Université de Montréal

Hypermethylation of the *retinoic acid receptor* β P2 promoter in lung and colon cancer cell lines: Multiple restriction enzyme digest - improved bisulphite genomic sequencing protocol.

par
Jane J. Pappas

Département de médecine
Faculté des études supérieures

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en biologie moléculaire

janvier 1998

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:

"Hypermethylation of the *retinoic acid receptor* β
P2 promoter in lung and colon cancer cell lines:
Multiple restriction enzyme digest - improved bisulphite
genomic sequencing protocol"

présenté par
Jane J. Pappas

a été évalué par un jury composé des personnes suivantes:

Président du jury: [redacted]

Dr. Richard Bertrand
Hôpital Notre-Dame, Montréal, CANADA

Directeur de recherche: [redacted]

Dr. W.E.C. Bradley
Hôpital Notre-Dame, Montréal, CANADA

Examinateur externe:

Dr. Richard L. Momparler
Hôpital Ste-Justine, Montréal, CANADA

Mémoire accepté le 26 mai 1998.

SUMMARY

Lung cancer is presently the number one cause of cancer-related mortalities in the industrialized world. It has been estimated that approximately 17, 000 deaths will be attributable to lung cancer in Canada, in 1997. Tobacco consumption is considered to be the main risk factor for lung cancer.

Genetic and epigenetic alterations involving the short arm of chromosome 3 are frequent cellular events in lung cancer. 100% of small cell lung carcinomas and 90% of epidermoid lung carcinomas have chromosomal deletions spanning 3p13/14-ter and 3p24-ter, respectively. This chromosomal region is therefore thought to harbour tumour suppressor genes. The ***RAR*β gene** which codes for the putative tumour suppressor gene product, **retinoic acid receptor β isoform 2** (RARβ2), is located in this region: its chromosomal locus is within 3p24. It is not surprising then, that most lung cancer cell lines, including small cell lung carcinoma and epidermoid lung cancer cell lines, do not express RARβ2, due to chromosomal deletions of these types. However, unlike the gain of function of a *dominant* oncogene which requires only <u>one</u> *activating* event, a *recessive* tumour suppressor gene requires <u>two</u> *in*activating events for complete loss of function. Thus, a chromosomal deletion may constitute *one* of these events and another type of alteration causing loss of expression may make up the *other*. Interestingly, most colon adenocarcinoma cell lines do not express RARβ2 either. But in contrast with lung cancer cells, colon cancer cells are not characteristically altered at chromosome 3p.

An established principle is that **DNA hypermethylation**, an epigenetic alteration, is associated with lack of gene expression. DNA methylation in higher eukaryotes involves the addition of a methyl group ($-CH_3$) to the carbon-5 of CpG-cytosines. It requires enzymatic catalysis by the methyl transferase enzyme. Hypermethylation is generally observed in specific CpG-rich regions of the genome, known as "CpG islands". When gene promoters, such as the P2 promoter of RARβ2, are located on CpG islands, they may potentially become hypermethylated. Consequently, gene expression may be reduced.

The primary objective of this project was to determine the methylation status of the RARβ2 P2 promoter in human lung cancer cell lines, human lung tumour cancer cells, and finally, human colon adenocarcinoma cell lines.

The hypothesis was that hypermethylation is associated with lack of RARβ2 expression in these cells. Hypermethylation may thus constitute one of the events involved in RARβ loss of function in RARβ2-*non*-expressing cancer cells.

The rationale behind investigating the status of methylation in colorectal carcinoma cell lines was threefold. First, our lab has previously shown that most colon cancer cell lines *do not* express RARβ. Second, genetic alterations involving chromosome 3p24 have not been associated with colon cancer cells, allowing for the possibility that hypermethylation plays a key role in RARβ inactivation. Third, hypermethylation of the RARβ P2 promoter has previously been demonstrated in certain human colon cancer cell lines, such as DLD-1. Thus, studying cell lines of another organ origin (colon versus lung) not only broadens our pool of information regarding RARβ methylation and inactivation, but also serves to strengthen any correlations which may be found in lung cancer cell lines specifically.

The method used was based on the *Bisulphite Genomic Sequencing Protocol* (Clark *et al.*, 1994), because of its powerful qualitative methylation analysis of all individual CpG-cytosines within a target sequence of choice. **The principle** is that bisulphite converts cytosines to uracils but leaves 5-methylcytosines *un*converted. Subsequent sequencing of PCR-amplified bisulphite-modified DNA, displays 5-methylcytosines as cytosines, but unmethylated cytosines as thymines.

Unfortunately, the protocol has intrinsic problems decreasing its usability. One major difficulty of the protocol is that the chemical conditions of the bisulphite conversion reaction are so harsh, that frequently the amount and/or the condition of recoverable DNA is *not PCR-amplifiable*. A second major difficulty of the protocol is that the pool of bisulphite-treated DNA consists of *multiple molecular species of various degrees of conversion*. Inefficient conversion is undesirable because the positive display of a CpG-cytosine as a cytosine could either reflect methylation (true positive), or lack of conversion (false positive). Since cytosines not within CpG dinucleotides are normally unmethylated, they are all expected to be converted to uracils under optimal conditions. The threshold of confidence, regarding conversion of non-CpG-cytosines, was established at 98.6% in this project.

The secondary objective of this project was to improve the genomic sequencing protocol, by increasing the amount of recoverable bisulphite-treated DNA, and by increasing the efficiency of conversion to the established threshold of 98.6% or greater.

The method developed was called the *Multiple Restriction Enzyme Digest (MRED)-Improved Bisulphite Genomic Sequencing Protocol*. First, the original protocol conditions were modified to maximize DNA recovery. Second, PCR-amplified bisulphite-treated DNA products were digested with several restriction enzymes pre-determined to have sites within *un*converted DNA, but *not* within *converted* DNA. Restriction sites were more or less evenly distributed across the target sequence and were not biased for CpG-cytosine methylation status. Products which remained undigested after MRED were selected for sequencing.

