlation"; the value -1, a "complete negative correlation"; and the value 0, the absence of any correlation or, equivalently, a training failure. The weight histogram depicts the distribution of activity per connection and indicates whether or not learning is correctly distributed over all PE's in the hidden layer. The various indicators discharge a very important duty: they show how the network performs during training and they give an evaluation of results during testing or recall.

Different recipes for the number of PE's in the hidden layer are available. We usually start with a maximal number (say twice that in the input or output layers) and then apply a "pruning" mechanism to eliminate those PE's that are not active in training. The activity of the weights connecting the PE's is always checked during training. The weights of smallest absolute magnitude are the least active and can be "pruned". The performance of the network is evaluated with the PE's and connection weights deleted. We have rarely used more than one hidden layer. An additional bias PE may be added to the input layer in order to ensure that the transfer function is maintained at a reasonable level of activation.

Weights are also adjusted during network adaptation. The choice of the pretraining weights will not affect the results. Therefore, the initial set of weights is randomly distributed in the range -0.1 to $+0.1$; that is, they take very small initial values. The values of the learning rate coefficient is less than 1: 0.3 for the hidden layer and 0.15 for the output layer (these values may be adjusted if the need to improve learning rates develops).

3.3. Information Content: Graceful Degradation and Interpretation of Radiation Signatures. The adaptation of weights during BPN training constitutes learning. That is, the information content specific to the solution of a given problem is contained in the connection weights and in the way they are distributed among the PE's of all layers. That distribution of information is responsible for the most important properties of ANN's, namely, their ability to deal with parallel distributed systems in a nonlinear manner and their utilization of the "graceful degradation" property which enables them to deal with incomplete data sets.

Graceful degradation refers to the smooth, controlled, skilled (graceful) degradation of network functioning when some PE's cease to function and their weight connections become irrelevant. There are two mechanisms that invoke such behavior: (1) training on uncertain or incomplete information sets, which causes "degradation" of some PE's because of the poor learning acquired during training and the resulting network incapacity to perform properly when the need to process more complete information sets occurs, and (2) training on complete information sets coupled with a posteriori disablement of certain PE's which are prerequisite to proper network performance. Since networks may still perform under both circumstances, information is surely not "localized" at some PE's but is more evenly distributed throughout the entire trained network.

Graceful degradation invoked by the second mechanism is sometimes associated with the "fault tolerance" concept in conventional nonadaptive systems. It occurs when pruning of parts of the network takes place or when hardware components fail. Fault tolerance, however, refers to a system which remains fully operational when a small number of the system elements fail but are "replaced" by other system elements.

In actuality, graceful degradation is best related to the property of "generalization", that is, the ability of an ANN to generalize from the input/output pairs on which it was trained and to produce sensible outputs from previously unseen inputs. If the testing and training sets are not very different, the performance of the network will gracefully degrade as it moves from training to testing, the "grade" of degradation being taken as a test of the network for that data set. As the training and testing sets are presented, the network must deal with "uncertainties" because information not in the training set may exist in the testing set, and vice versa. Conventional nondistributed programmable systems simply cannot handle such "distributed uncertainties".

We now query the information content of a trained BPN and whether or not it may facilitate the business of model extraction. We note first that ANN RS applications require little in the way of a priori assumptions about MLS's and their structures. We also note that the quantitative interpretation of network processing can be used for risk assessment, even when mechanistic models do not exist. We find, in other words, that a BPN which has been trained for RS extraction can be used to construct physico/chemical, even biological, models of radiation-induced effects. The success of ANN's indicates, then, that MLS's contain important discriminatory information about mechanisms of radiation-induced damage and that ANN's can use that information. This new finding suggests a new role, namely, the use of adaptive feature recognition and classification techniques to forecast mechanisms.

