(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097135 A1

(51) International Patent Classification[7]: C12Q 1/68, C12P 19/34, C12N 9/12, C07H 21/04

(21) International Application Number: PCT/US02/16967

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/294,890 30 May 2001 (30.05.2001) US

(71) Applicants and
(72) Inventors: **STAMATOYANNOPOULOS, John, A.** [US/US]; 15 Milford Street, #1, Boston, MA 02118 (US). **DORSCHNER, Michael, O.** [US/US]; 590 Galer Street, #415, Seattle, Washington 98109 (US). **MCARTHUR, Michael** [GB/GB]; Mill House; Mill Lane, Attleborough; Norfolk, NR17, Rockland All Saints (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **STAMATOYANNOPOULOUS, Georges** [US/US]; 702 35th Avenue, Seattle, Washington 98122 (US).

(74) Agents: **REMENICK, James** et al.; Intellectual Property Department, Heller Ehrman White & McAuliffe LLP, 101 Orchard Ridge Drive, Suite 300, Gaithersburg, MD 20787-1917 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/097135 A1

(54) Title: ACCURATE AND EFFICIENT QUANTIFICATION OF DNA SENSITIVITY BY REAL-TIME PCR

(57) Abstract: The sensitivity of regions of genomic DNA from chromatin to DNA-modifying agents are quantitated using real time PCR with careful use of reference regions and standard curves for improved accuracy over previous methods. A wide variety of DNA modifying agents are used to modify DNA, including single strand breakers, which allows analysis of a wider variety of DNA structures with greater precision than was previously possible. Methods also are provided that allow fine detailed analysis of hypersensitive DNA regions from as little as 100 cells. These methods also potentiate the investigation of more fine details of chromatin fiber folding and nucleosome-nucleosome contact.

Accurate and Efficient Quantification of DNA Sensitivity by Real-Time PCR

Reference to Related Applications

This application claims priority to U.S. Provisional Application No. 60/294,890 entitled "Accurate Quantification of DNaseI Sensitivity by Real-Time PCR" filed May 30, 2001.

Field of the Invention

The invention relates generally to methods of DNA analysis and more specifically to methods for analysis of genomic sequences having desirable biochemical activity.

Background of the Invention

It has been long appreciated that the sensitivity of chromatin to degradation by enzymes, chemicals or ultra violet light is a way of mapping chromatin structure in vivo. The most widespread used method is to digest nuclei with the non-specific endonuclease DNaseI. Combination of such digestions with Southern hybridizations have led to the general classification of DNaseI-insensitive, -sensitive and hypersensitive sites. Such hypersensitive sites are structurally and functionally interesting. Functionally they often coincide with the regulatory regions of proximal genes, making their accurate characterization important. Structurally they appear to be diverse. Laboratories have analyzed specific hypersensitive sites and described them as being nucleosome-free, bound co-operatively with transcription factors or containing a modified nucleosome. It is evident that a more accurate method is needed to describe DNaseI-sensitivity which will allow the detection and characterization of the myriad forms of structures which must underlie the transcriptionally active chromatin fiber.

Other limitations of the conventional use of Southern hybridization to assay DNaseI-sensitivity is that the method is largely technical. The assay requires up to 50 micrograms of genomic DNA to achieve a detectable signal following digestion, separation and blotting. Real-time PCR can routinely use ten thousand times less. Secondly, the detection of a hypersensitive site relies on the presence of a sub-band (which is generated by a cutting site for DNaseI and for the restriction enzyme used to digest the purified genomic DNA). These can be hard to detect if the DNaseI sites are disperse. Quantification of the signal between this sub-band and the

parental band (generated by two cutting sites for the restriction enzyme) can be compromised by several variables: there may be a different transfer efficiency in blotting the parental and smaller-sub band; disperse cutting will form a smeary sub-band; the efficiency of hybridization of the probe to the two bands may differ; analysis may become confused if more than one hypersensitive site exists on the same parental band.

Another problem with Southern hybridization assay is that this generally allows only analysis of relatively large segments of material (the size of the parental band) and mapping of the position of novel DNaseI-hypersensitive site depends upon the resolution of the gel. The PCR approach allows sequences to be amplified which can be tested directly as to whether or not they are hypersensitive to digestion. Also the distribution of cutting can be established by designing primer pairs amplifying adjacent, or overlapping, sequences. Structural information can be gathered for the first time about the status of sequences proximal to the hypersensitive sites and the effect of their presence to the local chromatin structure.

Semi-quantitative PCR as for example described by Kramer et al. (Bio Techniques 22: 879-882, 1997; and Gregory et al. (Nucleic Acids Res. 27: e32, 1999; see also Kramer et al., Development 125:4749-4755, 1998) has features that address these problems but generally suffers from background problems that make quantitation difficult and/or are not applicable to whole genome analyses. In general these previous procedures were developed to specifically study narrow aspects of hypersensitive sites and are very inaccurate. These procedures generally cannot provide copy number information for opened regions of nucleosomes and more accurate methods are needed.

Summary of the Invention

The present invention overcomes the problems and disadvantages associated with current strategies and designs and provides methods for accurately and efficiently analyzing a region of a genome.

An embodiment of the invention is a method for accurately and efficiently determining sensitivity of a candidate region of a genome to a DNA modifying agent comprising isolating chromatin from a population of eukaryotic cells containing said genome; treating at least one portion of said isolated chromatin with said DNA

modifying agent under conditions to cause DNA strand breakage; treating another portion with said DNA modifying agent under modified conditions; isolating treated DNA from the portions; amplifying the candidate region from isolated DNA by real-time PCR from each portion by real time PCR with a set of primers and obtaining a signal; determining a relative copy number of said candidate region within each isolated DNA portion by: determining the copy number of said candidate region in each of a plurality of DNA samples each containing a different amount of DNA in a fixed proportion with respect to each other and thereby calculating a first standard copy number curve; determining the copy number of a reference region in each of a plurality of DNA samples each containing a different amount of DNA in a fixed proportion with respect to each other and thereby calculating a second standard copy number curve from PCR amplification of said reference region; and comparing the signal to said first standard curve and said second standard curve to obtain the relative copy number ratio of the candidate region and the reference region for each isolated DNA; and determining sensitivity of said candidate region to said DNA modifying agent relative to the sensitivity of said reference region.

In another embodiment of the invention the genome is a human genome. In yet another embodiment the candidate sequence is less than 250 base pairs in length. In yet another embodiment the candidate region is between about 50 and about 2,000 base pairs in length. In yet another embodiment one portion of isolated chromatin comprises from 0.15 pg to about 5 ug of nucleic acid. In yet another embodiment the method of claim 1, wherein the one portion of isolated chromatin comprises one copy of said genome.

In another embodiment the conditions and the modified conditions are selected from the group consisting of different concentrations of the DNA modifying agent (e.g. pM, nM, mM), different times (e.g. seconds, minutes, hours), different temperatures (e.g. 15°C, 20°C, 25°C, 30°C, 35°C, 40°C, 45°C, 50°C, 55°C, 60°C, 65°C, 70°C), different buffer conditions (e.g. high to low pH, ion concentration or buffering capacity), different concentrations of chromatin (e.g. one copy of genome to 5 ug), and combinations thereof. In another embodiment the conditions and the modified conditions are performed at the same temperature (e.g. 20°C, 30°C, 37°C, 50°C). In another embodiment amplifying each isolated DNA by the real-time PCR

proceeds through at least eighteen cycles. In another embodiment amplifying each isolated DNA by the real-time PCR proceeds through at least thirty-five cycles. In another embodiment the set of primers amplify fragments of DNA that are from about 200 to 400 base pairs in length. In another embodiment the reference region is insensitive to the DNA modifying agent. In another embodiment the reference region is more sensitive to the DNA modifying agent. In another embodiment the reference region is less sensitive to the DNA modifying agent. In another embodiment the reference sequence is less than about 250 base pairs in length. In another embodiment the reference region is between about 50 and about 2,000 base pairs in length. In another embodiment the reference region is between 200 and 400 base pairs in length. In another embodiment a double strand DNA specific marker is used during PCR to detect the accumulation of PCR products as a function of the number of rounds of amplification. In another embodiment the DNA modifying agent is DNase I and the DNA modifying reactions utilize a single reaction time point and differing concentrations of DNase I. In another embodiment the DNA modifying agent introduces single stranded nicks into the DNA. In another embodiment the sensitivity of a genome of a eukaryotic species comprising: combining cells of a first eukaryotic species with a first genome with cells from a second eukaryotic species with a second genome having a combined cell population totaling at least about 10^8 cells, about 10^9 cells, or about 10^{10} cells; isolating chromatin from the combined cell population and dividing the isolated chromatin into one or more portions; treating one portion or sub-portion of chromatin with a first amount of said DNA modifying agent; isolating treated DNA from the portions; amplifying each isolated DNA by real time PCR with a set of primers that amplify a template sequence of the first eukaryotic species; and determining the sensitivity of said template sequence to the DNA modifying agent within the first genome.

In another embodiment the cells of the first eukaryotic species number less than 10^7. In another embodiment the cells of the first eukaryotic species comprise less than about 10^6 cells. In another embodiment the DNA modifying agent introduces single stranded nicks into the DNA.

Other embodiments and advantages of the invention are set forth, in part, in the following description and, in part, may be obvious from this description, or may be learned from the practice of the invention.

Description of the Figures

Figure 1 shows a map of the LCR of the murine ß-globin locus. DNaseI-hypersensitive sites are shown by vertical arrows, the estimated strength of which are represented by their relative thickness arid length. The position of restriction sites for HpaI are marked as a vertical line topped with an H. The names and positions of the amplicons that the Real-time primers are designed to amplify are shown. The letters 'HS' designate the hypersensitive sites and 'Fk' the Flanking sequences.

