Alfred Kraxenberger
Anna A. Friedl
Albrecht M. Kellerer

GSF-Forschungszentrum, Institut für
Strahlenbiologie, Neuherberg

Computer simulation of pulsed field gel runs allows the quantitation of radiation-induced double-strand breaks in yeast

A procedure for the quantification of double-strand breaks in yeast is presented that utilizes pulsed field gel electrophoresis (PFGE) and a comparison of the observed DNA mass distribution in the gel lanes with calculated distributions. Calculation of profiles is performed as follows. If double-strand breaks are produced by sparsely ionizing radiation, one can assume that they are distributed randomly in the genome, and the resulting DNA mass distribution in molecular length can be predicted by means of a random breakage model. The input data for the computation of molecular length profiles are the breakage frequency per unit length, α, as adjustable parameter, and the molecular lengths of the intact chromosomes. The obtained DNA mass distributions in molecular length must then be transformed into distributions of DNA mass in migration distance. This requires a calibration of molecular length *vs.* migration distance that is specific for the gel lane in question. The computed profiles are then folded with a Lorentz distribution with adjusted spread parameter Γ to account for band broadening. The DNA profiles are calculated for different breakage frequencies α and for different values of Γ, and the parameters resulting in the best fit of the calculated to the observed profile are determined.

1 Introduction

Exposure of living cells to ionizing radiation gives rise to reproductive cell death, chromosome aberrations and transformation. These effects are considered to be largely due to radiation-induced DNA double-strand breaks (DSB) (for a review, see [1]). DSB occur also as intermediate step in the processing of other types of DNA damage, for example furocoumarin-plus-UVA-induced damage [2], and during cellular processes such as replication, recombination, or apoptosis.

The broad interest in the quantitative analysis of DSB is reflected by the availability of various assays for their measurement, including neutral sucrose gradient sedimentation, neutral elution, unidirectional electrophoresis and pulsed field gel electrophoresis (PFGE) techniques (for reviews, see [3] and [4]). In recent years the PFGE assays have found increasing interest, as they are less expensive than sedimentation assays and more sensitive than elution procedures. Shortly after the first description of the PFGE technique [5] the method was shown to be applicable to the DSB evaluation in the yeast *S. cerevisiae* [6–8]. More recently several groups have described the use of PFGE for DSB measurement in mammalian cells [9, 12].

To make full use of the PFGE technique, we developed an improved method for the determination of the DSB frequency in yeast chromosomal molecules. Genomic DNA of irradiated and unirradiated yeast cells is isolated and subjected to PFGE. As the largest chromosome of the yeast *S. cerevisiae* is about 2.2 Mbp long, all intact chromosomal molecules and their radiation-induced fragments can be separated according to size by choosing appropriate electrophoresis conditions. Ethidium bromide staining of the PFGE gels and measurement of the fluorescence intensity distribution within the lanes yields DNA mass profiles which comprise intact chromosomal molecules and fragments of different lengths. The intensity profile depends on the frequency and the distribution of radiation-induced breaks and on the migration behavior of the molecules. A random breakage model and the knowledge of the migration behavior of molecules during electrophoresis permits the computation of DNA mass profiles that correspond to the observed profiles in the gels. The comparison of the observed with calculated profiles has been applied for the DSB quantification in yeast cells exposed to ^{60}Co-γ-radiation [13]. The details of the calculation and the fitting procedures, including certain improvements in the technique, will be presented.

2 Materials and methods

2.1 γ-Irradiation of yeast cells

Yeast cells (*S. cerevisiae*) are grown to stationary growth phase, washed twice in ice-cold 0.1 M sodium phosphate buffer and irradiated at a concentration of 5×10^8 to 2×10^9 cells per mL in a ^{60}Co-Gamma-Cell (Atomic Energy of Canada, Ltd.). During irradiation the cells are held on ice, and immediately after irradiation EDTA is added to inhibit repair enzymes.

2.2 DNA preparation

Genomic DNA is isolated as described earlier [13]. Shortly, cells are mixed with liquid agarose solution and zymolyase, and plugs are formed using a mold. Lysis of the cell walls takes place during incubation of the plugs for 1 h at 37 °C in a 2-mercaptoethanol-containing buffer solution. Subsequently, the buffer solution is substituted by a solution containing proteinase K and *N*-lauroylsarcosine. After incubation for 24 h at 50 °C, the plugs are rinsed several times with 10 mM TrisCl, 10 mM EDTA, pH 7.5, and stored at 4 °C.

2.3 PFGE

We use three different electrophoresis systems: Transverse alternating field electrophoresis (TAFE) [14], contour clamped homogeneous field electrophoresis (CHEF) [15], and zero integrated field electrophoresis (ZIFE) [16], all commercially available (Beckman, Bio-Rad and Q-Life Systems, respectively). Gels are prepared using 0.9 % or 1 % LE agarose (UltraPure, BRL) and electrophoresis buffers suggested by the manufacturers (TAFE: 10 mM Tris, 0.5 mM EDTA, 4.5 mM glacial acetic acid; CHEF: 90 mM Tris, 90 mM boric acid, 2 mM EDTA; ZIFE: 90 mM Tris, 90 mM boric acid, 3.2 mM EDTA). TAFE gels are run for 18 h with 60 s pulse time at a constant current of 160 mA. CHEF gels are run either for 8 h with 60 s pulse time, followed by 18 h with pulses ranging from 90 s to 150 s, or for 14 h with 60 s pulse time followed by 14 h with 90 s pulse time, each at 5.5 to 6 V/cm. The field angle is 120°, buffer temperature is kept at 14 °C. For running ZIFE gels the manufacturer provides ROM cards, which supply separation strategies for optimal resolution in specific ranges of molecular length. For our purposes we use ROM card # 5 for best separation in the range from 0.1 to 1.1 Mbp. After electrophoresis the gels are stained for 30 min with ethidium bromide (EB); subsequently they are destained and treated with RNAse A for several hours.