Results show that by using this technique, not only was the modified-DNA recoverable, but nearly 90% of subcloned PCR-amplifications selected for sequencing were converted with 98.6% or 100% efficiency. This is 9-fold greater than that observed when the DNA was treated as per original protocol. Among RARβ2-*non*-expressing cancer cell lines, CALU-1 and NCI-H596 (lung) and COLO-201, COLO-205, HCT-15 and LS-180 (colon) were *hyper*methylated. Among RARβ2-*expressing* cancer cell lines, C-19, C-59, NCI-H157, NCI-H520 and MM-1 (lung) and CACO-2 and SW 1222 (colon) were *hypo*methylated. Moreover, the normal (non-transformed) control cell line NBE-E_6E_7 was *not* hypermethylated. These observations support the hypothesis that RARβ2 P2 promoter hypermethylation is associated with lack of RARβ2 expression.

The main conclusion which can be drawn from these results is that the epigenetic alteration of hypermethylation may constitute one of the inactivating events resulting in the loss of function of the putative tumour suppressor gene, *RAR*β, both in lung cancer cell lines and in colon cancer cell lines.

Future perspectives, with regard to both lung and colon cancer, may include the use of inhibitors of methylation, such as 5-aza-2'-deoxycytidine and methyl transferase antisense oligodeoxynucleotides, to prevent hypermethylation of tumour suppressor genes, including *RAR*β. Inhibitors of methylation could be administered in combination with retinoic acid therapy to further boost RARβ2 expression. This approach could be clinically tested in chemoprevention trials, for example in patients at risk for lung cancer (such as smokers) and in anticancer therapy trials. Interestingly, recent findings suggest

that hypermethylation is associated with smoking history in non-small cell lung carcinoma patients. Again, smoking is a behaviour highly implicated in lung cancer. In the future, the use of such hypomethylating agents, which have carcinogenic effects of their own, may be engineered to be gene-specific or cancer cell-specific through the vehicle of gene therapy.

RÉSUMÉ

De nos jours, **le cancer du poumon** est la principale cause de la mortalité par cancer dans les sociétés industrialisées. On estime qu'il y aura eu environ 17 000 décès dûs au cancer pulmonaire au Canada en 1997. Il est généralement accepté que le facteur de risque principalement associé au cancer du poumon est le tabagisme.

Parmi les cellules pulmonaires cancéreuses, des altérations génétiques et épigénétiques impliquant le bras court du chromosome 3, sont des évènements cellulaires courants. Ainsi, cette région chromosomique fait l'objet de recherche de gènes suppresseurs de tumeur. Plus particulièrement, 100% des carcinômes pulmonaires à petites cellules et 90% des carcinômes pulmonaires épidermoïdes sont caractérisés par des déletions chromosomiques affectant les régions 3p13/14-ter et 3p24-ter, respectivement.

Ces lésions en particulier, impliquent la région 3p24, le locus du gène qui code pour **le récepteur β2 de l'acide rétinoïque** (RARβ2). L'isoforme β2 , par ses propriétés antinéoplasiques, donne au gène *RAR*β sa candidature pour être classifié gène suppresseur de tumeur. Alors, il n'est pas surprennant que la grande majorité des lignées cellulaires pulmonaires cancéreuses, incluant les carcinômes pulmonaires à petites cellules et les carcinômes épidermoïdes, *n'expriment pas* RARβ2. Aussi, la majorité des lignées cellulaires provenant d'adénocarcinômes du colon, n'expriment pas RARβ2 nonplus. Mais ces lignées colorectales ne sont pas associées avec des déletions affectant le chromosome au locus 3p24. En revanche, le récepteur RARβ2 *est* normalement exprimé par la majorité des tissus épithéliaux, incluant l'épithélium bronchique et l'épithélium colorectal.

Les récepteurs de l'acide rétinoïque permettent le relais de signaux extracellulaires directement au noyau, où ils agissent comme facteurs transcriptionnels. De plus, ces récepteurs sont multifonctionnels. Ils agissent au sur la différentiation, la prolifération et l'homéostasie cellulaires.

La nature récessive des gènes suppresseurs de tumeur prédit que deux évènements désactivants sont nécessaires afin que la fonction supprimante soit entièrement abolie. Un oncogène, par contre, a besoin d'un seul évènement pour l'activation de son potentiel tumorigène. Or, en ce qui concerne la perte de fonction d'un gène suppresseur de tumeur,

un évènement désactivant, telle une déletion chromosomique, pourrait être accompagné d'un autre évènement désactivant, telle une altération épigénétique.

Un dogme central généralement accepté, associe l'état de **l'hyperméthylation** d'un gène à sa baisse d'expression. L'hyperméthylation est la modification majeure que subit l'acide désoxyribonucléique (l'ADN) chez les vertébrés. Chez l'humain, la méthylation consiste en l'addition d'un groupement methyl ($-CH_3$) au carbone-5 de la base cytosine faisant partie du dinucléotide CpG. Cette réaction est catalysée par l'enzyme méthyltransférase, et nécessite la molécule donatrice de groupements méthyls, la S-adénosine méthionine. L'hyperméthylation est généralement observée au niveau des régions riches en CpG, nommées îlots CpG. Lorsque le promoteur d'un gène, i.e. le régulateur d'expression en *cis*, est localisé sur un îlot CpG, il est sujet à l'hyperméthylation. Ainsi, son expression peut être réduite.

L'objectif premier de ce projet était de déterminer l'état de méthylation du promoteur P2, responsable de la régulation de l'expression de l'isoforme β2 du gène *RAR*β, dans des lignées humaines pulmonaires cancéreuses, des cellules provenant de tumeurs pulmonaires malignes humaines, et dans des lignées provenant d'adénocarcinômes du colon humains.

L'hypothèse émise était que l'absence d'expression du récepteur nucléaire RARβ2 était associée à l'hyperméthylation du promoteur P2 du gène *RAR*β. L'hyperméthylation pourrait donc constituer un des évènements désactivants impliqués dans la perte de fonction de RARβ2 parmi les cellules cancéreuses n'exprimant pas ce récepteur.

Le raisonnement sur lequel était fondé l'étude du niveau de méthylation dans les lignées colorectales était basé sur trois faits. Premièrement, notre laboratoire a démontré que la plupart des lignées colorectales n'expriment pas RARβ. Deuxièmement, des altérations au niveau du locus 3p24 ne sont pas associées avec les lignées colorectales. Donc, il est possible que dans ces lignées, la méthylation joue un rôle primordial dans l'inactivation de l'expression de RARβ. Finalement, il a été démontré que le promoteur P2 de RARβ est hyperméthylé dans une lignée humaine colorectale, soit la lignée DLD-1. Conséquemment, l'étude de lignées provenant d'un autre site anatomique (le colon versus le poumon) permettrait la cumulation des informations à propos de la méthylation et l'inactivation génique, et, contribuerait à fortifier les associations formulées à partir des résultats obtenus.