Figure 2 shows the DNaseI-sensitivity of NF-M and HS2 in mouse fetal liver. Chromatograms were generated on the Lightcycler system (Roche) following Real-time quantitative PCR experiments performed on a series of genomic DNAs isolated from DNaseI-digested fetal liver nuclei amplifying (a), Nf-M and (b), HS2. Using the LightCycler FastStart DNA Master SYBR green I mix at 1x final concentration (containing PCR buffer, dNTPs, $MgCl_2$ and *Taq* polymerase) the QCPR reactions were assembled as follows: 0.3uM of each primer, additional $MgCl_2$ to 3mM and 5-30ng of template DNA. Reactions are thermal-cycled under two sets of conditions depending on amplicon size; 250 bp or 500 bp. For 250 bp amplicons the following amplification protocol is used: 95°C for 10 minutes, followed by 40 cycles of 95°C; 5 sec, 60°C; 5 sec and 72°C; 15 sec. For 500 bp amplicons the 72°C phase is extended to 25 sec. Both amplification protocols were followed by a melting curve analysis.

Each panel shows the progress of the PCR reactions plotted as a gain in fluorescence (on a logarithmic scale), due to the binding of the dsDNA-specific SYBR-green dye to the PCR product, as a function of the number of. cycles. Analysis, by the manufacturer's software, extrapolates from that part of the curve (using the gray crosses) where amplification is exponential to determine the number of cycles of PCR after which the fluorescent signal reached a threshold value (green line). The position of the intercepts from the standard curve samples are labeled with the amount of template used (ng), and the standard curves generated are plotted

in Figure 3c. The other intercepts are from DNaseI-digested samples, and the units of DNaseI used in the digestion labeled in panel b. It is evident that there was no digestion of the Nf-M amplicon as these intercepts are clustered around that of 50 ng (a). The intercepts for HS2 (b) are more widely distributed between that of 50 and 5. Both experiments contained a negative control (amplification with no template) and a signal is gained from this reaction in panel b, marked H_2O. Analysis confirmed that this signal was due to primer-dimers formed in the absence of template DNA (Figure 3a and b). The calculated DNaseI-digestion profile of HS2 is shown in Figure 4.

Figure 3 shows an analysis of the specificity and sensitivity of Real-time PCR with SYBR-green. (a), Melting curve analysis of the PCR products generated in Figure 2. Following the final step of amplification the samples were melted by slowly increasing the temperature to 95°C and continually monitoring the fluorescent signal. The rate of change of fluorescence is shown as a function of temperature. Two major peaks occur, centered at 84°C and 88°C due to the melting of the HS2 and Nf-M PCR products respectively. The smaller peak at 79°C, marked H_20, is derived from the negative control run in Figure 3b. These results are consistent with the primer pairs amplifying a single product from template DNA and the HS2 pair less stable primer-dimers in the absence of template. (b), Agarose gel electrophoresis of the PCR products of Nf-M, HS2 and the negative control. Single bands of the same size (450 bp) are evident for the template-dependent PCRs whereas for the negative control no single band is seen. (c), Standard curves generated from Figure 2 for (□) HS2 and (♦) Nf M. The data generated is linear over a range of 0.5 to 50 ng of template.

Figure 4 displays a plot of the DNaseI-digestion of fetal liver HS2. The standard curves in Figure 3c were used to calculate the percentage of copies of the HS2 amplicon remaining in 50 ng of DNaseI-treated genomic DNA. The amount of template DNA was standardized by correcting for amplification of the DNaseI-insensitive Nf-M sequence.

Figure 5 displays plots of the DNaseI-digestion profiles of amplicons throughout the mouse ß-globin LCR. The plots were generated as described in the legend to Figure 4. The primer pairs used were those shown in Figure 1; (a), Nf-M

and the DNaseI-hypersensitive sites HSI to HS4 and (b), the Flanking sequences Fk1 to Fk3 and the 'weak' HS6.

Figure 6 displays a plot of all the DNaseI-digestion profiles generated in this study. Four distinct classes of profiles are evident and are labeled as described in the text.

Figure 7 displays a schematic diagram to account for how DNaseI probes accessibility. The consequence of competition between DNaseI (shown as a solid circle) and a nucleosome (a hollow oval) for the extent of restriction of a molecule of duplex DNA in the case of (a), naked DNA, (b), partially bound and (c) covered DNA.

Description of the Invention

As embodied and broadly described herein, the present invention is directed to methods for accurately and efficiently analyzing a region of a genome.

Real-time quantitative PCR was performed as described in McArthur et al. (J. Mol. Biol. 313: 27-34, 2001) the content of which is hereby incorporated by reference in its entirety. Briefly, the Lightcycler system from Roche Molecular Biochemicals was used and all reactions were performed using the manufacturer's FastStart DNA SYBR-green master kit. Reaction conditions were optimized for standard oilgonucleotide pairs by varying the concentration of magnesium used in the reaction, as described in the manufacturer's protocol. The reactions were performed in a volume of 10 microliters with 0.5, 5, and 50 ng of genomic DNAs isolated from tissue culture cells in order to prepare standard curves for each of the amplicons, or 50 ng of sample from each DNaseI-digestion point. The DNaseI-treated genomic DNAs were prepared by the standard methodology. The number of copies of target remaining intact, corrected for the size of the fragment, was calculated by reference to the standard curve. These data were then plotted versus the units of enzyme used to digest the nuclei.

Embodiments of the invention include several desirable features that further alleviate problems in the field as cited above. A first particularly desirable embodiment is a method for analyzing chromatin structure from limiting amounts of tissue. This embodiment allows the use of as little as 10^5 cells for an analysis. The cells of interest can be mixed in a larger population of cells from a different species.

This material can be used to generate the DNaseI-digestion series of genomic DNAs used in the analysis. It is preferred to use up to 10^8 cells in each DNaseI-digestion. Otherwise estimation of the concentration of the isolated nuclei, which is preferred to calculate the amount of DNaseI to be added, becomes inaccurate. The primer pairs used in the quantitative Real-time PCR will be tested to show that they are species specific (and give no amplification from the DNA of cells used as the 'carrier' population). It will then be possible to proceed with analysis using an estimate of the amount of target DNA, present. Use of a DNaseI-insensitive control, as described in section 3, will accurately establish the amount of genomic DNA of interest present in each reaction, thus allowing an accurate determination of the kinetics of DNaseI-digestion of the amplicon of interest.

A second particularly desirable embodiment allows the detection of single-stranded nicks that have been introduced into chromatin sequences. The method need not just be applied to genomic DNAs isolated from nuclei treated with DNaseI but can use any DNA-modifying agent. Whereas Southern hybridization fails to detect single-stranded nicks introduced (as the DNA molecules migrate as duplexes) the PCR-based assay detects the cut as it destroys one half of the template. Several sensitive DNA modifying agents, such as hydrogen peroxide, which has the advantage over DNaseI of having no sequence specificity of cutting and being entirely soluble, allows much higher resolution in its cutting pattern and introduces single stranded nicks only.

A wide variety of DNA cutting agents and methods for cutting DNA are available and specifically contemplated for embodiments of the invention. For example, epigenetic modifications in chromatin such as histone acetylation and cytosine methylation may be used. Further optional treatments include contact with one or more of the following DNA-modifying agents or conditions: nucleases (both sequence-specific and non-specific); topoisomerases; methylases; acetylases; chemicals; pharmaceuticals (e.g., chemotherapy agents); radiation; physical shearing; nutrient deprivation (e.g., folate deprivation), and other agents that are commercially available and known to those of ordinary skill.

Other molecules which play a large role in regulation of genomic processes, including proteins and RNAs, which control the structure of the nucleus are being identified, and these also could be used as targets for modifying DNA. Still further, proteins that bind to a given DNA sequence or set of sequences may be employed to induce DNA modification such as strand breakage. Proteins can either be modified by many means, such as incorporation of ^{125}I, the radioactive decay of which would cause strand breakage (e.g., Acta Oncol. 39: 681-685 (2000)), or modifying cross-linking reagents such as 4-azidophenacylbromide (e.g., Proc. Natl. Acad. Sci. USA 89: 10287-10291) which form a cross-link with DNA on exposure to UV-light. Such protein-DNA cross-links can subsequently be converted to a double-stranded DNA break by treatment with piperidine.

Yet another approach to DNA modification relies on antibodies raised against specific proteins bound at one or more DNA sites, such as transcription factors or architectural chromatin proteins, and used to isolate the DNA from nucleoprotein complexes. An example of a currently used technique cross-links proteins and DNA within the eukaryotic genome following treatment with formaldehyde, for example, and after isolation of the chromatin and following either sonication or digestion with nucleases the sequences of interest are immunoprecipitated (Orlando *et al.* Methods 11: 205-214, 1997).

Another modification is cytosine methylation. The global pattern of methylation is relatively stable but certain genetic control regions become methylated if they are silenced or conversely demethylated if activated. Differential methylation can be detected by use of pairs of restriction endonucleases that cut the same site differently according to whether or not it is methylated (Tompa *et al.* Curr. Biol. 12: 65-68, 2002). Alternatively it is possible to distinguish generically between a methylated and non-methylated cytosine by genomic sequencing (a methodology developed by Pfeifer *et al.* Science 246: 810-813, 1989) that converts cytosines to uracil, which behaves similarly to thymine in sequencing reactions, and leaves methyl-cytosine unmodified. This material can be used as a template in PCR with primers sensitive to the C to U transition. Alternatively the potential mismatch

(G:U) between oligonucleotide and template can be cleaved by *E.coli* Mismatch Uracil DNA Glycosylase, and that fragment removed from the population.