2.4 Image analysis

The gels are placed atop of a UV source, and pictures are taken with a pixel synchronous monochrome charge coupled device (CCD) camera (XC77/PS, Leutron Vision, Germany) in combination with a real-time frame grabber (LFS-AT/PS/348kB, Leutron Vision, Germany). In front of the lens (MO-V-2508 0.95/25 mm C-mount, Olympus, Germany) a 590 nm glass filter is placed. Images are digitized on a grid of 512 × 512 pixels, which yields satisfactory spatial resolution to analyze our 20 × 20 cm gels. The system provides 256 gray levels (from 0 to 255, where 255 stands for the saturation level) for the intensity of the incident light at each pixel. After a single exposure of 40 ms the light intensities from typical gels cover the range from 0 (black) to about 240. Hence the light sensitivity of the camera is sufficient for our purposes, and there is no need for multiple exposure or longer exposure times. The distribution of light intensity along each lane is determined with a scan program written in "C", which is described in the next paragraph. The scan program contains routines to control the

camera functions, which are provided by the manufacturer of the frame grabber in form of a subroutine library. To simplify data collection, the gel lanes have to be arranged parallel to the columns of the camera pixel matrix. The width of one gel lane corresponds to approximately 20 pixel columns in the image. The left and right margins of a particular lane are marked with a cursor, and the mean brightness in every pixel row between the chosen borders is calculated by averaging the observed intensities. The averaging corrects for variations due to unequal loading of the wells and permits an improved intensity resolution. In a final step the relation between position (*i.e.* migration distance) in the gel and fluorescence intensity is written to an input file for the evaluation program PULSE described in the next section.

3 Results

It is assumed that sparsely ionizing radiation induces DSB independently and uniformly in chromosomal molecules of irradiated cells [17]. Hence the probability for breakage of the chromosomal molecules, and the size distribution of the resulting fragments can be calculated with a random breakage model, as will be described in Section 3.1. This size distribution is in a first step computed as a function of molecular length; it depends on the lengths of the original molecules and on the average frequency of DSB per unit molecular length. If the lengths of all intact molecules in the genome are known, one can derive the average frequency of radiation-induced DSB by comparing the calculated size distributions with the actual size distribution obtained from irradiated cells. The actual size distribution of DNA as a function of molecular length, however, cannot be observed directly. One determines instead the distribution of DNA mass along a gel lane as a function of migration distance by measuring the fluorescence intensity in EB-stained gels, which is assumed to be proportional to DNA mass density [18]. The calculated DNA mass distribution must, therefore, be transformed into a distribution in migration distances.

The migration behavior of nucleic acids in pulsed fields is not, as yet, fully understood. It is influenced by pulse duration, voltage, field angle, temperature, and agarose concentration [19]. However, the relation between molecular length and migration distance can be deduced from gels where yeast chromosomal molecules and/or length standards are separated. The positions of bands representing molecules of known lengths are determined and, with suitable interpolation, a calibration curve is derived (see Sections 3.2 and 3.3). One also needs to take into account that even molecules of identical length show deviations in their migration velocity; these deviations can be estimated from the widths of the observed bands (see Sections 3.4 and 3.5.1).

Computations with different assumed breakage frequencies together with an adequate description of the migration behavior of DNA molecules in a gel permits the computation of DNA mass distributions which can then be compared with the observed distributions. To obtain the best fit parameters an interactive computer program,

PULSE, has been designed for the evaluation of PFGE gels, where genomic DNA isolated from unirradiated and irradiated yeast cells was separated. The essentials of the program are described in the following sections.

3.1 Calculation of DNA mass profiles using a random breakage model

The fluorescence intensity after EB staining is taken to be proportional to the DNA mass; the profile of the fluorescence intensity along a gel lane then represents the DNA mass distribution along the lane. Figure 1 gives an example of a CHEF gel with DNA isolated from the diploid yeast strain BK0 [20], which was exposed to 0 (lane 1), 250 (lane 2), 500 (lane 3) and 750 Gy (lane 4) of γ-rays under hypoxic conditions. In the lane containing the unirradiated sample, each chromosome contributes to the total amount of genomic DNA in proportion to its molecular length. In the lanes containing irradiated samples, the number of intact chromosomal molecules decreases, and fragments of different lengths appear. The length distribution of the fragments can be deduced with a model, where each molecule is treated as a chain of residues with well-defined length. The probability, α, for breakage of the link between two residues is assumed to be constant over the whole genome, independent of the location and of other breaks.

In fact, we are not dealing with the simple case of a chain of single atoms as residues and covalent bonds as links, because radiation-induced DSB in chromosomal molecules result from two single-strand breaks (SSB) on opposite strands that can be a few bp apart from each other. In a strict sense one can, therefore, not assign the DSB to a specific covalent bond. But the subsequent solution of the random breakage model will be seen to be independent of the number of residues. It corresponds, in fact, to a model where uniform threads are randomly cut; the formulation in terms of discrete links is merely a convenient method to bring the mathematical derivation into a form that can be understood without sophisticated tools of probability theory.

Figure 1. Image of a gel loaded with DNA samples isolated from cells of the (*S. cerevisiae*) strain BK0 irradiated with 0 Gy (1), 250 Gy (2), 500 Gy (3) and 750 Gy (4) under hypoxic conditions. Gel run on a CHEF system as described in Section 2.3. Note that some of the bands are doublets.

The subsequent mathematical considerations will first give the solution for the case of a fixed length of the unbroken molecule; subsequently it will superimpose fragment spectra for molecules of different initial lengths. Exact solutions for the process were first given by Schulz [21] and Sakurada and Okamura [22] in the context of the decomposition of long polymer chains.

Suppose a chain containing L residues, *i.e.* $L - 1$ links. Assume that during irradiation s links are broken, yielding a breakage frequency of $\alpha = s/(L - 1)$. Hence, the probability for the breakage of one specific link is α, whereas the probability for one link not to be broken is $1 - \alpha$. A fragment of length l can be obtained in two ways: either by two breaks within the original chain, where the breaks are l residues apart from each other, or by one break localized l residues from the end of the molecule. (i) In the first case the number $n'(l)$ of obtained fragments of length l per initial molecule is: $n'(l)$ = (the number of sites, where the fragment can occur) $\times$ (the probability that two breaks occur at specific sites l links apart from each other) $\times$ (the probability that the chain between these two sites remains unbroken) $= ((L - 1) - l) \times \alpha^2 \times (1 - \alpha)^{l-1}$. (ii) The number $n''(l)$ of fragments of length l in the second case is: $n''(l)$ = (the number of possibilities that a fragment of length l is created at one end of the chain) $\times$ (the probability for a break at the l-th link) $\times$ (the probability that $(l - 1)$ links are unbroken) = $2 \times \alpha \times (1 - \alpha)^{l-1}$.