La technique employée fut fondée sur le protocole "*Bisulphite Genomic Sequencing Protocol*", ou le *Protocole de Séquençage Génomique au Bisulphite* (Clark *et al.*, 1994). **Le principe** de ce protocole est le suivant: les cytosines non-méthylées dans un fragment d'ADN cible *à simple brin*, réagissent avec le bisulphite pour être ultimement désaminées en uracils. Par contre, les 5-méthylcytosines ne réagissent pas avec le bisulphite et ne sont donc pas converties. Lors du séquençage de produits synthétisés par la réaction en chaîne de la polymérase (PCR), amplifiés à partir de l'ADN ainsi traité au bisulphite, les 5-méthylcytosines sont représentées comme des cytosines, tandis que les cytosines non-méthylées sont représentées comme des thymines. La technique au bisulphite permet l'analyse qualitative de l'état de méthylation de tous les dinucléotides CpG à l'intérieur de la séquence cible. Quant aux méthodes basées sur la digestion de restriction différentielle, type *Hpa* II/*Msp* I, seule une fraction de ces sites peut être analysée, puisque seuls les CpG à l'intérieur des séquences CCGG sont reconnus par ces enzymes. Parmi les 22 sites CpG à l'intérieur de la séquence cible analysée dans ce projet, seulement cinq faisaient partie d'une telle séquence CCGG.

Malgré sa puissance qualitative, la technique au bisulphite peut être caractérisée par deux inconvénients majeurs. De un, les conditions chimiques nocives causent une dégradation importante de l'ADN traité. Le degré de dégradation et/ou des modifications chimiques apportées peut être si élevé, que l'ADN *ne peut pas être amplifié* par la technique de PCR. De deux, les produits de la réaction forment un *ensemble d'espèces moléculaires à degrés variables de conversion*.

Il est considéré que la cause principale de l'absence de conversion est la présence de région(s) en double brin, ou, la formation de structure(s) secondaire(s). Quelques exemples sont, une dénaturation incomplète, le réappariement en double brin ou la formation de structure(s) "hairpin" entre des séquences complémentaires du même brin. Puisque toutes les cytosines ne formant pas le dinucléotide CpG ne sont pas normalement méthylées, nous nous attendons à ce qu'elles soient toutes converties. Si non, lors du séquençage, une cytosine pourrait résulter d'une 5-méthylcytosine (résultat vrai positif), ou bien d'une cytosine non-méthylée qui n'a pas été convertie (résultat faux positif). Conséquemment, le seuil de confiance par rapport à la proportion de conversion a été établi à 98.6% dans ce projet. Ce seuil était calculé en divisant le nombre de cytosines non-CpG converties par le nombre total de cytosines non-CpG dans la séquence cible x

100%. Cette formule exclut les cytosines de dinucléotides CpG, puisque l'état de méthylation de ces dernières est inconnu.

L'objectif second de ce projet était d'améliorer l'utilité du *Protocole de Séquençage Génomique au Bisulphite*, en contournant le degré élevé de la dégradation d'ADN et la multiplicité d'espèces moléculaires partiellement converties.

Le protocole ainsi développé fut nommé "*Multiple Restriction Enzyme Digest (MRED)-Improved Bisulphite Genomic Sequencing Protocol*". Premièrement, le protocole original a été changé par plusieurs modifications mineures afin d'augmenter la quantité d'ADN amplifiable par PCR. Deuxièmement, la modification majeure du protocole original consiste en la digestion des produits PCR (qui étaient amplifiés à partir de l'ADN traité au bisulphite pour les fins de séquençage) par de multiples enzymes de restriction. Ces enzymes de restriction ont été prédéterminées afin d'avoir des sites dans l'ADN *non*-converti, mais *pas* dans l'ADN efficacement *converti*. De plus, les sites de restriction étaient plus ou moins également répartis sur la séquence, au cas où il y avait formation(s) de structure secondaire ou réappariement(s) en double brin, affectant seulement une ou plusieurs régions de la séquence cible. D'autre part, les sites de restriction n'étaient pas biaisés pour l'état de méthylation de sites CpG, afin de ne pas sélectionner pour des patrons de méthylation spécifiques.

Les résultats démontrent que cette technique a permi l'atteinte du seuil de conversion (98.6% ou 100%) dans 90 % des sous-clones sélectionnés pour des fins de séquençages. Ceci représente neuf fois plus de clones efficacement convertis que lorsque le protocole original a été utilisé. Évidemment, les modifications apportées ont aussi permi l'amplification de l'ADN traité.

Parmi les lignées cellulaires cancéreuses *n'exprimant pas* le gène *RAR*β2, les lignées pulmonaires CALU-1 et NCI-H596 ainsi que les lignées colorectales COLO-201, COLO-205, HCT-15 et LS-180, étaient *hyper*méthylées au niveau du promoteur P2. Parmi les lignées cellulaires cancéreuses *exprimant* le gène *RAR*β2, les lignées pulmonaires C-19, C-59, NCI-H157, NCI-H520 et MM-1 ainsi que les lignées colorectales CACO-2 et SW 1222 étaient *hypo*methylées. De plus, la lignée normale (non-transformée) NBE-E_6E_7 n'était pas hyperméthylée. Ces observations supportent l'hypothèse que l'hyperméthylation du promoteur P2 du gène *RAR*β est associée à une absence d'expression de RARβ2.

En conclusion, l'hyperméthylation, une altération épigénétique, peut être un des deux évènements désactivants résultant en la perte de fonction du gène *RAR*β.

Il est à noter que des études récentes suggèrent que l'hyperméthylation est associée au tabagisme chez des patients atteints de carcinôme pulmonaire à non-petites cellules. Admettant que le tabagisme est le facteur de risque principalement associé au cancer du poumon, il est possible que la méthylation pathologique impliquée dans la carcinogénèse pulmonaire soit une des étapes moléculaires intermédiaires entre le comportement fumeur et le cancer pulmonaire.