A further approach is directed to the enzymatic machinery which gives rise to or maintains the epigenetic patterns. This machinery can also be labeled as described above so that it can be induced to cause detectable DNA modifications such as double stranded DNA breaks. Target proteins for this kind of approach would include the recently described HATs (Histone-Acetyl Transferases), HDACs (Distone De-Acetylase Complexes) whose effect on transcriptional induction has been recently described (Cell 108: 475-487, 2002), as well as DNA methyltransferases and structural proteins that bind to the sites of methylation, such as MeCP1 and MeCP2. Histones, and transcription factors, are also known to become methylated, phosphorylated and ubiquitylated. A range of covalent modifications, some of which have yet to be described, may be made to the structural and enzymatic machinery of transcription, replication and recombination. Our current understanding suggests that such modifications have a regulatory role and it has been demonstrated that these modifications can be positively and negatively correlated with the functional activity of the underlying sequence (Science 293: 1150-1155). The potential for combinations of modifications of the sequences overlays another layer of complexity of regulation on the underlying genome and it is possible to dynamically follow these epigenetic changes with immunoprecipitation of the DNA sequences from *in vivo* nucleoprotein complexes.

Yet another approach to selection and modification of DNA is suggested by the description of cytologically distinct regions of interphase nuclei such as the nucleoli which contain the heavily transcribed rRNA genes (Proc. Natl. Acad. Sci. USA 69: 3394-3398, 1972) and active genes may be preferentially associated with clusters of interchromatin granules (J. Cell Biol. 131: 1635-1647, 1995). Specific regulatory regions may become localized to distinct areas within the nucleus on transcriptional induction (Proc. Natl. Acad. Sci. USA 98: 12120-12125, 2001). By contrast specific areas of eukaryotic nuclei have been shown to be transcriptionally inert (Nature 381: 529-531, 1996) and associated with heterochromatin.

Fractionation of the nucleus on the basis of such and similar physical properties can be used to selectively cleave DNA.

A third particularly desirable embodiment allows the quantitative analysis of naturally occurring single-stranded DNA structures in vivo. The presence of single-stranded DNA in nuclei is unusual and can be caused by the action of enzymes, such as topoisomerase I, as a transient consequence of DNA replication or from formation of unusual DNA structures (such as Z-DNA or triplex DNA). The kinetics of digestion of single strand-specific cutters, such as potassium permanganate, or primer-directed restriction can be effectively monitored using the assay described above. For example, the formation of triplex DNA structures is of particular interest as these are believed to be involved in regulation of downstream genes and it would be of interest to establish if there were a correlation between their formation and expression.

Real-time PCR was used to allow quantification of the sensitivity of chromatin to digestion by DNaseI. This approach has three clear advantages to the more conventional use of the Southern hybridization assay: the accuracy of quantification is improved; the resolution of the assay is enhanced- by designing primers to amplify small amplicons so that it is possible to analyze sequences both eo-incident and proximal to sites of DNaseI-hypersensitivity; less material is needed, as little as 5 ng of treated genomic DNA. This method was applied in an analysis of the chromatin structure of the previously described mouse ß-globin locus control region (LCR) using fetal liver cells. The four hypersensitive sites of the canonical mouse LCR, HSI to HS4, are shown to have kinetics of digestion consistent with these sequences being nucleosome-free in vivo. A different pattern was seen for HS6 (a recently described 'weak' hypersensitive site). The site was also rapidly lost but more of the sites proved resistant, consistent with only a portion of HS6 being nucleosome-free. This finding implies that in vivo the LCR is structurally heterogeneous. Sequences proximal to the hypersensitive sites show a third pattern of intermediate sensitivity, consistent with the chromatin being unfolded but the sites still bound by a continual nucleosomal array. The results demonstrate that this method has the potential to achieve accurate and detailed mapping of chromatin structure from small amounts of tissue samples.

DNaseI-sensitivity assays have been used to demonstrate that chromatin structure can be a dynamic determinant of the transcriptional program (1). The bulk of the eukaryotic genomes are transcriptionally inert and occluded into tightly condensed chromatin that is insensitive to digestion by nucleases (2). Genes which are regulated in a tissue- or developmental-specific fashion, such as the murine proto-oncogenes c-fos and c-myc, undergo changes in general DNaseI-sensitivity that coincide with their transcriptional induction (3,4). Some sequences are extremely sensitive to digestion and are known as DNaseI-hypersensitive sites. The majority of these are gaps in the nucleosomal array, which leave 200- 400 bp of the underlying sequence vulnerable to digestion by DNasel and their cognate restriction enzymes (5). Recent studies have suggested that some DNaseI--hypersensitive sites, such as that of the human epsilon-globin promoter, may still be bound by a nucleosome but with an altered structure (6). The disruption or modulation of nucleosomes at these sites is caused by the co-operative binding of ubiquitous and tissue-specific transcription factors (7,8). The high density binding of transcription factors makes it likely that such sequences will be cis-acting regulatory elements, such as promoters and enhancers.

Clusters of hypersensitive sites can also form super-regulatory elements known as locus control regions (LCRs), which are capable of both regulating chromatin structure over large distances and enhancing transcription of a family of genes. To date twenty-five LCRs have been identified in humans (9) and for several of these homologues have been found in other eukaryotes. The best studied of these is that associated with the ß-globin locus which has, amongst others, a murine homologue. The human version consists of four tissue-specific DNasel-hypersensitive sites, HSI to HS4, within a 25 kb region upstream of the five ß-globin-like genes (10,11,12). Proof that the LCR is needed for directing transcription comes from two sources: naturally occurring deletions which remove the majority of the LCR, such as that which gives rise to the Hispanic $(\gamma\delta ß)^{o}$-thalassaemia syndrome, lead to both a failure to express the intact genes and a loss of DNasel-sensitivity in somatic hybrids containing the mutant chromosome (13,14); the presence of the LCR has been shown to be needed for proper expression of

human globin genes in transgenic mice (15), demonstrating its ability to counteract position effect variegation.

The sequences of these four hypersensitive sites and their relative positions are highly conserved throughout the mammalian kingdom (16). However, between upstream sequences the degree of homology is not so strong (17) and several of the additional DNasel-hypersensitive sites have no homologues. Bender and colleagues recently described two 'weak' sites in mouse, HS5 and HS6, that so far have not yet been ascribed a function (18). The human HS5 is in a similar position though it is not homologous to either of these sites, and has been shown to have chromatin insulator function (19). The characterization of such new hypersensitive sites is important as they may be as yet uncharacterized regulatory elements, which contribute to the suggested functional differences of the two LCRs (20).

The accurate quantification of conventional DNaseI-sensitivity assays is difficult. Southern hybridization assays are sensitive to many variables, such as probe choice, efficiency of transfer of the target DNA to the membrane and annealing efficiency to targets of different size. Also, the DNaseI-generated sub-band can be diffuse, and difficult to quantify, if the site of digestion is disperse. Quantitative PCR has been used to assess sensitivity, (21) and the application of real time PCR has the advantages of even greater accuracy and sensitivity. The double-strand DNA-specific fluorescent dye SYBR Green is used to detect the accumulation of PCR products as a function of the number of rounds of amplification. Primers are designed to separately amplify similar-sized products from either the LCR or a known DNaseI-insensitive gene, Nf-M (which is used as an internal control for the amount of template), from a series of genomic DNAs isolated from DNaseI-treated nuclei. By reference to a standard curve, it is possible to calculate the number of amplicons destroyed as a function of DNaseI concentration. These data allow an examination of the kinetics of digestion for each site and an accurate determination of the proportion of sites which proved DNaseI-insensitive. This approach was used to investigate the chromatin structure of the murine beta-globin locus as an exemplary embodiment.

The following examples are offered to illustrate embodiments of the present invention, but should not be viewed as limiting the scope of the invention.

Examples

The following materials and procedures were used for the subsequent examples.

Mouse fetal liver was DNaseI-digested in samples by harvesting twenty fetal livers from 11.5- 12.5 d.p.c mouse embryos. The livers were pooled and dispersed in a loose fitting homogenizer in 5 ml buffer A (15 mM TrisHCl pH 7.6, 60 mM KCl, 15 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.15 mM spermidine. Cells were lysed in the presence of 0.5% NP40, and nuclei collected and resuspended in digestion buffer (buffer A supplemented with 3 mM $CaCl_2$, 75 mM NaCl) at a concentration of 800 ug/ml chromatin. Digestions were carried out at 37°C with 0.25 to 8 units DNaseI (Sigma)- for 3 min before being stopped by the addition of an equal volume of stop buffer (0.1 M NaCl, 0.1% SDS, 50 mM Tris-HCl pH 8.0, 100 mM EDTA). The samples were treated with Proteinase K overnight and DNA recovered after phenol-chloroform extraction and ethanol precipitation. The DNA was then dialysed against two changes of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), re-precipitated and diluted in water to a concentration of 50 ng/μl as measured by DNA flurometry.