Hence the total number of fragments of length l obtained per initial molecule is

$$n(l) = n'(l) + n''(l)$$
$$= \alpha^2(1 - \alpha)^{l-1}\left((L - 1) - l + \frac{2}{\alpha}\right). \tag{1}$$

Consider now that the chain remains unbroken. The probability for this case is $(1 - \alpha)^{L-1}$. As there is only one possibility to obtain an unbroken molecule, their number per initial molecule is

$$n(L) = 1 \times (1 - \alpha)^{L-1}. \tag{2}$$

Up to now we have considered only the number of molecules of length l, for every $0 \leq l \leq L$, resulting from breakage of a chain of length L. In the genome of yeast cells we have to deal with K chromosome species with lengths L_i, where i runs from 1 to K. For example, in the case of haploid strains of *S. cerevisiae* holds $K = 16$, and in the case of diploid strains $K = 32$. Hence one has to calculate the sum over the number of molecules of length l resulting from breakage of each chromosome species. Finally, we are not interested in the number of molecules of each length but in their contribution to the DNA mass spectrum. The total DNA mass in a cell is given as $\mathscr{L} = \Sigma_{i=1}^{K} L_i$. Hence a molecule of length l contributes the fraction $l/\mathscr{L}$ to the total DNA mass of the genome. The fraction $q(l)$ of DNA mass represented by fragments of length l per genome is then

$$q(l) = \frac{l}{\mathscr{L}}\alpha^2(1 - \alpha)^{l-1}\sum_{i=1}^{K}\left((L_i - 1) - l + \frac{2}{\alpha}\right), \tag{3}$$

and the fraction $q(L_i)$ of DNA mass represented by chromosomes of length L_i that remained unbroken is

$$q(L_i) = \frac{L_i}{\mathcal{L}}(1 - \alpha)^{L_i-1}. \tag{4}$$

In our experimental conditions, the L_i's are of the order $10^6 bp = 1$ Mbp and α is of the order 1/Mbp for doses up to 1.000 Gray. The following approximations can thus be introduced: $L_i \pm 1 \approx L_i$, $l \pm 1 \approx l$ and $\ln(1 - \alpha) \approx -\alpha$. The DNA mass distribution $\tilde{m}(l)$ comprising fragments and unbroken chromosomal molecules can then be written as

$$\tilde{m}(l) = \frac{l}{\mathcal{L}} \, \alpha^2 e^{-\alpha l} \sum_{\substack{i=1 \\ l < L_i}}^{K} \left(L_i - l + \frac{2}{\alpha} \right) + \sum_{i=1}^{K} \frac{L_i}{\mathcal{L}} \, e^{-\alpha L_i} \delta_{l,L_i} \tag{5}$$

where δ_{l,L_i} is the Kronecker symbol ($\delta_{l,L_i} = 1$ for $l = L_i$, and 0 otherwise).

Using Eq. (5) one could, in principle, calculate the DNA mass profile after irradiation, but for computational effectiveness we chose a slightly different approach. Instead of calculating the differential mass distribution according to Eq. (5), we compute in a first step a sum distribution $\tilde{M}(l)$. The differential distribution is later derived by means of numerical differentiation. Integrating Eq. 5, one obtains for the sum distribution $\tilde{M}(l)$, *i.e.* the fraction of DNA mass that is contained in molecules of length less than l

$$\tilde{M}(l) = 1 - e^{-\alpha l} \sum_{\substack{i=1 \\ l < L_i}}^{K} \frac{L_i}{\mathcal{L}} \left(1 + \alpha l(1 - \frac{l}{L_i}) \right). \tag{6}$$

In the program PULSE $\tilde{M}(l)$ is computed on a finite grid covering the range $0 \leq l \leq L_{i,max}$ with $L_{i,max} = 2.2$ Mbp for *S. cerevisiae*. The number of grid points needs to be sufficiently high to obtain a resolution in the mass profile that is comparable to the resolution of the light intensity measurement. In practice we choose the number of grid points equal to the number of observed data points.

3.2 Determination of the relation between molecular length and electrophoretic migration distance

Observed DNA mass profiles depend on the mass distribution of the DNA in molecular length, l, and the migration distance, x, of molecules in a gel lane as a function of the molecular length. The relation between molecular length and migration distance in a gel can be substantially affected even by small deviations in gel run conditions (*e.g.* agarose concentration and buffer temperature), hence a calibration may need to be performed for every gel and each gel lane individually.

If the radiation doses are not excessively high, a sufficient fraction of each chromosome species remains intact, yielding distinct bands in a gel lane that are readily identified and can be used to define the calibration curve. The evaluation program PULSE reads the data of the fluorescence measurement from the input file and then displays the intensity profile. After background subtraction, the positions of peaks are marked interactively with

a cursor. This gives reference points for the determination of the calibration curve. The curve itself is then modeled in terms of a suitable interpolation procedure. As it turns out, the determination of the calibration function is crucial for the analysis.

Common interpolation schemes make use of polynomials to model a function, $x(l)$, between given reference points. The simplest option would be the representation of the calibration curve by a polygon that connects the reference points. Such a crude technique results, however, in artificial compression zones in the computed DNA mass profiles. These artifacts can be avoided if the calibration curve has continuous first and higher order derivatives, which is achieved by spline interpolation. In modeling the calibration curve, continuity through the second derivative appeared to be sufficient for our applications. Accordingly, cubic splines are used, which have the advantage of less artificial oscillations between the reference points than higher order splines or polynomials. Computation of a cubic spline interpolation requires the specification of two boundary conditions; to this purpose we equate the first derivatives at $x(L_1)$ and $x(L_K)$ with the slope to the adjacent reference points $x(L_2)$ and $x(L_{K-1})$, respectively.