Only a few strategies are available for probing ANN processing. One strategy lumps together all methods similar

to those used in pattern recognition and image preprocessing (e.g., dithering or jogging the input data): the input parameters are varied in order to ascertain the dependence of overall processing on particular input PE's. This sort of approach may well pinpoint those molecular lesions that can serve as radiation markers, and it may well identify relations between lesion groups and some specific physico/chemical properties. The second strategy puts the accent on network architecture, on specific nonlinear transfer functions, or on structure level adaptation of networks, etc. The structure level adaptation includes processes such as adaptive reorganization of the network connections and the generation and pruning of PE's during learning. These latter methods, originally developed for optical neural networks, are not used here.

4. RESULTS AND DISCUSSION

4.1. Molecular Lesion Spectra and Computational Spaces for Back-Propagation Networks. Questions about the computational space for RS's relate directly to the representation and preprocessing of MLS's. We shall focus on MLS's and RS's induced by EM radiation, specifically VIS, UV, and γ radiation, for two reasons: (1) the fact that somewhat complete lesion assays exist for these three EM radiation types and (2) the fact that the mechanisms of EM interactions with matter are much better known than for other radiation types. We have recently studied a computational space for particulate radiation (specifically α's, protons, and electrons). While the approach was slightly different from that for EM radiation, the two computational spaces can be unified. The reason for a different strategy is that the first task of a neural network is to identify sets of radiation types, and the particulates are best categorized by linear energy transfer (LET).

The elements of the computational space for EM radiation types are MLS's or histograms, and it is on that space that the BPN is trained. We will show the limits of that computational space in detail and discuss the information content of BPN's and graceful degradation.

The choice of molecular lesion spectra for these studies of EM computational space and network training was based on coverage of as large a radiation frequency (wavelength) range as possible. For example, experimental data are available for low-frequency nonionizing visible [VIS (limit of UV-A), 405 nm; Peak and Peak (1991)[13]], intermediate ionizing [UV-C, 254 nm; Setlow and Setlow (1972)[14]], and highly ionizing radiations [γ-ray, ^{137}Cs ($\sim$0.6 MeV), ^{60}Co ($\sim$1.1 MeV); Ward (1988)[15]].

The molecular lesions assayed were

index	name	abbreviation
1	**cytidine hydrate**	**ch**
2	**thymine dimer**	**d**
3	**pyrimidine adduct**	**pa**
4	**DNA-protein cross-link**	**dpc**
5	**DNA-DNA cross-link**	**ddc**
6	**single strand break**	**ssb**
7	**double strand break**	**dsb**
8	**locally multiply damaged sites**	**lmds**

The source authors[13-15] and Cadet and Vigny[16] should be consulted for definitions of the lesion types. The MLS's (or histograms) for these three radiation have been constructed as bar graphs of lesion frequency versus lesion type, the bar sum in any one graph being normalized to 100.

Intercomparison of the MLS's (Figure 2) is revealing. The 405-nm distribution is quite similar to the γ-ray distributions if the **lmds** lesion category, index **8**, of the latter is disregarded, and both the 405-nm and γ-ray distributions are markedly different from the 254-nm spectrum. According to radiation quality considerations, on the other hand, it is the 405- and 254-nm distributions which should exhibit similarity, and both should be quite different from the γ-ray distribution. It is clear, then, that radiation quality does not provide an apt classification of DNA damage at the molecular level and that the nature and loci of the energy deposition events is the significant descriptor. In specific, the 254-nm distribution reflects direct absorption by, and deposition in, the DNA bases, events which are totally impossible at 405 nm and statistically improbable at $\sim$1 MeV; the similarity of the 405-nm and γ-ray spectra, the **lmds** category of the latter excepted, is probably a consequence of energy deposition in the aqueous environment, directly for γ-rays and through the intermediacy of some third-party-chromophore for 405 nm.

RS's as represented by MLS's define both the causal radiation and the manner in which it interacts with matter (i.e., with DNA). Thus, there is a reason to expect that ANN processing will lead to recognition of RS classes, that is, that it will recognize radiation types which have some degree of commonalty (e.g. ionizing, nonionizing, etc.) Such a classification ability when coupled with an understanding of mechanisms will simplify radiation risk assessment and the correlation of RS's with final health outcome.