DNaseI-sensitivity studies were carried out by performing real-time quantitative PCR on the DNA samples using the Roche Molecular-Biochemical's Lightcycler system. PCR reactions were performed using the manufacturer's SYBR-green master kit optimized for use with the following primer pairs (see Figure 1 for the positions of the amplicons within the ß-globin LCR; the co-ordinates of the sequence of the amplified fragment and the GI accession number of the sequence file are shown):

HSIf, 5'-AGA TTA TAT TGC CAT GGT ACA CTT GAA-3' (SEQ ID NO 1);

HSIr, 5'-ACT GGA CCA ATT TTC TCC CTC C-3' (SEQ ID NO 2), (297I-3441; GI: 50150);

Fklf, 5'-GGC TTT TGA GGC AAC ACT AT-3' (SEQ ID NO 3);

Fk1r, 5'-CAT TGG CAG AAA GCT CTC ATA CA-3' (SEQ ID NO 4), (5709-6179; GI: 50150);

Fk2f 5'-GGA. TTT TAC TAT ATA ACT ATG CTA TCA-3' (SEQ ID NO 5);

Fk2r 5'AGA AAG TAA GGG ATGACG TGT AAT ACA AC-3' (SEQ ID NO 6), (11130-11601; GI: 50150);

HS2f, 5'-AGT GTC AGC ATA TTA CCG-ATG-TTC C-3' (SEQ ID NO 7);

HS2r, 5'-CAC ACA GCA AGG CAG GGT C-3' (SEQ ID NO 8), (11767-12237; GI: 50150);

HS3f, 5'-TGT AAG TGT AAA TTT TGG AGC ACA GG-3' (SEQ ID NO 9);

HS3r; 5'-CTG AAA GAC TAA AGT TCC CGG C-3' (SEQ ID NO 10), (5639-6119; 50515);

Fk3 5'-GTA CCA TGT GTT TGT GTG AAG TAG A-3` (SEQ ID NO 11);

Fk3r 5'-TTT TTG AGG ATT TCC ATC AGC AT-3' (SEQ ID NO 12), (7423-7943; GI: 50515);

HS4f, 5'-TGT TTG TGG TTT TTC TGT TGT ATG TTT-3' (SEQ ID NO 13);

H54r, 5'-AAG AGC AGA AAG GAA TTA AAT ACA CAC A-3' (SEQ ID NO 14), (62758-63257; GI: 11908223);

HS6f, 5'-CAG AGC ATT GTT GAA AGA TGA GGA-3' (SEQ ID NO 15);

HS6r, 5`-GGG TTA GCA GAA ATG TAG AGC TCC-3' (SEQ ID NO 16), (57283-577733; GI: 11908223);

Nf-Mf 5'- GCT GGG TGA TGC TTA CGA CC-3' (SEQ ID NO 17);

Nf-Mr, 5'-GCG GCA TTT GAA CCA CTC TT-3' (SEQ ID NO 18), (988-1438; GI: 53357).

PCR was performed using FastStart DNA SYBR-green kit (Roche) as per the manufacturer's instructions. The reactions were performed in a volume of 10 μl with 0.5, 5 and 50 ng of genomic DNAs isolated froze the mouse erythroleukemia line (MEL) in order to prepare standard curves for each of the amplicons, or 50 ng of sample from each DNaseI-digestion point. The number of copies of target remaining intact, corrected for the size of the fragment, was calculated by reference to the standard curve. These data were then plotted versus the units of DNaseI used to digest the nuclei.

An alternative protocol is to use the LightCycler FastStart DNA Master SYBR green I mix at 1x final concentration (containing PCR buffer, dNTPs, $MgCl_2$ and *Taq* polymerase), and to assemble the QPCR reactions as follows: 0.3 uM of each primer, additional $MgCl_2$ to 3 mM and 5-30 ng of template DNA. Reactions are

thermal-cycled under two sets of conditions depending on amplicon size; 250 bp or 500 bp. For 250 bp amplicons the following amplification protocol is used: 95°C for 10 minutes, followed by 40 cycles of 95°C; 5 sec, 60°C; 5 sec and 72°C; 15 sec. For 500 bp amplicons the 72°C phase is extended to 25 sec. Both amplification protocols are followed by a melting curve analysis.

Quantification of DNaseI-digestion by Real-time PCR

Previous studies have used semi-quantitative PCR to estimate the extent of DNaseI-sensitivity in the HPRT gene of tissue culture cells (21). With the development of Real-time PCR it has become possible to perform routinely more accurate and reproducible quantitative PCR. SYBR-green was used as a generic probe for double-stranded DNA and it was possible to detect as little as 20% difference in the number of templates present in separate samples. The approach taken was to separately amplify fragments of interest from 50 ng of DNaseI-treated DNA, and calculate the amount of template destroyed in the sample by reference to a standard curve. Primers were designed to amplify 450 bp fragments throughout the mouse β-globin LCR (Figure 1), including the DNaseI-hypersensitive sites and intervening sequences.

The genomic DNA samples that were tested had been harvested from mouse fetal live nuclei that had been treated with increasing amounts of DNaseI, in order to establish the sensitivity of their chromatin structure to digestion. Typically 50 ng of genomic DNA was used for each reaction but it was possible to use as little as 5 ng. A serial dilution of undigested genomic DNA was also analyzed (using 0.5 ng of material as the earliest point) in order to produce a standard curve so that the number of copies of template in each of the samples could be calculated. In order to correct for small differences in the amount of DNA the separate amplification of a similarly-sized fragments from a known DNaseI-insensitive gene, Nf-M (22), was used as an internal control for the amount of DNA present in the reaction. A Southern hybridization assay confirmed that there was no detectable digestion of this gene under the conditions used (data not shown).

Chromatograms from the Roche Molecular Biochemical's Lightcycler Instrument are shown in Figure 2 for the amplification of (a) Nf-M and (b) HS2. Each experiment contains reactions to calculate a standard curve (using 0.5, 5 and 50

ng of untreated genomic DNA), 50 ng of samples from each of the digestion pints of the DNaseI series (using 0.25, 0.5, 1, 2, 4 and 8 united of enzyme) and a negative control with water added instead of template DNA. The machine measures the gain of fluorescence, due to the proportional binding of SYBR-green to the PCR product, and plots the log of this value as a function of cycle number. The more copies of amplicon present in the sample the fewer cycles of PCR will be needed to achieve a threshold value of fluorescence (indicated by the horizontal green line and set to include the first values above background). This number of cycles, the C_T value, is calculated by extrapolation from the linear part of the curve (identified by the two gray crosses) where amplification is exponential and plots the intercept with the threshold line, represented by the red crosses. Hence the data used to generate the standard curve for each experiment (the intercept is identified with the amount of template used) show a regular increases in the C_T value as the amount of template is decreased by a factor to ten. A plot of the log of concentration of template against C_T value generates a straight line which is used to calculate the percentage of remaining amplicons in the DNaseI-digested samples (Figure 3c).

In figure 2a all of the DNaseI-treated samples have C_T values close to that of the value for 50 ng of untreated DNA, consistent with none of the amplicons being lost due to digestion by DNaseI. The pattern for HS2 is different. Less and less amplicons remain intact as increasing amounts of DNaseI were used to treat the samples. This is consistent with the HS2 site being preferentially destroyed by DNaseI digestion.

To confirm that the signals recorded were due to the amplification of the desired product, melting curve analysis (Figure 3a) and agarose gel electrophoresis (Figure 3b) were performed. Melting curve analysis denatures the PCR product after the final step of PCR by raising the temperature and continually monitors the dissociation of DNA stands by measuring the loss of fluorescent signal. The melting curve for each of the PCR products consisted of a single peak (centered at 84°C and 89°C representing the HS2 and Nf-M products respectively) consistent with the presence of a single species of product. The positive signal for water recorded in the HS2 experiment is shown to have a much lower melting temperature than that of the HS2 product (79°C) suggesting that these products are non-specific primer-dimers.

Agarose gel electrophoresis of samples recovered from the experiments confirmed this; the products of Nf-M and HS2 were single bands of 450 bp, which were not evident in the HS2 negative control.

The digestion profile can be seen by plotting the percentage of copies of HS2 remaining, corrected for DNA content, against the number of units of DNaseI used in the digestion of nuclei (Figure 4). Two general features of the curve are apparent; the initial rate of loss of copies is very fast and the curve reaches a plateau, representing the number of sites that are not accessible to digestion. The first property is an indication of the level of sensitivity of the site to digestion (a less accessible site would be expected to digest more slowly, leading to a slower rate). The second establishes the proportion of material which remains inaccessible, due to the cells being derived from the non-erythroid compartment of the tissue where the hypersensitive site does not form.

Comparison of digestion profiles

The experiments were repeated using the primer pairs spanning the LCR region and their digestion profiles are shown in Figure 5. It can be seen that there are four distinct types of profiles. The simplest is that of Nf-M (Figure 5a), which under these conditions remains undigested so that there are no copies of the amplicon lost. If the amount of nuclease used was steadily increased then these Nf-M sites would eventually be destroyed. In contrast to this profile the primers directed towards the DNaseI-hypersensitive sites (HS1, HS2, HS3 and HS4) show rapid loss of copy number to a stable number (Figure 5a). As discussed previously it is believed that this represents the rapid loss of accessible sites but the persistence of inaccessible sites. The sites most likely are inaccessible as the formation of DNaseI-hypersensitivity is erythroid dependent and the fetal liver will be composed of approximately 20% non-erythroid cells at the time of harvest (12.5 d.p.c.).

The profiles of the intervening sequences (Fk1-3) are similar to each other and are plotted in a shaded compartment of Figure 5b. They show a less steep rate of loss of copy number than with the hypersensitive sites and the curve plateaus later and a higher level. The final profile recorded is that of the 'weak' DNaseI-hypersensitive HS6 (Figure 5b); the rate of loss is similar to that of the other

hypersensitive sites and it reaches a plateau early, but it is at a higher level. All four types of profile are plotted together in Figure 6.

The two properties of these profiles, rate of digestion and proportion of insensitive material, can be interpreted to reveal the underlying structure of the chromatin. A fast rate of digestion is consistent with the site being extremely accessible in the nuclei to digestion, as it is expected to occur at a hypersensitive site where the nucleosomal array is interrupted allowing the DNaseI enzyme free access to cut. As shown in Figure 7a the enzyme can independently cut twice to completely destroy the site. A slower rate of digestion is presumably caused by the site being blocked in vivo, most likely by a nucleosome, and this competing with DNaseI for access to the site (Figure 7b). If cutting does occur here it may be more likely to be a single-stranded nick. This will be detected by quantitative PCR (as 50% of the template has been destroyed) but would not be detected in a Southern hybridization assay (as the nicked molecule would migrate through the gel as a duplex). In the case of the inaccessible site the nucleosome does not have a fast on-and-off rate, say as it is a tightly folded part of the chromatin fiber, and it effectively blocks access for DNaseI (Figure 7c).