Figure 2 shows a calibration curve for a gel lane loaded with DNA isolated from unirradiated cells. The gel was run on a CHEF system. The positions of identifiable bands in the intensity distribution that correspond to intact chromosomal molecules are represented by dots, the interpolated calibration curve is drawn as a solid line. In this case the resulting function is smooth, but the cubic spline can in some cases lead to strong oscillations in one or two intervals. Examining several generalized spline algorithms we found that exponential splines are the most suitable tool to avoid such oscillations. A parameter (p_i) is introduced for every distance interval, $[L_i, L_{i+1}]$, that describes the deviations of the exponential
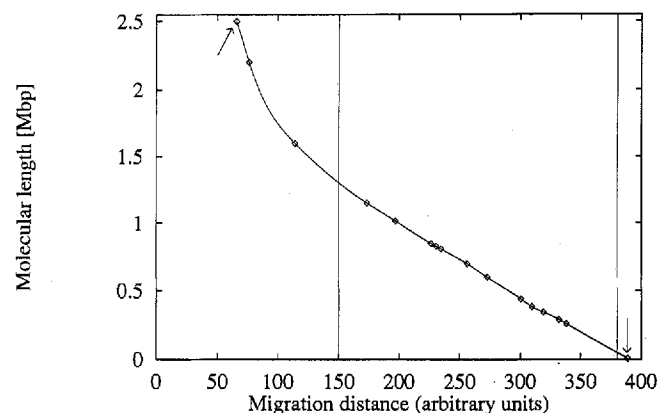

Figure 2. Calibration curve for a gel run on a CHEF system. The dots represent the observed bands corresponding to chromosomal molecules with known lengths. The calibration function obtained by interpolation with an exponential spline interpolation algorithm between these reference points is shown as a solid line. The points indicated by arrows were determined by extrapolation as described in the text. The vertical bars delimit the region of interest (see Section 3.5.2), where the comparison between observed and calculated profile is performed.

spline from a cubic spline. When the cubic spline shows oscillations in one interval, one can choose a value of p_i between 0 and ∞, where $p_i \to \infty$ leads to a cubic spline and $p_i = 0$ to a straight line. In practice, $p_i = 50$ is chosen by default for every interval, and the calibration curve is checked visually on the display. This value is sufficient to yield almost no deviations from a cubic spline.

With the default values for the p_i, for a large number of lanes calibration curves with no oscillations or artificial large curvatures can be established automatically. Only in rare cases will the default calibration curve need to be corrected manually. Nevertheless, an interactive check is necessary for every lane. In case an oscillation is seen in an interval, the corresponding p_i is decreased and set to the highest possible value that supresses the oscillations. Values for the calibration function are computed for every grid point, x, in the gel lane where intensity data were measured, and stored for further use.

3.3 Transformation of the DNA mass distribution in molecular length into a distribution in migration distance

The sum distribution in molecular length, $\tilde{M}(l)$, is calculated in terms of Eq. (6). It is then transformed into a sum distribution in migration distance, $M(x)$, using the information provided by the tabulated calibration function $x(l)$. Since one deals with sum distributions, the transformation is simply

$$M(x) = M(x(l)) = \tilde{M}(l). \tag{7}$$

A linear interpolation is sufficient in this calculation. From the sum distribution, $M(x)$, the differential distribution, $m(x)$, is obtained.

3.4 Consideration of band broadening

In the preceding treatment, molecules of identical length were assumed to have identical migration velocity. In fact, the migration distances vary around a mean value for molecules of identical length, yielding bands of finite width. This band broadening reflects various random factors in the preparation and migration processes, such as finite thickness of the plugs or any spatial variation in the system conditions (*e.g.* temperature, field or gel concentration gradients), and thermal diffusion, which was recently discussed in detail by Slater [23]. The theory of diffusion by Slater, which is based on the biased reptation model (for a review, see [24]) predicts nearly the same thermal diffusion constant, D, independent of molecular length in the size regime of yeast chromosomal molecules and for gel run conditions as described in Section 2.3. This leads to a constant band width along each gel lane. Our experimental data are in good agreement with these predictions, at least in the region of interest (see Section 3.5.2).

To account for band broadening, a smooth function of finite specified width is folded into the computed profile. The kernel should be a good model for the shape of peaks corresponding to bands which represent unbroken

molecules. We found that, in a comparison with the use of a Gaussian distribution, the Lorentz distribution

$$C(x_0, x) = \frac{\Gamma^2}{(2(x - x_0))^2 + \Gamma^2} \tag{8}$$

provided better fits to observed intensity profiles. Γ is here the full width at half peak value, x_0 the center of the Lorentz distribution. Folding of the DNA profile with a Lorentz distribution finally gives the calculated mass distribution $m_{calc}(x)$ in the gel lane,

$$m_{calc}(x) = \frac{\int_{x - \Delta x}^{x + \Delta x} m(x') \times C(x', x) dx'}{\int_{x - \Delta x}^{x + \Delta x} C(x', x) dx'} \tag{9}$$

In the computation it is found sufficient to extend the integration of Eq. 9 to the range $\Delta x \leq 9 \times \Gamma$, which contains at least 96% of the area under the Lorentz distribution.

In practice, Γ can vary for different lanes even in one gel; hence the appropriate value for the width of peaks in the intensity profile needs to be determined for every single lane. Peak widths can be varied interactively in order to obtain the best fit of the calculated to the observed profile. Figure 3 shows a comparison of calculated profiles obtained by folding with different kernels. In both panels the observed intensity profile is drawn as a solid line. The best fit profiles calculated with the Gauss and Lorentz distributions are shown as dashed lines. The upper panel compares the observed distribution with the distribution computed with a Gaussian distribution, whereas in the lower panel the Lorentz distri-
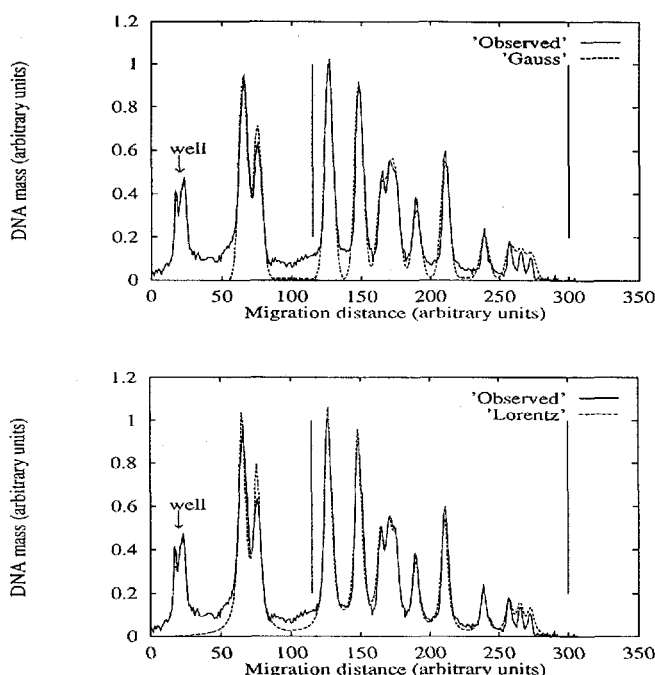

Figure 3. A comparison of calculated profiles computed by folding the DNA mass distribution in units of migration distance with a Gaussian (upper panel) and a Lorentz (lower panel) distribution. The observed profile (shown in each panel as a solid line) was obtained by separation of DNA isolated from unirradiated cells on a CHEF system. Profiles calculated with best fit values for breakage frequency and peak width are drawn as dashed lines.

bution was used. There is clear improvement in modeling the minima, but also adjacent maxima, such as the three maxima at migration distance ~ 170. In this unirradiated sample the breakage frequency was 0 breaks/Mbp and the peak width was adjusted to the optimal value in both panels.