At this time, there is no unique and generally acceptable standard for biological dose. The situation is even more complicated in nanodosimetry. The data used here are based on a biological normalization related to the "efficiency of killing", namely, the dose (D_{37}) or fluence (F_{37}) which permits a survivor fraction of 37%. This type of normalization presents at least two difficulties. Firstly, it appears that the ratio of lesion production to mortality is a function of the radiation. For example, if 405 nm is, as it appears to be, inefficient in killing, a D_{37} dose at 405 nm implies measurement of the lesion production at very high fluence. That is, (D_{37}/F_{37}) for γ-ray radiation may well be more than 10^6 times larger than (D_{37}/F_{37}) for 405-nm radiation. This question, in turn, generates the second difficulty: if the quadratic component in the linear-quadratic dependence[16] of effect on dose is dominant, or merely important, the distribution of lesion intensity in the lesion spectra will vary with dose (and fluence), and the concept of a signature will run into difficulty. Thus, it is imperative that the effect of fluence on lesion distribution be investigated in detail. If it is found that lesion distributions are dose-dependent, this will require an extension of the concept of a signature promulgated here to one which acquires multidimensionality in the sense that, for electromagnetic radiation of wavelength λ, the radiation signature $RS = RS(\lambda)$ will have to be replaced by $RS = RS(\lambda,F)$. In terms of the adaptive parallel distributed processing approach presented here, such a change poses no problem. In fact, it should produce better signature discrimination characteristics.

The description of a radiation type for the purpose of RS's is certainly an important task for future RS studies. In the case of EM RS's, the main descriptors include wavelength, energy, and fluence, although polarization and other variables can also play a role. There is no a priori distinction between EM radiation types, however, from the point of view of the computational space for neural networks.

Unfortunately, only three lesions, **{dpc, ssb, dsb}** or **{4, 6, 7}** are common to all the actual measurements of the three sets of authors cited under "Source of Data". We refer to this set of lesions as "the common experimental set (CES)". However, this set of three lesions can be expanded to five, **{d, dpc, ssb, dsb, lmds}** or **{2, 4, 6, 7, 8}**, such that this set of five can also be considered to represent actual measurements. We refer to this set of five lesions as "the verifiable experimental set (VES)". This expansion from three to five and, then,

Chart 1

one lesion + seven "0's":	two lesions + six "0's":	three lesions + five "0's":	etc.
{**ch**,0,0,0,0,0,0,0}RT1	{**ch**,**d**,0,0,0,0,0,0}RT1	{**ch**,**d**,**pa**,0,0,0,0,0}RT1	
{**ch**,0,0,0,0,0,0,0}RT2	{**ch**,**d**,0,0,0,0,0,0}RT2	etc.	
{**ch**,0,0,0,0,0,0,0}RT3	{**ch**,**d**,0,0,0,0,0,0}RT3		
{0,**d**,0,0,0,0,0,0}RT1	{**ch**,**0**,**pa**,0,0,0,0,0}RT1		
{0,**d**,0,0,0,0,0,0}RT2	{**ch**,**0**,**pa**,0,0,0,0,0}RT2		
{0,**d**,0,0,0,0,0,0}RT3	{**ch**,**0**,**pa**,0,0,0,0,0}RT3		
{0,0,**pa**,0,0,0,0,0}RT1	etc.		
{0,0,**pa**,0,0,0,0,0}RT2			
{0,0,**pa**,0,0,0,0,0}RT3			
etc.			

from five to eight has been justified elsewhere.[3,4]

Our interest now is 2-fold: first, to investigate the information content of the network and its graceful degradation, and second, to investigate the minimal lesion sets which suffice for provision of a high-fidelity signature. The first part is best carried out using an experimentally "soft" lesion data set. The second goal is best accomplished within the context of an experimentally "hard" lesion data set. Consequently, we shall first test and report on the eight-lesion data set (i.e., the assumed experimental set, AES) and then we will down-size to the five-lesion verifiable experimental set, VES.