De par sa nature réversible, la méthylation représente une altération qui peut être ciblée de façon thérapeutique. Ce fait pourrait donc être exploité dans la chimioprévention (chez les patients à risque pour le cancer pulmonaire, par exemple, les fumeurs) et peut-être même dans la chimiothérapie.

Une avenue de recherche intéressante serait l'étude de l'hypométhylation exogène et gène-spécifique par le véhicule de la thérapie génique. En ce sens, des inhibiteurs de la méthylation, tel l'analogue non-méthylable de la dCTP, la 5-azadésoxycytidine, seraient dirigés vers le promoteur d'un gène spécifique, par exemple le gène *RAR*β, où une expression réduite ainsi qu'un état d'hyperméthylation seraient préalablement déterminés/diagnostiqués. Aussi, une hypométhylation globale thérapeutique pourrait être envisagée, mais en ciblant des agents anti-hyperméthylation (par exemple, des désoxynucléotides antisens de la méthyltransférase) vers des cellules cancéreuses préférentiellement. La raison d'être du ciblage gène-spécifique et/ou cellule néoplasique-spécifique est que l'effet global d'hypométhylation pourrait causer l'activation d'un ou de plusieurs oncogènes en activant son/leur expression. En effet, les inhibiteurs de méthylation genre 5-azacytidine et 5-aza-2'-désoxycytidine ont des effets carcinogènes documentés. Finalement, cette thérapie pourrait être combinée à une administration d'acide rétinoïque afin d'augmenter, peut-être de façon synergistique, le niveau d'expression de RARβ2.

INDEX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

A	Adenine
ACTH	Adenocorticotrophic hormone
AF-1 & 2	Activation factor -1 & 2
ADH	Antidiuretic hormone
APC	*Adenomatous polyposis coli* gene
AR	Androgen receptor
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
*at*RA	*all-trans* retinoic acid
5-azaC	5-azacytidine
5-azadC	5-aza-2'-deoxycytidine
BAL	Bronchioalveolar lavage
bp	base pairs
BP	Benzopyrene
BPE	Bovine pituitary extract
BSA	Bovine serum albumin
C	Cytosine
C-	Carboxy-
CC	cubic centimeter
CCK	Cholecystokinin
CDKN2/p16	*Cyclin-dependent kinase inhibitor 2* gene (*p16*)
CEA	Carcinoembryonic antigen
$-CH_3$	Methyl group
CH_4	Methane
cm	centimeter
CO_2	Carbon dioxide
COUP-TF I & II	Chicken ovalbumin upstream promoter-TF I & II
CpG	Cytidine-Guanosine dinucleotide
*c*RA	9-*cis* retinoic acid
CRABP I & II	Cellular retinoic acid binding protein I & II
CRBP I & II	Cellular retinol binding protein I & II
DBD	DNA binding domain
DCC	*Deleted in colon cancer* gene
ddH_2O	Double distilled water

DMSO	Dimethyl sulfonyl oxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide-5'-triphosphate
dTTP	deoxythymidine-5'-triphosphate
EDTA	Ethylene diamine tetraacetic acid
EGF	*Epidermal growth factor* gene
ER	Estrogen receptor
*erb*B	*Erythroblastosis* B gene
EtOH	Ethanol
FCS	Fetal calf serum
f	Frequency of occurrence
FHIT	*Frequently hypomethylated in tumours* gene
G	Guanine
GDP	Guanosine diphosphate
GR	Glucocorticoid receptor
GRP	Gastrin releasing peptide
GTP	Guanosine triphosphate
H	Hydrogen atom
ha-*ras*	Harvey *ras* gene
HCl	Hydrochloric acid
H_2O	Water
H_2O_2	Hydrogen peroxide
HNF4	Hepatocyte nuclear factor 4
HRE	Hormone response element
kb	Kilobase
KCl	Potassium chloride
kDa	Kilodalton
ki-*ras*	Kirsten *ras* gene
KOAc	Potassium acetate
K-SFM	Keratinocyte-Serum Free Medium
LB	Luria Bertani
LBD	Ligand binding domain
LOH	Loss of heterozygosity
M	Molar
5-mC	5-methylcytosine
MeTase	Methyl Transferase

min.	minute
mL	milliliter
mm	millimeter
mM	millimolar
MCC	*Mutated in colon cancer* gene
mg	Milligram (10^{-3}g)
Mg^{++}	Magnesium divalent cation
$MgCl_2$	Magnesium chloride
MHC	Major Histocompatibility Complex
mol.wt.	Molecular weight
Mϕ	Macrophage
MR	Mineralocorticoid receptor
MRED	Multiple Restriction Enzyme Digest
myc	*Myelocytoma* gene
N-	Amino-
Na	Sodium
NaAc	Sodium acetate
NaCl	Sodium chloride
NaOH	Sodium hydroxide
$NaHSO_3$	Sodium bisulphite
N-CoR	Nuclear receptor corepressor
NFκB	Nuclear factor κB
ng	Nanogram (10^{-9}g)
NGFI-B	Nerve growth factor I-B
NMU	Nitrosomethyl urea
NNN	N'-nitrosonornicotine
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
NSCLC	Non small cell lung carcinoma
O	Oxygen atom
O_2	Oxygen molecule
O_2^-	Superoxide anion
Oligo	Oligodeoxynucleotide
O/N	Overnight
p	Chromosomal short arm
p53	*Protein 53* kDa gene
PAH	Polycyclic aromatic hydrocarbon

PBS	Phosphate buffered saline
pBS KS	Plasmid BS KS (Bluescript-based)
PCR	Polymerase chain reaction
pg	picogram (10^{-12} g)
pH	Log_{-10} of the hydrogen ion concentration
PPAR*a*	Peroxisome proliferator activator receptor *a*
PR	Progestin receptor
Pu	Purine (Adenine or Guanine)
Py	Pyrimidine (Cytosine, Thymine or Uracil)
q	Chromosomal long arm
RAL	Retinal
RAR	Retinoic acid receptor
RARE	Retinoic acid response element
Rb	*Retinoblastoma* gene
RNAase	Ribonuclease
ROL	Retinol
RPM	Revolutions per minute
RT	Room temperature
RXR	Retinoid X receptor
S.A.M.	S-adenosylmethionine
SCLC	Small cell lung carcinoma
SDS	Sodium dodecyl sulphate
SF-1	Steroidogenic factor-1
SMRT	Silencing mediator for RAR and TR
STP	Signal transduction pathway
T	Thymine
TA	Thymine-Adenine
TAE	Tris-acetic acid-EDTA
TE	Tris-EDTA
ter	Chromosomal terminus
TF	Transcription factor
TIF	Transcription intermediary factor
TK	Tyrosine kinase
TPA	12-O-tetradecanoyl phorbol-13-acetate
TpG	Thymine-Guanine dinucleotide
TR	Thyroid hormone receptor

Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
U	Unit of enzymatic activity
UV	Ultra violet
VAD	Vitamin A deficiency
VDR	Vitamin D_3 receptor
VHL	*Von Hippel Lindau* gene
V	Volt
WHO	World Health Organization
wt	Wild type
X-Gal	5-Bromo-4-chloro-indolyl-β-D-galactoside
α	alpha (*Greek*)
β	beta (*Greek*)
γ	gamma (*Greek*)
κ	kappa (*Greek*)
μg	microgram (10^{-6} gram)
μm	micron (10^{-6} meter)
μM	micromolar
μL	microliter
-/-	Homozygous null mutant
-/+	Heterozygous null mutant
3'	Downstream through DNA sequence
5'	Upstream through DNA sequence
°C	Degree Celcius
%	Percent

ACKNOWLEDGEMENTS

I would like to thank...

My mom and dad, Sophie and Jean, for their intelligent advice, their experience, their devotion, their patience and their unflagging moral support.

My sister, Catherine, for her abundant encouragements, faith and so-called "gut-instincts" throughout all my endeavours.

Dr. Ted Bradley, my Master's supervisor and mentor, for his endless support, his enthusiastic interest, and the sincere dedication with which he approaches everything in life, including his profession.

Dr. Benoit Houle, for having built the foundation for this project (and many more), and for his constant feedback, patience and moral support.

André Toulouse, for having initiated my interest for this project, and for his constant availability and aid throughout.

Johane Morin, for her extraordinary technical know-how and speed-of-light assistance, but mostly for her lucidity regarding many different aspects of life.

Miguel Landers, for his steady and stable support, his sardonic humour and his sharp perception.

Edlie St-Hilaire, for her support, her patience, her gentle kindness, and her serene and wise outlook on life.

Graham Dellaire, for his extremely unique and helpful insights, but mostly for the ease with which he shares ideas.

Dr. Li, for his competence and availability.

For
E
N
F
and
B

I. INTRODUCTION

A. LUNG PHYSIOLOGY

1. STRUCTURE

1.1 Gross anatomy of the lungs

The left and right lungs comprise two and three lobes, respectively (Figure 1). They are located in the thoracic cavity, within the rib cage. The rib cage comprises the ribs which are joined by intercostal muscles and connective tissue. Beneath the lungs is the diaphragm, a sheet of skeletal muscle. Surrounding the lungs are sheets of epithelial tissue known as the pleura. The parietal pleura lines the thoracic cavity; the visceral pleura envelopes the lungs. In order to minimize friction during respiration, the pleural tissues secrete a lubricant which fills the intrapleural space.

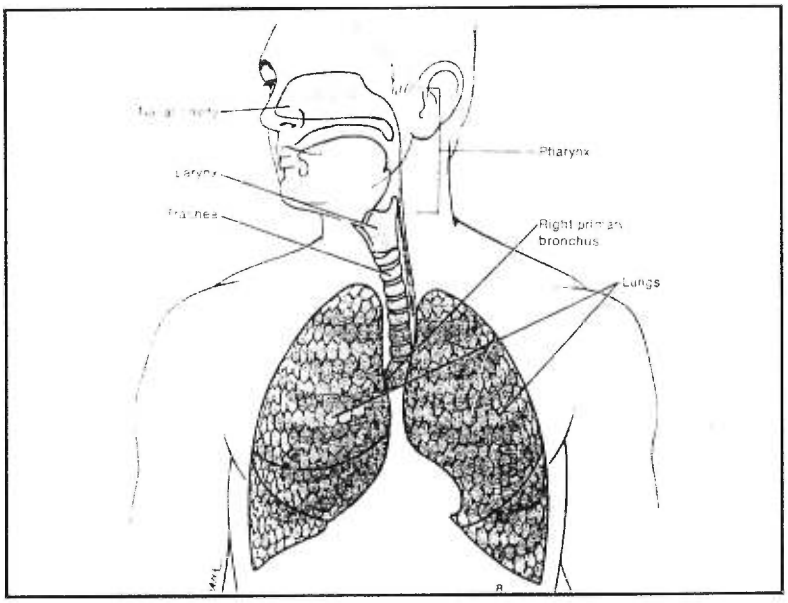

Figure 1: Diagram of the organs of the respiratory system in relation to their surrounding structures (taken from Tortora and Anagnostakos, 1987).

1.2 Gross anatomy of the airways

The airways comprise the naso-pharynx, the larynx, the trachea, the left and right primary and secondary bronchi, the segmental bronchi and the bronchioles (Figure 2). The multiple series of segmental bronchi (approximately 12 generations) supply distinct regions of the lobes with inhaled air. The final branchings result in approximately 150 000 terminal bronchioles (Schauff *et al.*, 1990). The terminal bronchioles which are smaller than 0.5 mm in diameter contain alveoli in their walls and are known as respiratory bronchioles.