3.5 Determination of the breakage frequency α

The breakage frequency α in an irradiation experiment can be determined through an optimization procedure, where the parameters of calculated profiles are first estimated and then adjusted until the best fit between the observed and computed profile is achieved.

3.5.1 Profile describing parameters

If the DSB are uniformly and independently distributed within the DNA, only one parameter, the breakage frequency α, is required to determine the resulting distribution of DNA fragments. The random distribution of breaks appears to hold for sparsely ionizing radiation. It is an open question whether deviations from a random distribution occur with higher linear energy transfer (LET) radiation such as neutrons, α-particles or heavier ions, or whether they will become apparent for a remaining fraction of DSB after repair.

A description of the distribution of DNA mass within a gel lane requires more than the mere computation of the length distribution of DNA molecules. Additionally, one has to account for the migration behavior of the molecules. Varying any of the electrophoresis conditions causes deviations in the migration behavior of DNA molecules. To account for such variations, a calibration curve needs to be determined for every particular gel lane. In addition, band broadening may vary somewhat, although we use a constant buffer temperature and plugs of identical thickness. Thus the parameter Γ needs to be fine-tuned for every lane to adjust for factors such as the pore size distribution within the gel, which may be non-uniform. The calibration of migration distance *vs.* molecular length accounts implicitly for a variety of parameters; it reduces the optimization problem to the adjustment of only two parameters, α and Γ. To obtain the optimum fit values of the two parameters, one needs to quantify the difference between the observed profile, $m_{obs}(x)$, and the calculated profile $m_{calc}(x; \alpha, \Gamma)$. This difference is expressed in terms of a least square deviation of either the differential DNA mass distribution, $m_{calc}(x)$, or of the sum distribution $M_{calc}(x) = \int_0^x m_{calc}(x')dx'$.

3.5.2 Region of interest

In practice, observed and calculated profiles are not compared over the full range of the observed profile, but only in a user-defined region of interest. This restriction is necessary, since, depending on the electrophoresis system in use and the running conditions, band inversion and/or self-trapping of molecules can occur [19]. Using hybridization assays we found that, in our experiments, migration velocity is a monotonous function of

molecular length for all chromosomes, except for the two longest chromosomes of *S. cerevisiae* with lengths of 2.2 Mbp and 1.6 Mbp. A clear-cut classification of bands and, therefore, a unique determination of the calibration curve is only possible in this monotonous region. Using PULSE, the region of interest is defined interactively on the displayed profile; it typically spans the region of molecular length from 1 kb to 1.2 Mbp.

3.5.3 Deviation functions

The mean square deviation

$$F_{diff}(\alpha, \Gamma) = \frac{1}{\langle m_{obs}\rangle} \sqrt{\frac{\int_{x_1}^{x_2} [m_{obs}(x) - m_{calc}(x; \alpha, \Gamma)]^2 dx}{x_2 - x_1}}$$
(10)

is a suitable measure for the deviation between observed and computed DNA profiles. $m_{obs}(x)$ is the observed DNA mass density in the gel lane, which is proportional to the measured fluorescence intensity, and $m_{calc}(x)$ is the calculated intensity. x_1 and x_2 are the migration distances at the borders of the region of interest. $\langle m_{obs}\rangle = \int_{x_1}^{x_2} m_{obs}(x)dx/(x_2 - x_1)$ is the mean observed intensity in the region of interest which is used for normalization. The use of relative, rather than absolute, deviations did not appear to provide an improved criterion for the fit, nor did we find improvements by using x-dependent weighting factors in the deviation function.

Comparing sum distributions of DNA mass in a gel lane is an alternative approach to quantify deviations between observed and calculated profiles. We choose as deviation function

$$F_{sum}(\alpha, \Gamma) = \frac{1}{M_{obs}(x_2)} \sqrt{\frac{\int_{x_1}^{x_2} [M_{obs}(x) - M_{calc}(x; \alpha, \Gamma)]^2 dx}{x_2 - x_1}}$$
(11)

The global minima of F_{diff} and F_{sum} are clearly apparent; they closely coincide in almost all cases in the two-dimensional space of the parameters α and Γ. A single deviation function would in fact be adequate, but we found that the minimum of F_{diff} is most exactly localized in the Γ-direction, yielding somewhat more precise information about band widths, whereas the breakage frequency α can be determined with best accuracy due to the good localization of the minimum of F_{sum} in the α-direction.

3.5.4 Determination of α

The deviation functions are calculated on a discrete grid of parameter values α and Γ, and the values for deviations are displayed on the terminal. Maximum and minimum values for α and Γ as well as the grid size are specified interactively. In practice, in a first step, the deviation functions are computed in a large domain of α and Γ on a coarse grid. During optimization, the grid is gradually narrowed around the global minimum of F. In this way, estimates for the breakage frequencies are determined to an accuracy of about $\pm$ 0.01/Mbp. Figure 4 shows a contour plot of F_{diff} computed for a (9 $\times$

9)-matrix of (α, Γ)-values in the vicinity of the global minimum of F_{diff}. A gel lane of a sample irradiated with ^{60}Co-γ-radiation at a dose of 500 Gy under hypoxic conditions was chosen as an example. The well-defined global minimum allows an immediate determination of the best fit values. $\alpha_o = 0.48$ breaks/Mbp and $\Gamma_o = 0.041$ Mbp.