These data show that the hypersensitive sites HS1 to HS4 are nucleosome-free in the vast majority of erythroid tissue. The intervening sequences are part of an 'open' nucleosomal array, where nucleosomes compete with DNaseI for access to underlying sequences. As the nuclease will only nick the DNA, the higher plateau reached in their profiles does not represent a higher proportion of entirely inaccessible sequences but all the sites in erythroid tissues receiving a single cut. These interpretations allow us to determine the structure of HS6. The relative 'weakness' of this site could be due to either it forming in only half of the erythroid tissues or not fully displacing a nucleosome from the array. The data presented here demonstrate that the former conclusion is correct; crucially the site is rapidly degraded, consistent with it being nucleosome-free, implying that the higher plateau truly represents a greater proportion of inaccessible sites.

Evaluation of the mouse β-globin LCR structure

The data generated in the above examples show that the DNaseI-hypersensitive sites HS1 to HS4 are formed in all the erythroid cells of mouse fetal

livers. This is an important observation as it was possible that individual sites may have formed with different efficiencies. If such were the case it would imply that the canonical LCR in erythroid tissue could be functionally heterogeneous, as transgenic studies which delete individual hypersensitive sites from the human β-globin locus show that these deleted LCRs sponsor altered transcriptional programs (23, 24). In order to address this question of whether the hypersensitive sites form independently or co-operatively it would be preferred to use a system where the sites form sub-optimally. The approach taken here is to study transgenic mice carrying a deletion of one of the DNaseI-hypersensitive sites. In this system the level of expression is lower and not all hypersensitive sites form.

It was estimated that the 'weak' DNaseI-hypersensitive site HS6 is only formed in half of the erythroid tissue, implying that in vivo the chromatin upstream of the canonical LCR is structurally heterogeneous. However, it is not known whether this structural heterogeneity will have any functional consequences as initial studies have yet to define any cis-acting regulatory role to HS6 (18).

These studies also allowed the investigation of the chromatin structure of sequences between the DNaseI-hypersensitive sites. The digestion profiles of these sequences was different from that of the hypersensitive sites suggesting that the functional LCR could be imagined as being covered by an 'open' nucleosomal array studded with gaps corresponding to the appearance of the hypersensitive sites. The formation of such gaps would be predicted to alter local chromatin structure, as folding of the chromatin fiber involves nucleosome-nucleosome contact. This localized loosening of chromatin structure may play a role in domain-wide alterations associated with LCR activity. The increase in accessibility of sequences in the LCR may allow binding of histone acetyl-transferases or chromatin-modifying complexes (25) capable of propagating the chromatin-opening activity.

Other embodiments and uses of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. All references cited herein, including all U.S. and foreign patents and patent applications and U.S. Provisional application number 60/294,890 filed May 30, 2001, are hereby specifically and entirely incorporated herein by reference.

It is intended that the specification and examples be considered exemplary only, with the true scope and spirit of the invention indicated by the following claims.

References

1. Felsenfeld, G. (1992). Chromatin as an essential part of transcriptionally active genes. Nature 355, 219-224.

2. Sperling, K., Kerem, B. S., Goitein, R., Kottusch, V., Cedar, H. & Marcus, M_ (1985). DNasel sensitivity in faculative and constitutive heterochromatin. Chromosoma 93, 38-42

3. Chen, T. A. & Allfrey, V. G. (1987). Rapid and reversible changes in the nucleosome structure accompany the activation, repression and superinduction of the murine proto-oncogenes *c-fos* and *c-myc*. *Proc. Nail. Acad. Sci. USA* 84, 5252-5256.

4. Chen, T. A., Sterner, R., Cozzolino, A. & Allfrey, V. G. (1990). Reversible and irreversible changes in nucleosomal structure along the *c-fos* and *c-myc*. oncogenes following inhibition of transcription. *J. Mol. Biol.* 212, 481-493.

5. Elgin, S. C. R. (1988). The formation and function of DNasel hypersensitive sites in the process of gene activation. *J. Biol. Chem.* 263, 19259-19262.

6. Gui, C. Y. & Dean, A. (2001) Acetylation of a specific promoter nueleosome. accompanies activation of the epsilon-globin gene by the beta-globin locus control region. *Mol. Cell Biol.* 21, 1155-1163.

7. Stamatoyannopoulos, J. A., Goodwin, A., Joyce, T. & Lowrey, C. H. (1995). NF-E2 and GATA binding motifs are required for the formation of DNaseI-hypersensitive site 4 of the human beta-globin locus control region. *EMBO J.* 14, 106-116.

8. Adams, C. C. & Workman, J. L. (1995) Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell Biol.* 15, 1405-1421.

9. Li, Q., Haru. S. and Peterson, K. R. (1999). Locus Control Regions: Coming of age at a decade plus. *Trends. Genet.* 15, 403-408.

10. Tuan, D., Solomon, W., Li, Q. & London, M. (1985). The ß-like globin gene domain in human erythroid cells. *Proc. Natl. Acad. Sci.* U.S.A. 82, 6384-6388.

11. Forrester, W. C., Thompson, C., Elder, J. T. and Groudine, M. (1986). A developmentally stable chromatin structure in the human ß-globin gene cluster. *Proc. Natl. Acad. Sci.* U.S.A. 83, 1359-1363.

12. Forrest, W., Takagawa, S., Papayannopoulou, T., Stamatoyannopoulos, G & Groudine, M. (1987). Evidence for a locus activation region: The formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucl. Acids Res*, 15, 10159-10177.

13. Driscoll, M. C.; Dobkin, C. S. & Alter, B. P. (1989).γδß-thalassemia due to a *de novo* mutation deleting the 5' ß-globin gene activation-region hypersensitive sites. *Proc. Natl. Acad. Sci.* U.S.A. 86, 7470-7474.

14. Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M-, Papayannopoulou, T. & Groudine, M. (1990). A deletion of the human ß-globin locus activation region causes a major alteration in chromatin structure and replication across the entire ß-globin locus. *Genes Dev.* 4, 1637-1649.

15. Grosveld, F., van Assendelft, G. B, Greaves, D. R. & Kollias, G. (1987). Position-independent, high level expression of the human ß-globin gene in transgenic mice. *Cell* 51, 975-985.

16. Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N. & Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene* 205, 73- 94.

17. Bulger, M., von Doorninek, J. H_, Saitoh, N., Telling, A., Farrell, C., Bender, M. A., Felsenfeld, G., Axel, R. & Groudine, M. (1999). Conservation of sequence and structure flanking the mouse and human ß-globin loci: ß-globin genes are embedded within an array of odorant receptor genes. *Proc. Natl. Acad. Sci.* USA 96, 5129-5134.

18. Bender, M. A., Reik, A., Close, J., Telling, A., Epner, E., Fiering, S,, Hardison, R_ & Groudine, M. (1998), Description and targeted deletion of 5' hypersensitive site 5 and 6 of the mouse ß-globin locus control region. Blood 92, 4394-4403.

19. Li, Q. & Stamatoyannopoulos, G. (1994). Hypersensitive site 5 of the human betaglobin locus control region functions as a chromatin insulator. *Blood* 84., 1399-1401.

20. Higgs, D. R. (1998). Do LCRs open chromatin domains? Cell 95, 299-302.

21. Pikaart, M., Feng, J. and Villponteau, B. (1992). The polyomavirus enhancer activates chromatin accessibility on integration into the HPRT gene. *Mol. Cell- Biol.* 12, 5785-5792.

22. Gribnau, J., Didrech, K., Pruzina, S., Calzolari, R. and Fraser, Q. (2000). Intergenic transcription and developmental remodelling of chromatin subdomains in the human ßglobin locus. *Mol. Cell* 5, 377-386.

23. Peterson, K. R., Clegg, C. H., Navas, P. A., Norton, E. J., Kimbrough, T. G. & Stamatoyannopoulos, G. (1996). Effect of deletion of 5'HS3 or 5'HS2 of the human betaglobin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proc. Natl. Acad. Sci.* U.S.A. 93, 6605-6609.

24. Navas, P. A_, .Peterson, K. R., Li, Q., Skarpidi, E., Rohde, A., Shaw, S. E., Clegg, C. H., Asano, H., Stamatoyannopoulos, G. (1998). Developmental specificity of the interaction between the locus control region and embryonic or fetal globin genes transgenic mice with an HS3 core deletion. *Mol. Cell Biol.* 18, 4188-4196.

25. Travers, fA. (1999). Chromatin modification by DNA tracking. *Proc. Natl. Acad. Sci.* U.S.A. 93, 13634-13637.

Claims

1. A method for accurately and efficiently determining sensitivity of a candidate region of a genome to a DNA modifying agent comprising:

a) isolating chromatin from a population of eukaryotic cells containing said genome;

b) treating at least one portion of said isolated chromatin with said DNA modifying agent under conditions to cause DNA strand breakage;

c) treating another portion with said DNA modifying agent under modified conditions;

d) isolating treated DNA from the portions;

e) amplifying the candidate region from isolated DNA by real-time PCR from each portion by real time PCR with a set of primers and obtaining a signal;

f) determining a relative copy number of said candidate region within each isolated DNA portion by:

(i) determining the copy number of said candidate region in each of a plurality of DNA samples each containing a different amount of DNA in a fixed proportion with respect to each other and thereby calculating a first standard copy number curve;

(ii) determining the copy number of a reference region in each of a plurality of DNA samples each containing a different amount of DNA in a fixed proportion with respect to each other and thereby calculating a second standard copy number curve from PCR amplification of said reference region; and

(iii) comparing the signal to said first standard curve and said second standard curve to obtain the relative copy number ratio of the candidate region and the reference region for each isolated DNA; and

g) determining sensitivity of said candidate region to said DNA modifying agent relative to the sensitivity of said reference region.

2. The method of claim 1, wherein the genome is a human genome.

3. The method of claim 1, wherein the candidate sequence is less than 250 base pairs in length.

4. The method of claim 1, wherein the candidate region is between about 50 and about 2,000 base pairs in length.

5. The method of claim 1, wherein the one portion of isolated chromatin comprises from 1.5 pg to about 5 ug of nucleic acid.