4.2. Graceful Degradation and Radiation Signatures. We have defined graceful degradation as the ability of adaptive systems to deal with poor or incomplete information. We have also identified the training that occurs with uncertain information as one of two mechanisms that can be used to invoke and test graceful degradation. We now discuss the eight-lesion set of MLS's. We will use training and testing data sets that contain incomplete information. We will construct new MLS's that artificially replace the actual values of a certain lesion by a "no number" or "zero" value. We will not, however, introduce deliberately "wrong" data. Other "uncertain" MLS's can be constructed using an "average value" replacement of the actual value. We emphasize that we are primarily concerned with the graceful degradation and are also interested in the evaluation parameters for the network: the R correlation coefficients for output layer PE's for different training and testing sets and the percentage of inaccurate predictions.

We start with the eight lesions, **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8**, and with the three RS's for the three EM radiation types used for construction of the computational space. The BPN was first trained to recognize RS classes for 405-nm, 256-nm, and γ radiation. The results, using the sigmoid transfer function, consist of three RS's: the corresponding vectors are (0.96, <0.02, <0.02), (<0.02, 0.93, <0.02), and (<0.02, <0.02, 0.92). The degree of orthogonality of these vectors is better than 90%. We then investigated graceful degradation using the same eight-lesion types but with a tanh(T) transfer function. The use of the tanh function is associated with situations where exceptions are more important than the average, whereas sigmoid functions are used when the average is more important than the exceptions. Since the data base in this case contains incomplete MLS's, zeros being used for certain lesion frequencies, we are dealing with the presentation of "fragments" or "pieces" of the MLS's used in the first instance. The newly constructed MLS's were as given in Chart 1, where RTi denotes the EM radiation: 405 nm ≡ RT1; 254 nm ≡ RT2; γ radiation ≡ RT3. This data base consists of 37 sets of 3-fold MLS's when MLS's with all zero values are eliminated. This situation is similar to the presentation of fragments of figures or sections of patterns to an ANN system in order to determine the continuing ability to identify them as the original unfragmented figure or unsect pattern. Of course, there is a limit beyond which one cannot reduce information without losing the ability to identify.

Table 1

		Pearson's R coefficients					
		training set			testing set		
case no.	% inaccurate	PE1	PE2	PE3	PE1	PE2	PE3
1	6	0.97	1.0	0.96	0.98	0.92	0.96
2	15	0.99	0.99	0.99	0.85	0.84	0.77
3	6	0.99	0.99	0.99	0.96	0.85	0.89
4	18	0.99	0.99	0.99	0.96	0.16	0.79

After construction of the "fragment" MLS's, many choices exist for the composition of the training and testing data sets. The ratio of training set size to testing set size was also varied. We shall discuss the following cases:

case 1	100% are in the training set
case 2	1:1 ratio; 50% in training set (19), 50% in testing set (18)
case 3	1:2 ratio; 33% in training set (12), 66% in testing set (25)
case 4	1:1 ratio; 50% in training set (19), 50% in testing set (18)

Cases 2 and 4 differ only in the sets used for training and testing.

Since PE outputs lie between −1 and +1, the average being 0, a correct prediction is associable with the sign of the output value. Inaccurate predictions (%), then, are calculated as the ratio of inaccurate predictions (wrong sign) to the total number of MLS's. The results are given in Table 1.

The result for case 1 is expected because the two sets are identical; the small variations are due to noise induced training. The accuracy of performance is reflected in the rate of successful classification: only 6% were classified in the wrong RS class. When presented with a training set that contains only 50% of the data base, case 2, but which is still balanced, the RS classification inaccuracy increases to 15% and the R values fall to R = (0.85, 0.84, 0.77) for the three output PE's representative of the three RS classes. When a lesion type or RS class is very badly represented, as in case 4, where only a few lesions are used, the inaccuracy of the 50% set increases to 18% and the R values become unacceptably poor, being R = (0.99, 0.16, 0.76) for the testing sets.