Observed intensity profiles and the profiles computed with best fit parameter values are shown in Fig. 5. The observed profiles were obtained from a gel run on a CHEF system with running conditions as described in Section 2.3. DNA was isolated from cells (strain BK0) irradiated under hypoxic conditions at doses of 0 Gy, 250 Gy, 500 Gy and 750 Gy. The vertical bars indicate the region of interest within the gel lanes. Note the good fit between observed and calculated profiles in the region of interest, and the lack of agreement at smaller migration distances, which is the region of possible band inversion and self-trapping of DNA molecules. The deviation function shown in the contour plot (Fig. 4) corresponds to the lane containing DNA from cells irradiated at a dose of 500 Gy.

4 Discussion

We present here an approach to the quantitative analysis of radiation-induced DSB in yeast, which is based on a comparison of observed DNA mass distributions in agarose gels to calculated DNA mass distributions. Observed distributions are obtained from electrophoretic separation of genomic DNA isolated from irradiated and unirradiated cells; calculated distributions are determined assuming that DSB are randomly distributed within the genome, *i.e.* DSB occur independently and with equal probability throughout the DNA molecules. The close agreement between observed and calculated profiles indicates that the above assumption is valid for
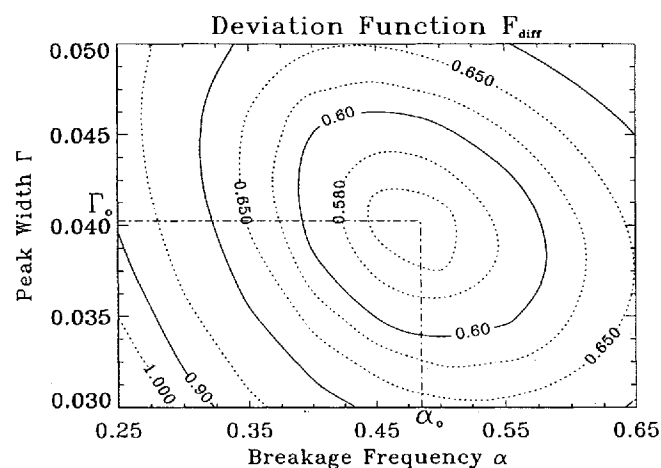

Figure 4. Contour plot of the deviation function $F_{\text{diff}}(\alpha, \Gamma)$ in the vicinity of its global minimum; the observed DNA mass profile in question was derived from a gel lane loaded with DNA isolated from cells of *S. cerevisiae* strain BK0, which were irradiated at a dose of 500 Gy under hypoxic conditions. Values for F_{diff} were calculated on a finite grid. F_{diff} reaches a minimum for assumed values for the breakage frequency $\alpha = 0.48$/Mbp and for the band width $\Gamma = 0.041$ Mbp.
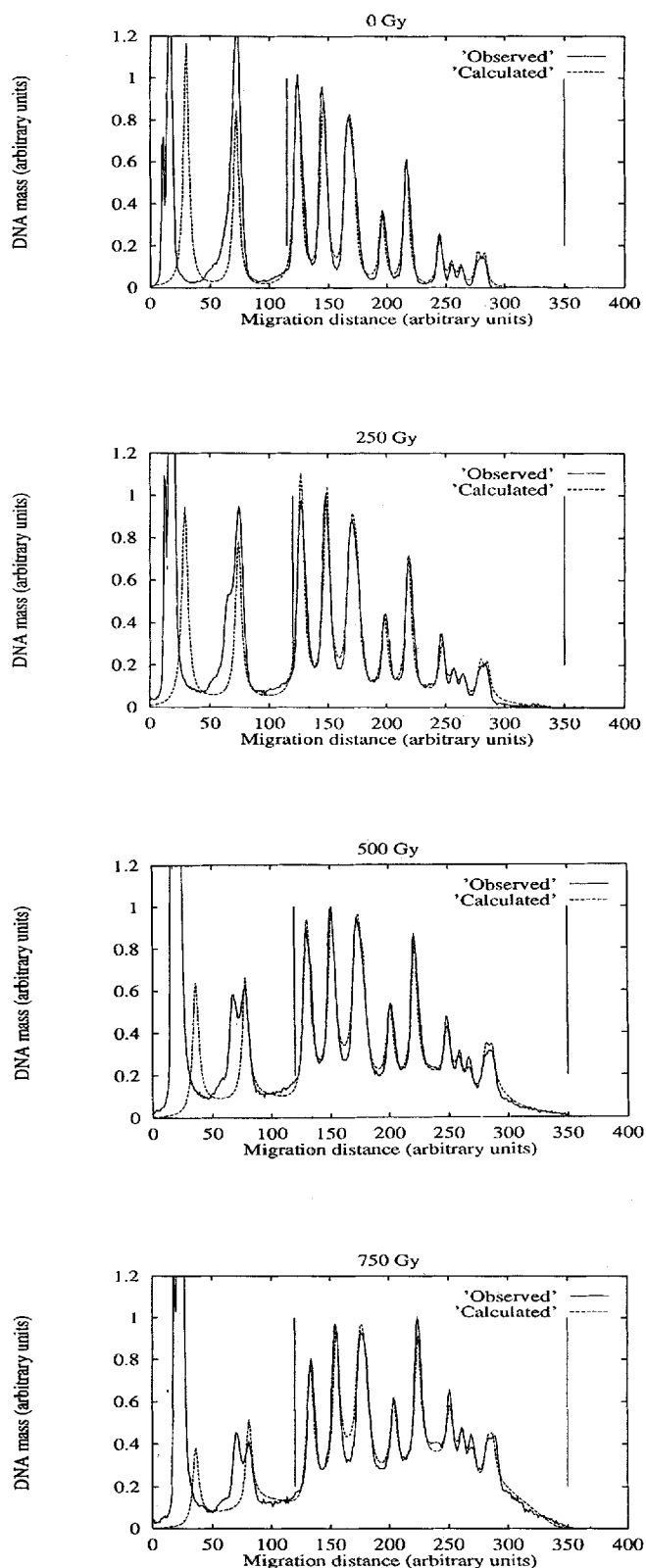

Figure 5. Observed and calculated DNA mass distributions. The observed profiles obtained by separation of DNA isolated from cells irradiated under hypoxic conditions (0, 250, 500 and 750 Gy, respectively) are represented as solid lines, whereas calculated profiles are shown as dashed lines. α and Γ were adjusted to the optimal values in every calculated profile. (0 Gy: $\alpha_o = 0$ breaks/Mbp, $\Gamma_o = 0.039$ Mbp; 250 Gy: $\alpha_o = 0.15$ breaks/Mbp, $\Gamma_o = 0.044$ Mbp; 500 Gy: $\alpha_o = 0.48$ breaks/Mbp, $\Gamma_o = 0.041$ Mbp; 750 Gy: $\alpha_o = 0.79$ breaks/Mbp, $\Gamma_o = 0.040$ Mbp).