6. The method of claim 1, wherein the one portion of isolated chromatin comprises one copy of said genome.

7. The method of claim 1, wherein the DNA modifying agent is selected from the group consisting of a non-specific endonuclease, a sequence-specific endonuclease, a DNAse, DNase I, S1 nuclease, micrococcal nuclease, mung bean nuclease, P1 nuclease, a topoisomerase, topoisomerase II, a methylation-sensitive enzyme, DpnI, MspI, HpaII, a chemical DNA modifying agent, hydrogen peroxide, potassium permanganate, a DNA-modifying chemotherapeutic agent, radiation, UV radiation, histone acetylation, cytosine methylation, nuclease, topoisomerases; methylases; acetylases; chemotherapy agents that effect DNA; radiation; physical shearing; nutrient deprivation, folate deprivation, and combinations thereof.

8. The method of claim 1 wherein the conditions and the modified conditions are selected from the group consisting of different concentrations of the DNA modifying agent, different times, different temperatures, different buffer conditions, different concentrations of chromatin, and combinations thereof.

9. The method of claim 1, wherein the conditions and the modified conditions are performed at the same temperature.

10. The method of claim 1, wherein amplifying each isolated DNA by the real-time PCR proceeds through at least eighteen cycles.

11. The method of claim 1, wherein amplifying each isolated DNA by the real-time PCR proceeds through at least thirty-five cycles.

12. The method of claim 1, wherein the set of primers amplify fragments of DNA that are from about 200 to 400 base pairs in length.

13. The method of claim 1 wherein the reference region is of the same sensitivity to the DNA modifying agent.

14. The method of claim 1 wherein the reference region is more sensitive to the DNA modifying agent.

15. The method of claim 1 wherein the reference region is less sensitive to the DNA modifying agent.

16. The method of claim 1, wherein the reference sequence is less than about 250 base pairs in length.

17. The method of claim 1, wherein the reference region is between about 25 and about 2,000 base pairs in length.

18. The method of claim 1, wherein the reference region is between 200 and 400 base pairs in length.

19. The method of claim 1, wherein a double strand DNA specific marker is used during PCR to detect the accumulation of PCR products as a function of the number of rounds of amplification.

20. The method of claim 1 wherein the DNA modifying agent is DNase I and the DNA modifying reactions utilize a single reaction time point and differing concentrations of DNase I.

21. The method of claim 1, wherein the DNA modifying agent introduces single stranded nicks into the DNA.

22. The method of claim 1, wherein reactions are cycled under multiple sets of conditions depending on amplicon size.

23. A method for determining the sensitivity of a genome of a eukaryotic species comprising:

a) combining cells of a first eukaryotic species with a first genome with cells from a second eukaryotic species with a second genome having a combined cell population totaling at least about 10^8 cells;

b) isolating chromatin from the combined cell population;

c) treating one portion of said isolated chromatin with a first amount of said DNA modifying agent;

d) isolating treated DNA from the portions;

e) amplifying each isolated DNA by real time PCR with a set of primers that amplify a template sequence of the first eukaryotic species; and

f) determining the sensitivity of said template sequence to the DNA modifying agent within the first genome.

24. The method of claim 23, wherein the cells of the first eukaryotic species number less than 10^7.

25. The method of claim 23, wherein the cells of the first eukaryotic species comprise less than about 10^6 cells.

26. The method of claim 23, wherein the DNA modifying agent introduces single stranded nicks into the DNA.

27. The method of claim 23, wherein reactions are cycled under multiple sets of conditions depending on amplicon size.

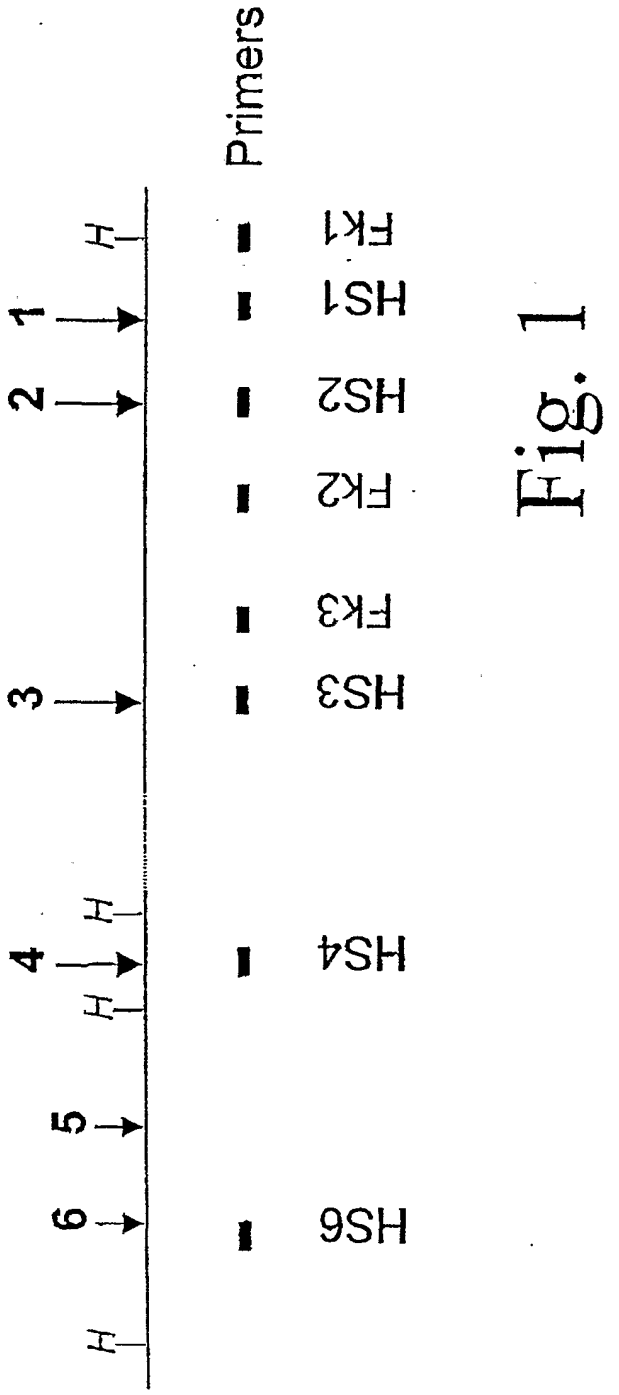


Fig. 1

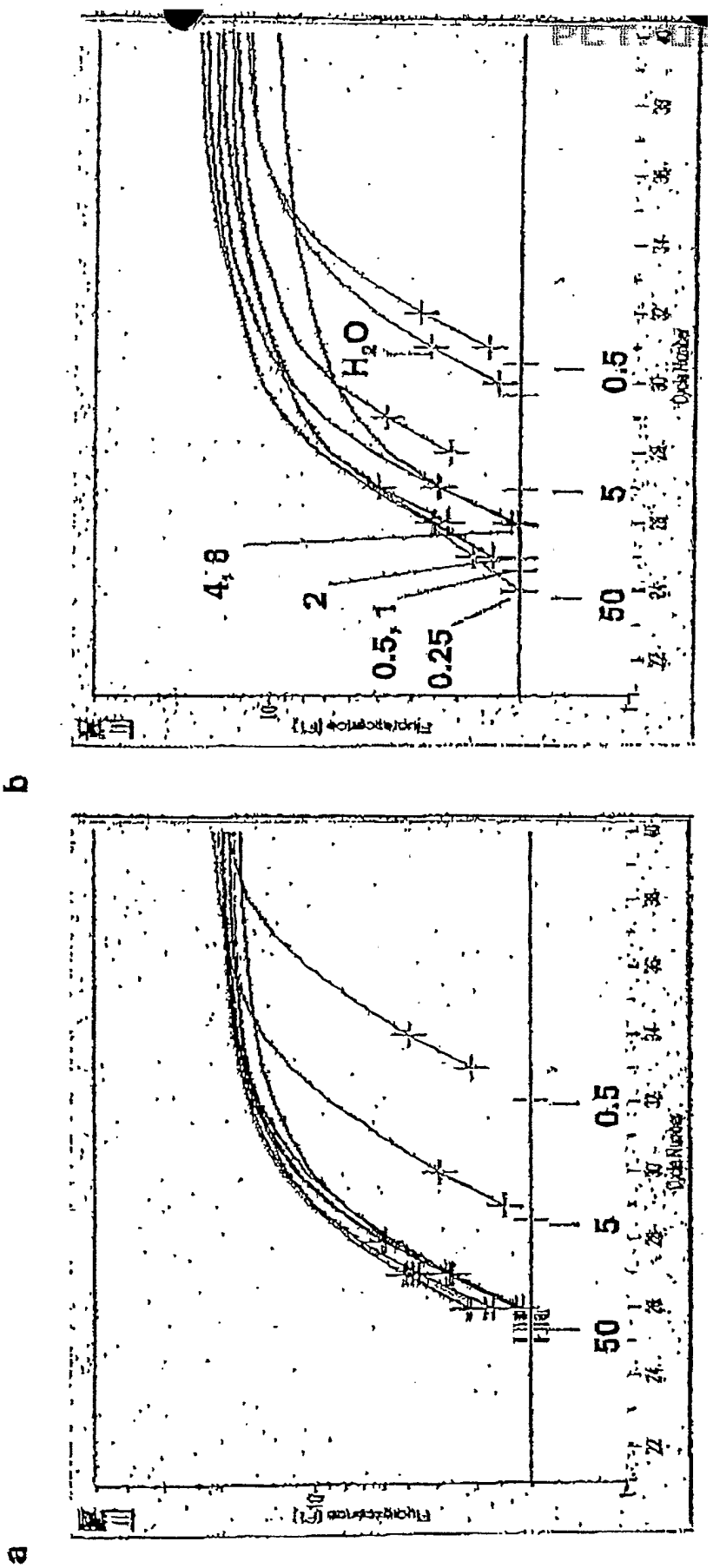
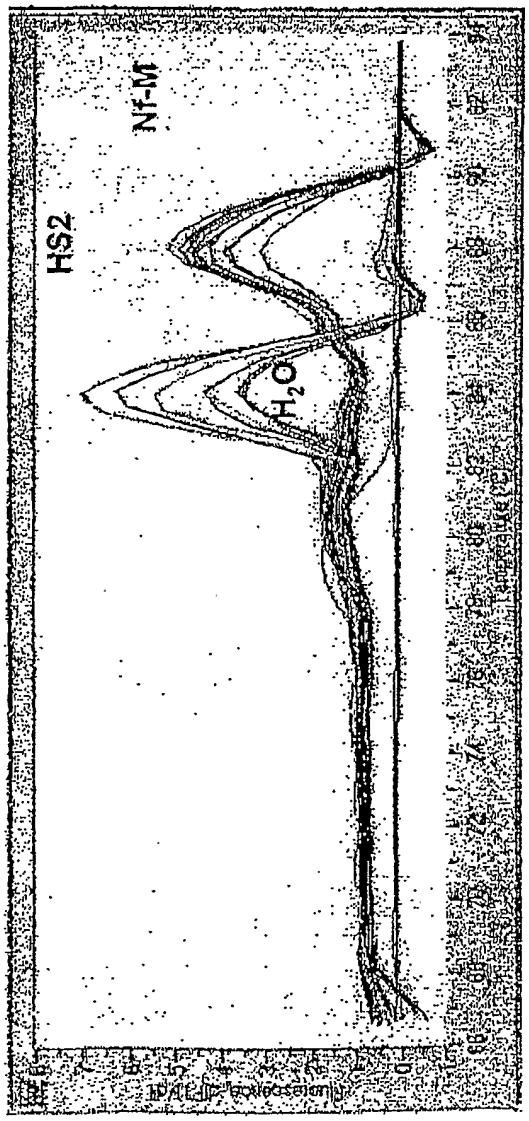