There is a boundary between the usefulness of graceful degradation and performance failure. The moderate, balanced reduction of the number of lesions in the training set does not usually degrade performance. Indeed, BPN's can sustain feature extraction capacity even when the training set is quite small: for example, with only one-third of the original data base incorporated into the training set, only a 6% inaccuracy is induced, and the best values are R = (0.99, 0.99, 0.99) for the training and (0.96, 0.85, 0.89) for the testing sets. Similar results (i.e., ~6% inaccuracy) were also obtained for random, balanced "half" training sets. Consequently, the BPN's used here exhibit the property of graceful degradation, which implies that (1) the network learns to extract complex, characteristic features (signatures) or models from RS's and describes them in terms of weights; and (2) balanced mixtures of good experimental data, poor experimental data, and calculational data may be used in the search for signature capability. This last property is important for particulate radiation types where data bases are limited, fragmented, and "soft".

4.3. Experimental Data Sets and Criteria for Radiation Signatures and Radiation Markers. We now consider the serach for RS's: In specific, we seek the size of the minimal lesion set which permits a high-fidelity RS. Only good

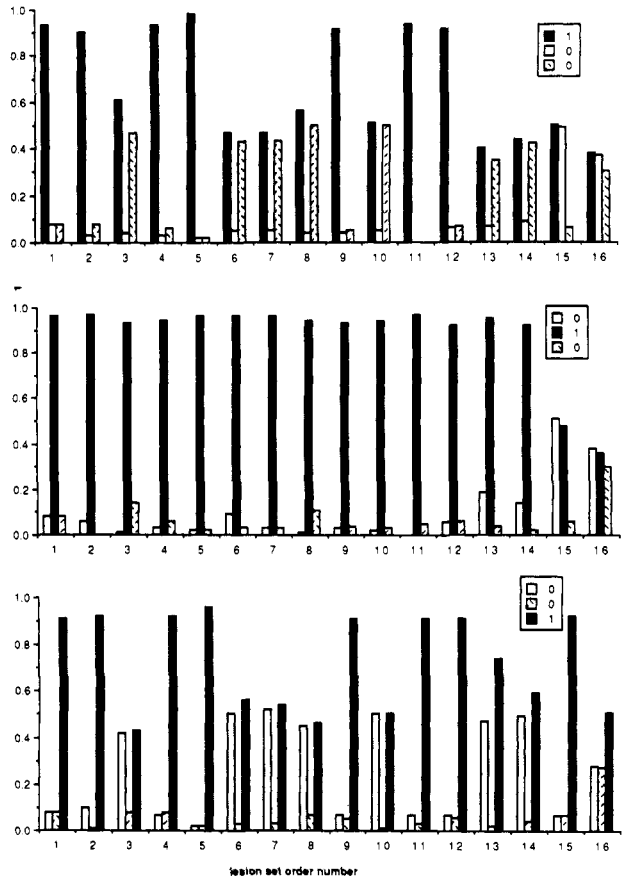

Figure 4. Application of a BPN to five-, four-, and three-lesion sets. A total of 16 sets is considered. The abscissal number designates the lesion set. The values of output layer PE's for each set are shown for three EM radiations: 405 nm (upper); 254 nm (middle); γ (lower). The desired values for the three EM radiations are the vectors (1,0,0) (upper), (0,1,0) (middle), and (0,0,1) (lower), which are shown in side boxes. A value 1 is denoted by dark shading. The criteria for RS are as follows: only if the component that replaces "1" in any vector is greater than 0.9 for all three and the component that replaces "0" is less than 0.1 for all three vectors is the set in question considered to provide a high-fidelity signature. This criterion is equivalent to Pearson's R at a 90% level. It is clear from the diagrams that lesion set numbers 1, 2, 4, 5, 9, 11, and 12 are RS's. These lesion sets are **{2,4,6,7,8}**, **{4,6,7,8}**, **{2,4,7,8}**, **{2,4,6,8}**, **{2,4,8}**, **{2,6,8}**, and **{2,7,8}**, respectively.