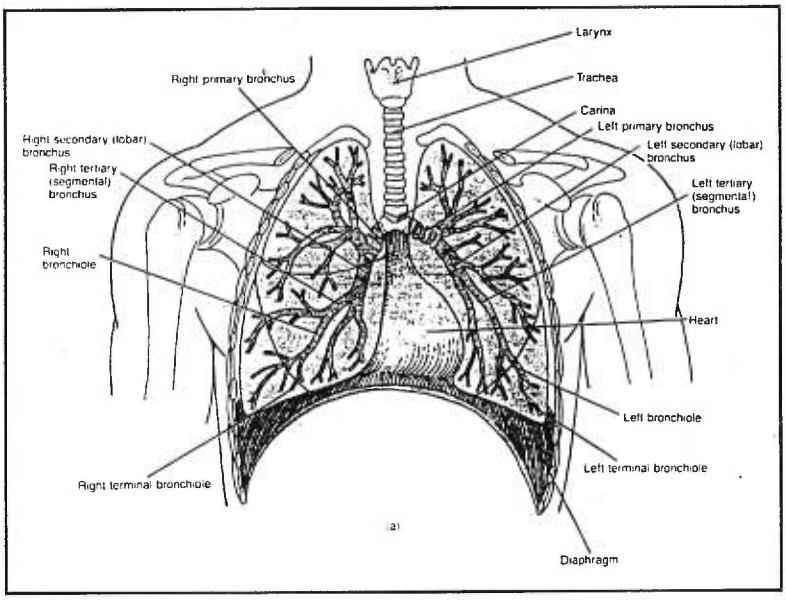

Figure 2: Diagram of the respiratory airway and the bronchial tree (taken from Tortora and Anagnostakos, 1987).

1.3 Microscopic anatomy of the respiratory bronchioles

The respiratory bronchioles proximal to the last generation of segmental bronchi are composed of a simple layer of cuboidal epithelial cells. As they develop, they proceed distally and differentiate into squamous epithelial cells. These respiratory bronchioles branch out into 2-11 alveolar ducts, each (Tortora and Anagnostakos, 1987). These ducts

are surrounded by approximately 20 alveolar sacs, each (Schauff *et al.*, 1990). An alveolar sac (Figure 3) is made of epithelial cells known as pneumocytes and supported by a thin elastic basal membrane. Type I pneumocytes are large and membranous; type II pneumocytes are small and granulous. Type II pneumocytes secrete a phospholipid substance known as surfactant which decreases superficial tension. Among the cell types also contained within the alveolar sac are alveolar macrophages, monocytes, leukocytes and fibroblasts. A monolayer of endothelial cells creates the capillary network from which gaseous exchanges are conducted via diffusion between the lungs and the blood.

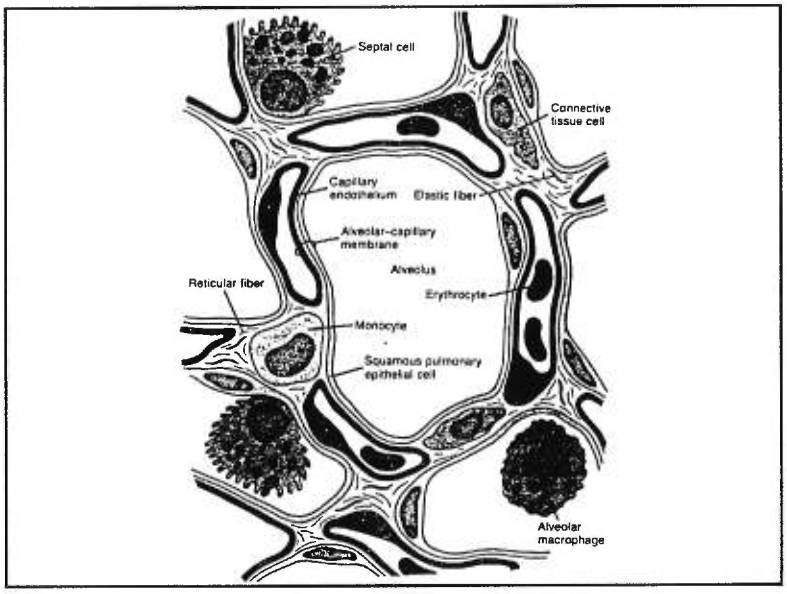

Figure 3: Detailed diagram of an alveolar sac (taken from Tortora and Anagnostakos, 1987).

2. FUNCTION

2.1 Basic functions of the respiratory system

The lung is the organ which links the circulatory system to the respiratory system, allowing ventilation (inhalation and exhalation) to take place. The primary function of the respiratory system is the gaseous exchange between the atmosphere and the circulatory system, through external respiration, and between the circulatory system and the cells,

through internal respiration. Carbon dioxide (CO_2) is excreted and oxygen (O_2) is replenished via diffusion across the alveolar-capillary membrane (Figure 4).

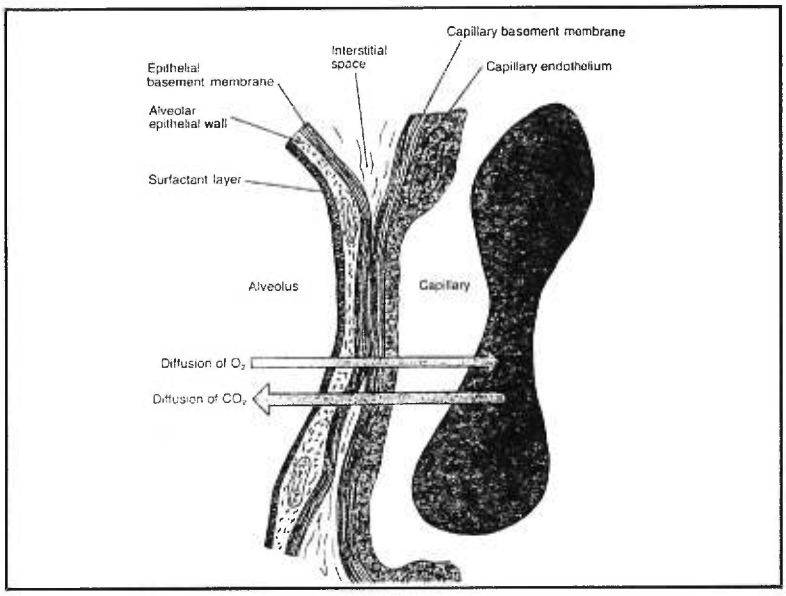

Figure 4: Diagram of the alveolar-capillary membrane (taken from Tortora and Anagnostakos, 1987).