DSB caused by sparsely ionizing radiation. It remains to be seen whether deviations from randomness will be found with densely ionizing radiations or after repair incubation. Investigations along these lines are in progress. If there are substantial deviations from randomness, they should be detectable with our evaluation procedure. DNA profiles obtained following partial restriction enzyme digest of yeast DNA could not be modeled adequately with the assumption of randomly distributed breaks (Friedl, unpublished results). In such cases more complex considerations will need to be introduced instead of the simple random breakage model that is now being used in the computer program PULSE.

Application of the described formalism requires information on the molecular lengths of unbroken chromosomes. The length of molecules up to 1 Mbp can be determined precisely by a calibration of migration distances vs. molecular length during a gel run using concatemeres of phage λ as a length standard, whereas the sizes of longer molecules need to be estimated. Recently it has been shown that the sizes of yeast chromosomal molecules can be inferred from the sizes of restriction fragments obtained by a digest from isolated yeast chromosomes using rarely cutting restriction enzymes [25]. However, this method is time-consuming and needs to be performed individually for each analyzed yeast strain, as yeast karyotypes show considerable chromosome length polymorphism [26].

Information on the lengths of the chromosomal molecules as well as on exact positions of bands within a gel lane is necessary for the determination of the calibration curve. In principle, it would be desirable to derive peak positions (i.e. migration distances of chromosomal molecules) automatically from the observed intensity profile. This may lead to better reproducibility than the presently used interactive procedure. However, it is at present impractical because of the poorly resolved peaks corresponding to molecules of different chromosome species with almost similar length. Furthermore, it would be difficult to provide for diploid strains with homologous chromosomes of distinguishable lengths (e.g. chromosome III of the strain BK0), a decision algorithm that classifies reliably homologous chromosomes of similar lengths and homologous chromosomes of different lengths. Work on these tasks is currently under way, but we feel that the interactive procedure for peak classification is at present most reliable.

The relation between migration distance and molecular length within one gel lane is established in terms of a spline interpolation between the chosen reference points. An exponential spline interpolation algorithm was found most suitable for this purpose. The number of reference points in the range from 50 kb to 1 Mbp can be increased by taking into account migration distances of λ-concatemeres which are run on the same gel on an adjacent lane. Nevertheless the interpolated calibration function may deviate slightly from the actual relation between molecular size and migration distance.

The modeling of DNA mass profiles with more than 1.5 DSB/Mbp is still not satisfying. With a high breakage

frequency, a large fraction of DNA mass is contained in fragments. As the determination of the distribution of fragments depends critically on the interpolation algorithm for the calibration curve, even minor deviations in the calibration from the actual relation cause large errors in the calculated profile. To minimize these problems, we recommend the use of electrophoresis conditions leading as closely as possible to linear relations between molecular size and migration velocity in the region of interest. Special attention needs to be paid to possible band inversions and zones where self-trapping of molecules may occur. We checked by hybridization experiments, that with our electrophoresis conditions, within the region of interest no inversions occur and that self-trapping is negligible.

The calculated DNA mass profiles depend on the following two parameters: the breakage frequency, α, and the width of observed bands, Γ. The shape of peaks in the observed profiles are in close agreement with Lorentz distributions; within one lane the parameter Γ does not depend on the migration distance. The Lorentz distribution is merely an approximation for the actual shape of peaks in an intensity profile. At present we are not able to give a mechanistic model for the evolution of band shapes during gel runs. Having tested several kinds of functions empirically we found that the use of Lorentz distributions yields best results. Nevertheless, further experimental and theoretical work is needed to attain an even more thorough description of band shapes to improve further the agreement of calculated with observed profiles.

The parameters α and Γ are optimized until the best fit between observed and computed profile is reached. The quality of the fit is quantified in terms of the deviation functions $F_{diff}(\alpha, \Gamma)$ and $F_{sum}(\alpha, \Gamma)$ that relate to the differential distributions and the sum distributions, respectively. The minimum of $F_{sum}(\alpha, \Gamma)$ is most suitable to

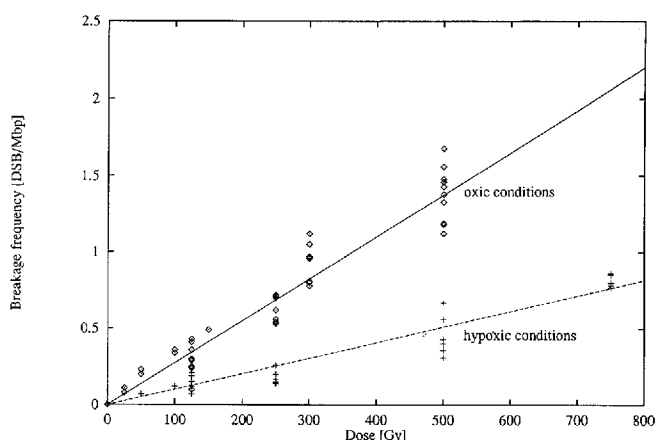Figure 6. Dose-effect relation for the irradiation of S. cerevisiae (strain BK0) with sparsely ionizing radiation (^{60}Co-γ-rays) under oxic (◇) and hypoxic (+) conditions. Indicated are values for α obtained in seven independent irradiation experiments evaluating several gels for every experiment. The relation between dose and breakage frequency is linear for irradiation under oxic as well as under hypoxic conditions. The dose-effect relation found by linear regression is drawn as a solid line with a slope of 2.75×10^{-9} breaks bp^{-1} Gy^{-1} for oxic conditions and as a dashed line with a slope of 1.02×10^{-9} breaks bp^{-1} Gy^{-1} for hypoxic conditions.

determine the optimum value for the breakage frequency α, and $F_{\text{diff}}(\alpha, \Gamma)$ defines with best precision the peak width Γ. In most cases, however, the minima of the two deviation functions coincide.