experimental data should be used to train the BPN, which suggests the 5-member MLS's. Sigmoidal transfer functions were used. Similar results are obtained with tanh transfer functions. The input data consisted of normalized histograms. The desired results, in terms of orthogonal vectors are (1,0,0), (0,1,0), and (0,0,1) for the three EM radiation types.

All four-, three-, and two-lesion sets which can be constructed from the five-lesion VES set, a total of 25, have been processed. The results for four- and three-lesion sets are given in Figure 4. Only one two-lesion set provides an acceptable RS. At the level of the three-lesion sets, the various sets branch into two categories, those which are CES and those which are VES. While the vectors of Figure 4 are not pertinent to a linear space, and normal concepts of orthonormality do not apply, the criteria for radiation signature remain stringent: Only if the component under "1" in any one vector is greater than 0.9 for all three vectors and the components under "0" are less than 0.1 for all such components in all three vectors is the set in question considered to provide a high-fidelity signature. This criterion, in association with RMS convergence to zero, ensures "orthonormalities" at the 95% level, as measured for example by confusion matrices and Pearson's R value.

The five-lesion data set, **{d, dpc, ssb, dsb, lmds}** or **{2, 4, 6, 7, 8}**, provides a good radiation signature. A BPN, the results of which are shown in Figure 3, was successfully trained to evaluate signature fidelity; the output vectors are (0.93, <0.08, <0.08), (<0.08, 0.96, <0.08), and (<0.08, <0.08, 0.91). We have used several configurations for this network; we have varied the number of PE's in the hidden layer from 5 to 12; we have used sigmoid and hyperbolic tangent transfer functions; and we have modified the learning coefficients in order to establish the limits of network adaptability, to improve learning rates, and to identify instabilities in the learning process. While it is clear that the present configuration does give satisfactory results, it is also evident that there is considerable leeway for further improvement. Indeed, it is certain that preprocessing of data and automatic pruning procedures would improve efficiency, particularly for larger data sets. However, the salient point is that we have not identified any problem that could detract from our results.

Five four-lesion data sets exist and three of these possess signature capability. For example, the **{2, 4, 6, 8}** set provided excellent signature, the output vectors being (0.98, <0.02, <0.02), (0.02, 0.96, <0.02), and (<0.02, <0.02, 0.96). The counterpropagating network was also successful for this set. For **{2, 4, 6, 7}**, however, neither network could discriminate between the 405-nm and γ-ray MLS's, because both of these EM radiation types have very similar **{2, 4, 6, 7}** spectra.

A very large number of three- and two-lesion sets exists. Three of the three-lesion sets and one of the two-lesion sets, none of them belonging to the common experimental set, provide excellent signatures.

A number of conclusions result:

None of the three-lesion sets that provide signature belong to CES. Thus, insofar as distinction of the three radiation types is the objective, it is clear that the maximally efficient three-lesion sets have not been investigated in adequate experimental detail. On the other hand, these results also point up the power of neural networking as intervenor at the point of experiment in the signature recognition process.

No marker is available for 405-nm radiation, although **DNA–protein crosslinks** come close to fulfilling some requirements.

It is surprising that a two-lesion set should serve to identify each of three separate radiation types. This, nonetheless, is the case. The simplest situation in which this can occur is when each one of the two lesions serves as a basis vector for each of two different radiation types and the third radiation produces neither lesion. Such is the case here. The important point is that the procedure has identified two radiation markers.[18-19] Thus, one concludes from the set **{2,8}** that **thymine dimers** are a marker for 254-nm radiation and **locally multiply damaged sites** are a marker for γ-radiation.