2.2 Protection against inhaled particulate materials

Upon inhalation, the lungs are introduced to a variety of different particulate materials, such as dust, pollen, spores, viruses, and bacteria. The airways possess several different mechanisms to ensure constant and maximal protection. First, the intake of particles larger than 5-10 microns in diameter is impeded by cilia and mucus in the nasal passages. Second, sneezing and coughing are reflexes instigated by sensory nerve endings along the nasal passages, the trachea and the bronchi. Third, epithelial and glandular cells lining the airways give rise to the mucociliary elevator mechanism. This is a continuous secretion and upward moving flow of accumulated mucus and entrapped particles towards the pharynx from which it will be expulsed. The mucociliary process is inhibited by tobacco smoke, cold air and certain chemicals. Fourth, alveolar macrophages phagocytose dust and debris. Finally, pulmonary lymphocytes provide the lungs with immunity against antigenic pathogens and their products.

2.2.1 Common injurious materials

Particles smaller than 2-5 microns in diameter may reach the alveoli and never be expulsed. In industrialized countries, coal, asbestos, radon, silica and textile mill products are common causes of lung fibrosis, or permanent scarification of the lung. However, even prior to visible effects of the environment on the lung, daily exposure to ubiquitous noxious agents such as carcinogens takes place. Long term exposure has important effects in pulmonary pathology.

B. LUNG CANCER

1. PATHOLOGIES OF THE LUNG

Diseases of the lung can be categorized into eight different categories: (1) atelectasis (alveolar collapse); (2) pediatric lung diseases; (3) obstructive lung diseases; (4) restrictive lung diseases; (5) vascular lung diseases; (6) pulmonary infections; (7) benign pleural and upper respiratory lesions; and (8) benign and malignant lung tumors (Lipscomb, 1992). Lung cancer is presently the most important cause of cancer-related deaths in men and women in North America.

2. BRONCHOGENIC CARCINOMAS

Approximately ninety-five percent of primary lung tumors are bronchogenic carcinomas: they originate from the bronchial epithelium (Figure 5). The remaining 5% of lung tumours consist of bronchial carcinoids, mesotheliomas, bronchial gland neoplasms, mesenchymal malignancies and lymphomas, and benign lesions (Lipscomb, 1992).

Figure 5: Photograph of a bronchogenic carcinoma (taken from Kumar *et al.*, 1992).

3. LUNG CANCER HISTOLOGIES

Lung cancers are classified according to their histological properties (Table I). The four main classes of epithelial bronchogenic carcinomas are: (1) squamous cell (epidermoid) carcinoma; (2) small cell carcinoma; (3) adenocarcinoma; and (4) large cell carcinoma (WHO, 1981).

A.	BENIGN
	1.Papillomas
	a. Squamous cell papilloma
	b. Transitional papilloma
	2.Adenomas
	a. Pleomorphic adenoma ("mixed tumour")
	b. Monomorphic adenoma
	c. Others
B.	DYSPLASIA
	Carcinoma *in situ*
C.	MALIGNANT
	1. Squamous cell carcinoma (epidermoid carcinoma)
	Variant:
	Spindle cell (squamous) carcinoma
	2. Small cell carcinoma
	a. Oat cell carcinoma
	b. Intermediate cell type
	c. Combined oat cell carcinoma
	3. Adenocarcinoma
	a. Acinar adenocarcinoma
	b. Papillary adenocarcinoma
	c. Bronchiolo-alveolar carcinoma
	d. Solid carcinoma with mucus formation
	4. Large cell carcinoma
	Variants:
	a. Giant cell carcinoma
	b. Clear cell carcinoma

Table I: Histological classification of epithelial bronchogenic carcinomas (based on WHO, 1982).

Small cell lung carcinomas (SCLCs) account for approximately 10% of all pulmonary malignancies (Ginsberg *et al.*, 1997). Squamous cell carcinomas (30%), adenocarcinomas (40%) and large cell carcinomas (15%) are distinct from SCLCs histologically and clinically. They are thus grouped under the classification of non small cell lung carcinomas (NSCLCs). The remaining 5% are of mixed histologies.

Bronchogenic carcinomas are normally aggressive, invasive, producers of a variety of bioactive products and widely metastasizing. They usually arise in the epithelial lining of the major bronchi (Kumar *et al.*, 1992).

Lung cancers of the four main histologies frequently exhibit phenotypes characteristic of two or more histological classifications. This heterogeneity, though not unexpected, hinders the accuracy of prognosis. Moreover, understanding the basic science behind the development of a lung cancer requires understanding its histogenesis. Unfortunately, it is not known whether all bronchogenic carcinomas arise from a common precursor stem cell or from precursor stem cells of different lineages. Though one school of thought claims that all lung cancers share a common endodermal origin (Vinocour and Minna, 1989), another school of thought claims that the different phenotypes exhibited by the various carcinomas are indicative of different precursor stem cells. Such a postulation hypothesizes, for example, the neurectodermal origin of SCLCs. This is because the typical secretion of different neuroendocrine factors by neuroendocrine tumors such as SCLCs is a feature of neuroendocrine cells located in the basal mucosa. These cells are known to differentiate from the ectoderm, not the endoderm.

Nonetheless, what is known is that all bronchogenic carcinomas are associated to cigarette smoking. SCLCs and squamous (epidermoid) cell lung carcinomas are the most closely correlated with cigarette smoking.

3.1 Squamous cell carcinoma (epidermoid carcinoma)

Squamous cell carcinoma (Figure 6) usually arises in the major bronchi, in the central region of the lungs (Lipscomb, 1992). It originates as a carcinoma *in situ*: malignant squamous cells (approximately 20 to 30 μm in diameter) replace the normal squamous epithelium without invading the basement membrane. However, as the malignancy progresses, invasion of the basement membrane occurs causing obstruction of the lumen. Squamous cell carcinoma may comprise poorly, moderately or well differentiated epithelial cells and stains positive for keratin upon biopsy and subsequent immunohistochemistry (Ginsberg *et al.*, 1997). These tumours grow relatively slowly and can be detected at their earliest form from cytological examination of bronchioalveolar lavage (BAL). For these reasons, squamous cell carcinomas tend to have the best prognoses.