The calculated DNA profiles for the optimal values α and Γ agree well with the experimentally derived profiles. To obtain best results, attention needs to be paid to the preparation of DNA. In most of our experiments we infer a value of less than 0.01 DSB/Mbp in the unirradiated control samples. Figure 6 shows a plot of breakage frequencies derived with our method after exposure of cells of the strain BK0 (*S. cerevisiae*) to different doses under oxic and hypoxic conditions. Values of α derived from seven independent irradiation experiments with different doses are shown. For every experiment several gels were evaluated. The solid line represents the dose-effect relation for irradiation under oxic conditions; the dashed line the dose-effect relation for irradiation under hypoxic conditions. Both lines were obtained by a linear regression algorithm. A breakage frequency of $(1.02 \pm 0.06) \times 10^{-9}$ breaks bp^{-1} Gy^{-1} was obtained for irradiation under hypoxic conditions and $(2.75 \pm 0.11) \times 10^{-9}$ breaks bp^{-1} Gy^{-1} for irradiation under oxic conditions, which implies an oxygen enhancement ratio (OER) of 2.7. The results from experiments under hypoxic conditions are in good agreement with recently published data obtained with a preliminary PULSE version and densitometric scanning of gel photonegatives [13]. However, the DSB frequencies after λ-irradiation under hypoxic and oxic conditions that are determined with the presented approach are about 20% lower than the DSB yields that are determined in neutral gradient sedimentation assays [27]. At present this difference remains unexplained.

Note that clusters of DSB separated by only a few bp cannot be detected with any of the methods currently in use for the analysis of DSB. It is expected that clustered DSB occur especially in the case of densely ionizing radiation, but it cannot be ruled out that they are induced as well by sparsely ionizing radiation. Hence, it is possible that the actual value for the breakage frequency is somewhat larger than determined with our method.

Our evaluation procedure is, in principle, applicable for all PFGE assays. It was tested for the CHEF, TAFE and ZIFE systems and can be used immediately for the evaluation of gels obtained with any system that yields straight lanes. If PFGE assays yield distorted lanes in "smiling" or "frowning" gels, the process of data collection will need to be modified, but the evaluation program requires no major changes in such cases.

The presented approach is a convenient method for the quantitative analysis of DSB induction in organisms whose chromosomes can be separated by PFGE. Adaption of the evaluation procedure for the quantitation of DSB in mammalian cells is currently in progress. The approach may also bring further insight into the processes of damage induction, as it permits the test of various breakage models, and of assumptions underlying these models.

The program PULSE is available on request. Along with a stand-alone executable file, we provide a short manual, data files allowing a sample program run (as a tutorial), and some plots to check the results obtained with the run.

5 References

[1] Ward, J. F., *Int. J. Radiat. Biol.* 1990, *57*, 1141–1150.
[2] Averbeck, D., Dardalhon, M. and Magana-Schwencke, N., *J. Photochem. Photobiol. B* 1990, *6*, 221–236.
[3] Elia, M. C., DeLuca, J. G. and Bradley, M. O., *Pharmac. Ther.* 1991, *51*, 291–327.
[4] Sutherland, J. C., Chen, C. Z., Emrick, A., Hacham, H., Monteleone, D. C., Ribeiro, E., Trunk, J., Sutherland, B. M., in: Sutherland, B. M. and Woodhead, A. D. (Eds.), *DNA Damage and Repair in Human Tissues*, Plenum Press, New York 1990, pp. 45–61.
[5] Schwartz, D. C. and Cantor, C. R., *Cell* 1984, *37*, 67–75.
[6] Contopoulou, R., Cook, V. E. and Mortimer, R. K., *Yeast* 1987, *3*, 71–76.
[7] Geigl, E. M., Eckardt-Schupp, F. and Hagen, U., *Yeast* 1986, *2*, 126.
[8] Geigl, E. M. and Eckardt-Schupp, F., *Molec. Microbiol.* 1990, *4*, 801–810.
[9] Blöcher, D., Einspenner, M. and Zajackowski, J., *Int. J. Radiat. Biol.* 1989, *56*, 437–448.
[10] Stamato, T. D. and Denko, N., *Radiat. Research* 1990, *121*, 196–205.
[11] Ahn, S. Y., Nevaldine, B. and Hahn, P. J., *Int. J. Radiat. Biol.* 1991, *59*, 661–675.
[12] Iliakis, G. E., Metzger, L., Denko, N. and Stamato, T. D., *Int. J. Radiat. Biol.* 1991, *59*, 321–341.
[13] Friedl, A. A., Beisker, W., Hahn, K., Eckardt-Schupp, F. and Kellerer, A. M., *Int. J. Radiat. Biol.* 1993, *63*, 173–181.
[14] Gardiner, K., Laas, W. and Patterson, D., *Somatic Cell Mol. Genet.* 1986, *12*, 185–195.
[15] Chu, G., Vollrath, D. and Davis, R. W., *Science* 1986, *234*, 1582–1587.
[16] Noolandi, J. and Turmel, C., in: Burmeister, M. and Ulanovsky, L. (Eds.), *Pulsed-field Gel Electrophoresis*, The Humana Press, Totowa, NJ 1992, pp. 451–467.
[17] Blöcher, D., *Int. J. Radiat. Biol.* 1982, *42*, 314–328.
[18] Ribeiro, E. A., Larcom, L. L. and Miller, D. P., *Anal. Biochem.* 1989, *181*, 197–208.
[19] Gemmill, R. M., *Adv. Electrophoresis* 1991, *4*, 1–48.
[20] Kunz, B. A. and Haynes, R. H., *Mutation Res.* 1982, *93*, 353–375.
[21] Schulz, G. V., *Z. physikal. Chem. (B)* 1942, *51*, 127–141.
[22] Sakurada, I. and Okamura, S., *Z. physikal. Chem. (A)* 1940, *187*, 289–296.
[23] Slater, G. W., *Electrophoresis* 1993, *14*, 1–7.
[24] Noolandi, J., *Adv. Electrophoresis* 1992, *5*, 1–57.
[25] Link, A. J. and Olsen, M. V., *Genetics* 1991, *127*, 681–698.
[26] Carle, G. F. and Olson, M. V., *Proc. Natl. Acad. Sci. USA* 1985, *82*, 3756–3760.
[27] Frankenberg-Schwager, M., Frankenberg, D., Blöcher, D. and Adamczyk, C., *Int. J. Radiat. Biol.* 1979, *88*, 261–270.