The frequency (probability) of a marker has an importance that may supersede its signature abilities. It can, for example, provide an internal measure of dose, fluence, etc., for use with certain neural network architectures which have particular relevance to dose and or fluence functionality extraction.

The value of the neural network approach becomes very obvious as the size of the lesion set is progressively reduced in the series **8 → 5 → 4 → 3 → 2** lesions. In specific, it has been found that the five-lesion set provides excellent signature fidelity. This fidelity is shown also for some of the four-lesion and three-lesion sets and, surprisingly, for one of the two-lesion sets. At the same time, it becomes patently obvious that the identification of signature capability is facilitated by the use of neural networking.

It has been shown that a complete lesion set is not always necessary to the provision of a signature characteristic and that, using neural networking, a minimal lesion set that will serve signature purposes can be selected. Given the difficulty of lesion measurements, the ability to minimize demand is an important consideration. In the present instance, it appears that a two-lesion set will suffice for the three radiation types of interest. Nor does the quality of this signature suggest that a specific three lesion set, such as {**2**, **4**, **8**} is any better. Of course the evaluation of success is inseparable from an evaluation of the experimental data bases, a subject which requires further discussion.

The supposition that certain lesions may serve as markers for specific radiation types has been raised by the neural network analysis. Radiation markers, **d** for 254 nm and **lmds** for γ, have been identified. This is an important observation because it suggests the application of another neural networking property, namely, its ability to deduce the functional dependencies of lesion frequency on dose, fluence, etc.

5. CONCLUSION AND RESEARCH DIRECTIONS

The goal of our preliminary studies[1-4] of RS capabilities was to establish the RS concept and to verify that the distribution of molecular lesions (i.e., MLS's) generated in DNA by exposure to a particular radiation was a characteristic of (i.e., was a signature for) the causal radiation.

We have initiated the formation and expansion of lesion data bases. The U.S. DOE has a number of research projects concerned with radiation damage in DNA. However, in response to our requests for data relevant to RS studies, we were surprised to find that very little pertinent data have been generated. We surmise that this happenstance is a consequence of the absence of models and ideas for which such data would have had relevance. Now, that the MLS as RS paradigm has been developed, we hope these data deficiencies will be corrected.

It is now clear that there may well be two "classes" of MLS's: the first class consists of lesions formed in an isolated, cold environment (0 °C approximation) where enzymatic repair does not occur, whereas the second class consists of lesions formed in a warm environment where the repair apparatus does function.

The most important accomplishments are as follows:

(1) A trained network can be used to extract signatures and markers from molecular lesion sets of various sizes. The success of neural network processing is based on two distinguishing characteristics: the ability to deal with parallel distributed systems in a nonlinear manner and the property of "graceful degradation".

(2) It is possible to quantify the "sensitivity" and precision of the various signatures and to identify the minimum lesion sets required to distinguish different but limited groups of radiation types.

(3) Neural networks can suggest the molecular lesion sets that are most appropriate for particular purposes. The results of such studies should enhance the formation of molecular lesion data bases for the application of RS's to radiation risk assessment.

(4) Neural network processing indicates that MLS's contain important discriminatory information about the mechanisms of radiation-induced damage. This finding suggests a new role for RS research: namely, the use of adaptive models for the interpretation of results. We are investigating situations where it is possible to show how an ANN can predict the biological consequences of radiation and the physico/chemical and/or biological processes which are involved.

(5) We have initiated formation of lesion data bases for RS research and biomedical data bases pertinent to the correlation of MLS's with clinical outcomes.

(6) We have sought signatures and markers of a nonlethal nature which arise from sublesion branching into mutagenic but not carcinogenic channels. Such signatures or markers should be good, personal, cumulative biological dose indicators. Preliminary study is in progress.

REFERENCES AND NOTES

(1) McGlynn, S. P.; Rupnik, K. How to Model Molecular Lesions *and* Adaptive Lesion Models. *The 24th DOE Contractors' Meeting on Radiation Signatures: Facts or Fancy*? April 15–17, 1991; Columbia University: New York, 1991.