Fig. 2

Fig. 3A



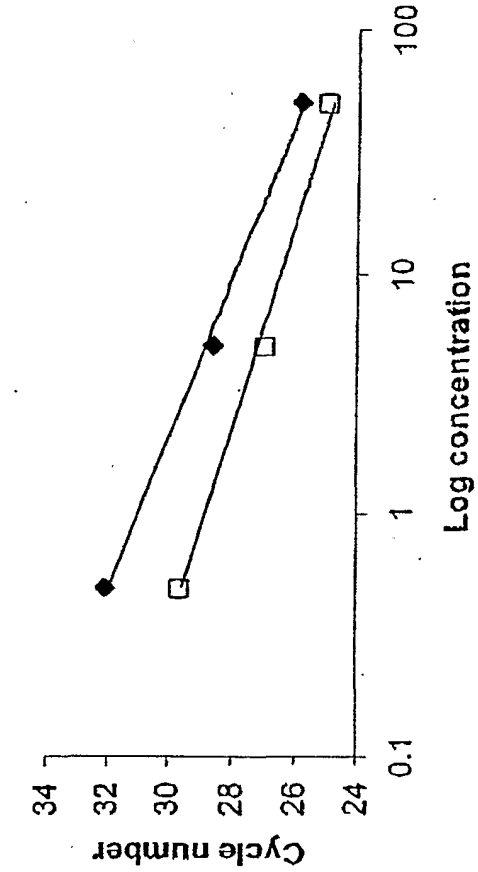


Fig. 3C

Fig. 3B

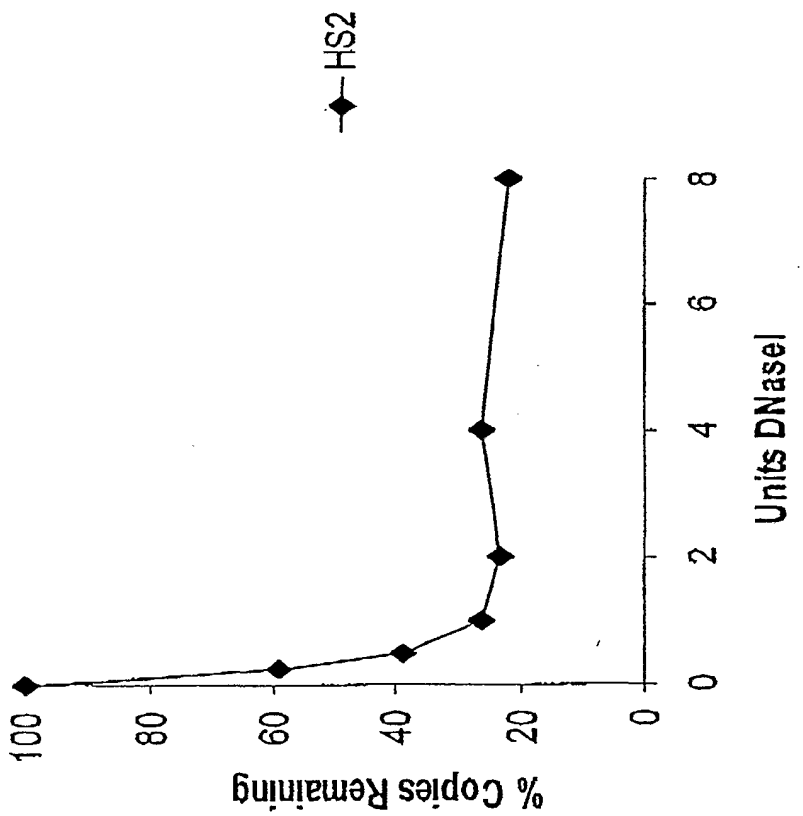


Fig. 4

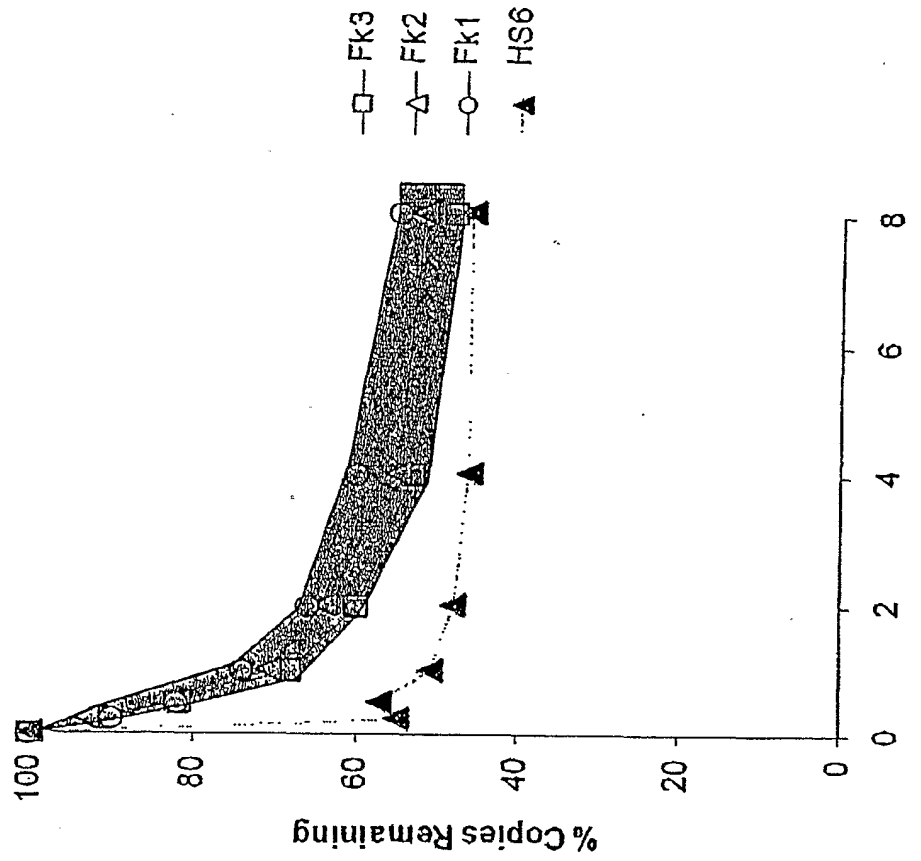

Fk3
Fk2
Fk1
HS6
% Copies Remaining
100
80
60
40
20
0
0
2
4
6
8
Units DNaseI

Fig. 5B

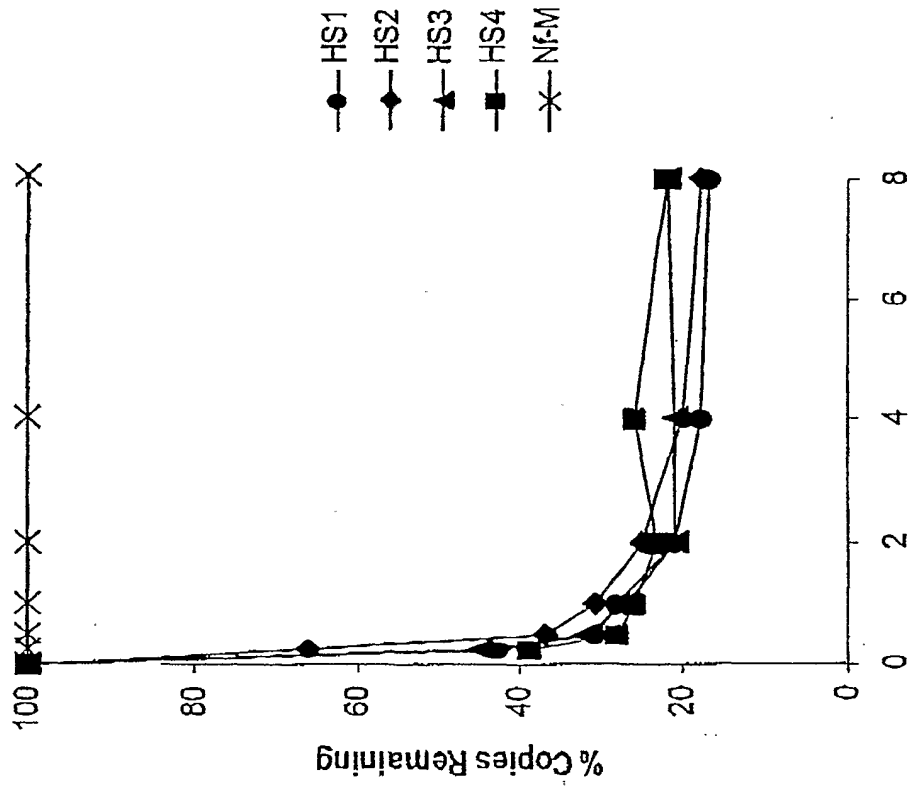

HS1
HS2
HS3
HS4
NF-M
% Copies Remaining
100
80
60
40
20
0
0
2
4
6
8
Units DNaseI

Fig. 5A

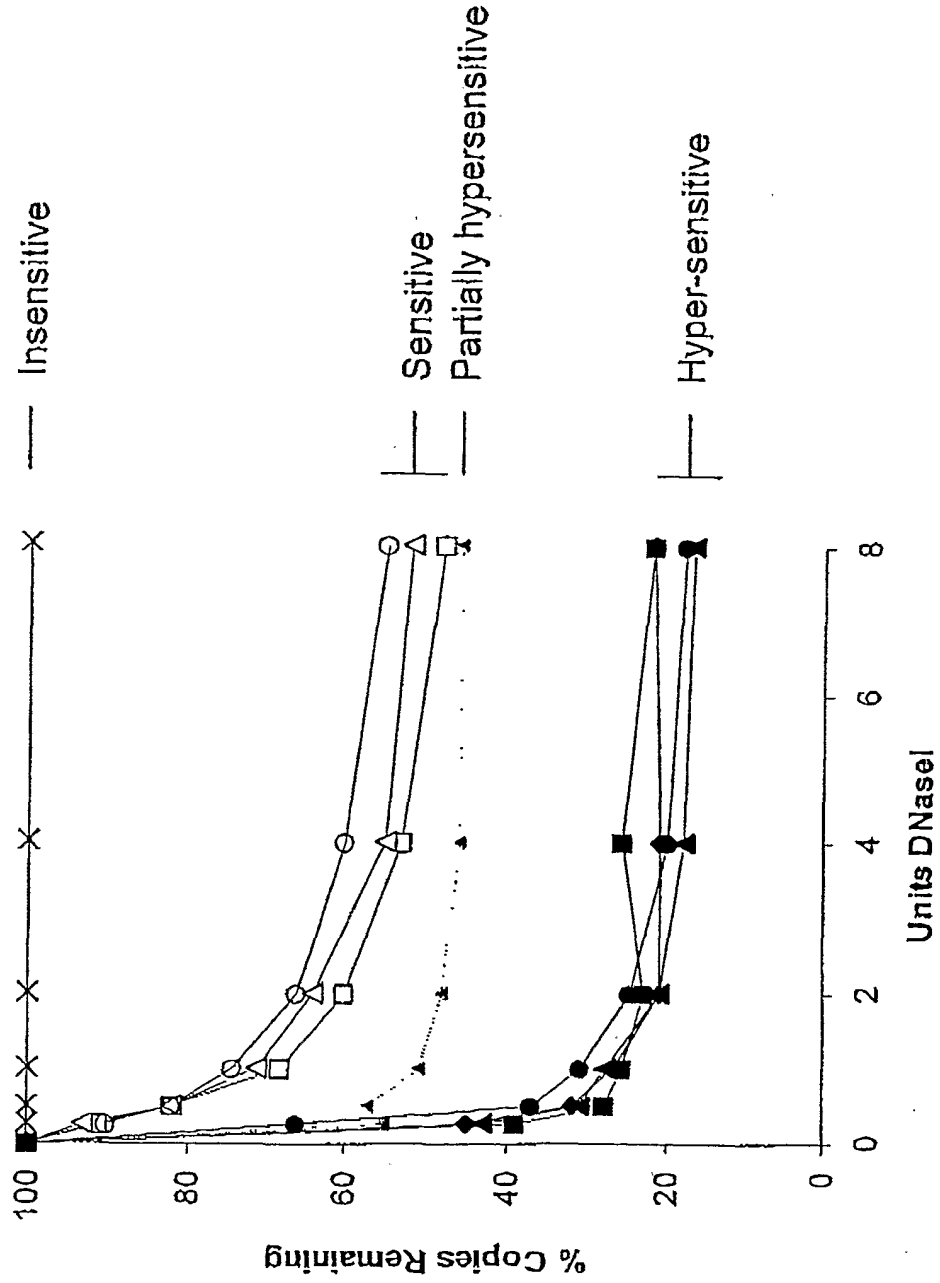

Insensitive
Sensitive
Partially hypersensitive
Hyper-sensitive
% Copies Remaining
100
80
60
40
20
0
0
2
4
6
8
Units DNaseI

Fig. 6

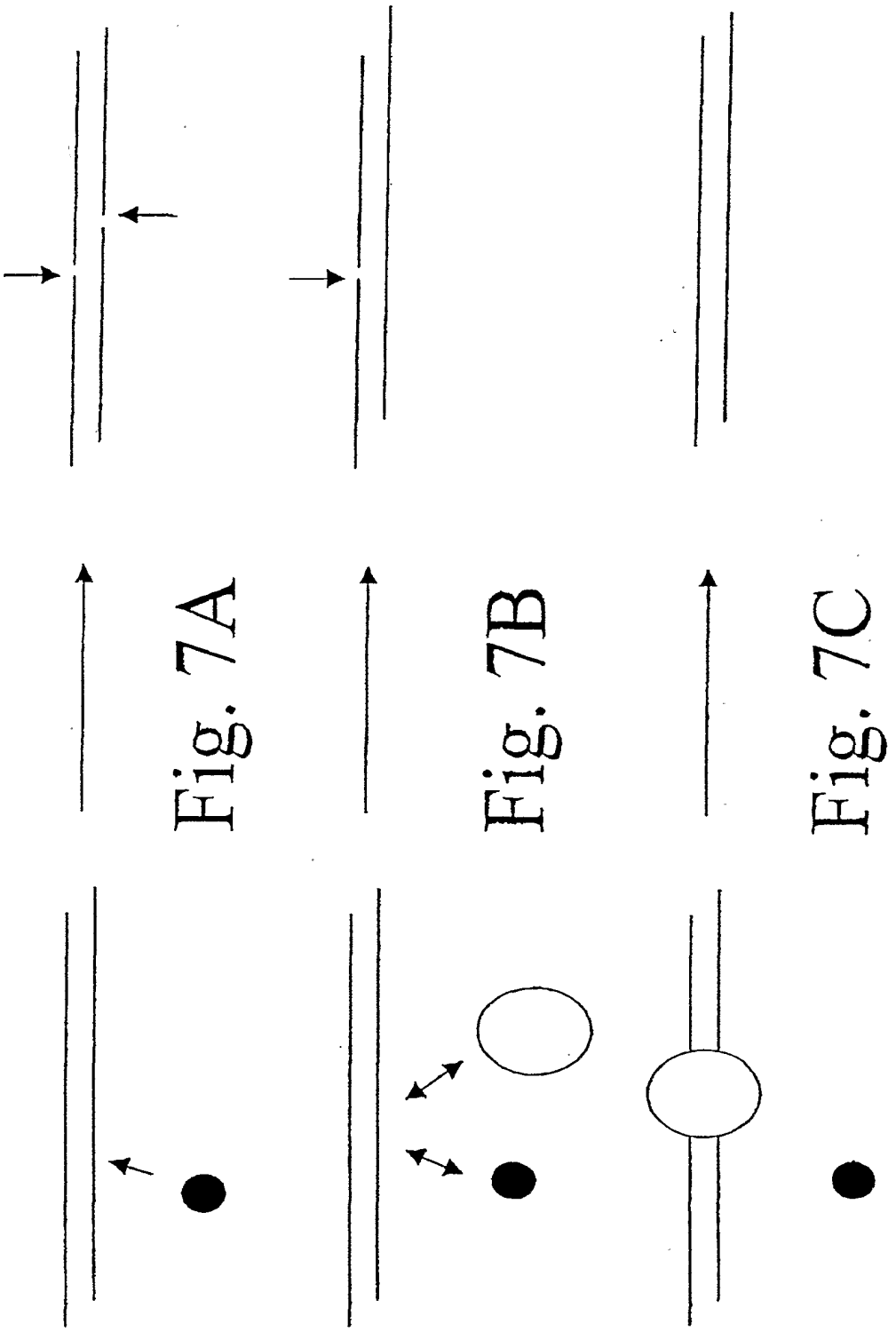


Fig. 7A

Fig. 7B

Fig. 7C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/16967

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12Q 1/68; C12P 19/34; C12N 9/12; C07H 21/04

US CL : 435/6, 91.2, 196; 536/23.1, 24.33

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2, 196; 536/23.1, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, Dialog ONESEARCH: dna, damage, methyla+, radiation, copy number, real time PCR, PCR, chromatin, topoisomerase, peroxide, permanganate, UV, histone, et al.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,180,349 B1 (GINZINGER et al) 30 January 2001, see entire document, especially claims	1-27
Y	US 5,180,666 A (STATES et al) 19 January 1993, see entire document, especially columns 6-7	1-27
Y	US 6,210,878 B1 (PINKEL et al) 03 April 2001, see entire document	1-27

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 01 JULY 2002	Date of mailing of the international search report 02 AUG 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHANIE ZITOMER Telephone No. (703) 308-3985

Form PCT/ISA/210 (second sheet) (July 1998)★