(2) McGlynn, S. P.; Varma, M. N. Radiation Signatures, Presented at the Council of the European Community Workshop on Biophysical Modeling of Radiation Effects, Padova, Italy, September, 1991. In *Biophysical Modelling of Radiation Effects*; Chadwick, K. H.; Moschini, G., Varma, M. N., Eds.; Adam Hilger (IOP): New York, 1992; pp 293–300.

(3) Rupnik, K.; McGlynn, S. P. Molecular Lesion Spectra as Radiation Signatures. *Spectrosc. Lett.* **1993**, *26*, 873–886.

(4) McGlynn, S. P.; Rupnik, K.; Varma, M. N.; Klasinc, L. Radiation Signatures and Radiation Markers. *Radiat. Prot. Dosim.* **1994**, *52*, 155–169.

(5) Battista, J.; Rupnik, K.; McGlynn, S. P. Proposal submitted to the Centers for Disease Control (CDC), Atlanta, GA, 1993.

(6) McClelland, J. L.; Rumelhard, D. *Parallel Distributed Processing*; MIT Press: Cambridge, MA, 1988. {Some feedforward BPN architectures avoid supervision by using another layer. There are also many network architectures for nonsupervised learning, where back-propagation is not used at all. The advantages/disadvantages of various networks and learning rules are beyond the scope of this simple presentation. The interested reader is inferred to refs 6–11.}

(7) Hecht-Nielsen, R. *Neurocomputing*; Addison-Wesley: Reading, PA, 1990.

(8) NeuralWorks; NeuralWare, Inc., Penn Center West, Building IV, Suite 227, Pittsburgh, PA 15276, 1991 and 1993.

(9) Kohonen, T. *Self-organization and associative memory*; Springer: New York, 1988.

(10) Grossberg, S. *Structures of mind and brain*; Reidel: Boston, 1982.

(11) Widrow, B.; Hoff, M. In *IRE WESCON Convention Record*, Part 4, Los Angeles, Aug. 23–26, 1960; pp 96–104.

(12) Vähäkangas, K. M.; Samet, J. M.; Metcalf, R. A.; Welsh, J. A.; Bennett, W. P.; Lane, D. P. C.; Harris, C. C. Mutations of p53 and ras Genes in Radon-associated Lung Cancer from Uranium Miners. *Lancet* **1992**, *339*, 576–580.

(13) Peak, J. G.; Peak, M. J. Comparison of initial yields of DNA-to-protein crosslinks and single-strand breaks induced in cultured human cells by far- and near-ultraviolet light, blue light and X-rays. *Mutat. Res.* **1991**, *246*, 187–191.

(14) Setlow, R. B.; Setlow, J. K. Effects of radiation on polynucleotides. *Annu. Rev. Biophys. Bioeng.* **1972**, *1*, 293–346.

(15) Ward, J. F. DNA damage produced by ionizing radiation in mammalian cells: Identities, mechanisms of formation, and repairability; In *Progress in Nucleic Acid Research and Molecular Biology*; Cohen, W. E., Moldave, K., Eds.; Academic Press: New York, 1988; Vol. 35, pp. 95–125.

(16) Cadet, J.; Vigny, P. The photochemistry of nucleic acids; In *Bioorganic Photochemistry*; Morrison, H., Ed.; Wiley: New York, 1990; Vol. 1, pp 1–272.

(17) Chadwick, K. H.; Leenhouts, H. P. *The molecular theory of radiation*; Springer-Verlag: New York, 1981.

(18) Weinstein, I. B.; Perera, F. P. *Banbury Report 13: Indicators of Genotoxic Exposure*; Cold Spring Harbor Laboratory: New York, 1982; pp 3–15.

(19) Hulka, B. S.; Wilcosky, J. C.; Griffith, J. D. *Biological Markers in Epidemiology*; Oxford University Press, New York, 1990; pp 214–225.