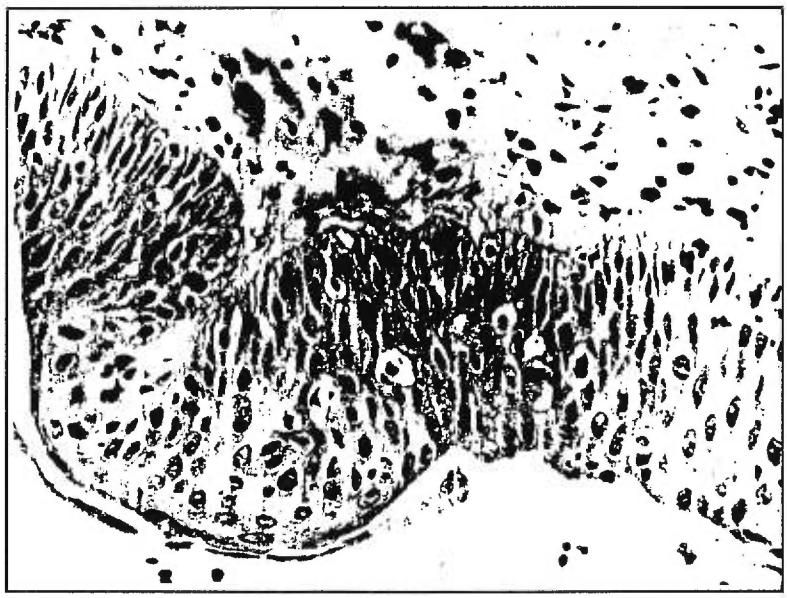

Figure 6: Light micrograph of a moderately differentiated squamous cell carcinoma. Cross-section: Hematoxylin + Eosin stain, magnification: x180 (taken from WHO, 1981).

3.2 Adenocarcinoma

Adenocarcinoma (Figure 7) usually arises from alveolar epithelium or bronchial mucosal glands (Ginsberg *et al.*, 1997) in the periphery of the lungs. This carcinoma forms adenoid structures (i.e. glands) which may secrete mucin. These structures are revealed upon light microscopy. However, adenocarcinomas also stain positively for carcinoembryonic antigen (CEA) and keratin, thus facilitating the immunohistochemical analysis. Adenocarcinoma has a worse prognosis than squamous cell carcinoma.

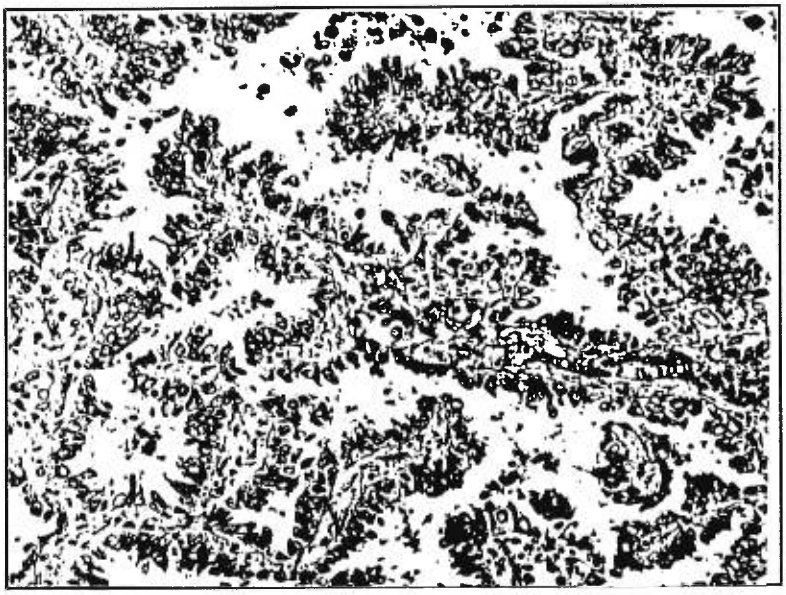

Figure 7: Light micrograph of a papillary adenocarcinoma. Note the prominent glandular structures. Cross-section: Hematoxylin + Eosin stain, magnification: x180 (taken from WHO, 1981).

3.3 Large cell carcinoma

Large cell carcinomas (Figure 8) are subdivided into two varieties: large cell and clear cell carcinomas. These cells are normally larger than other NSCLCs, approximately 30 to 50 μm in diameter, *ergo* their classification. Large cell carcinomas present neuroendocrine features previously solely attributed to small cell lung carcinomas, and/or may express features characteristic of squamous cell carcinomas and/or adenocarcinomas. These tumors have yet a worse prognosis than adenocarcinomas and squamous cell carcinomas.

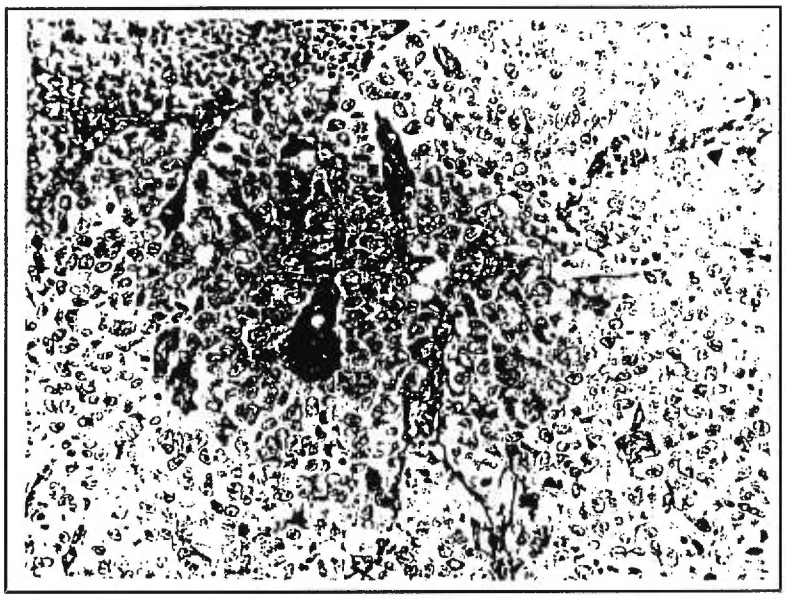

Figure 8: Light micrograph of a large cell carcinoma. Cross-section: Hematoxylin + Eosin stain, magnification: x180 (taken from WHO, 1981).

3.4 Small cell lung carcinomas

Small cell lung carcinomas (Figure 9) present characteristic cytological features such as: (1) uniformly small cells (10-20 μm in diameter); (2) large and round nuclei; and (3) neuroendocrine granules. Moreover, they express and secrete several different neuroendocrine factors, such as adenocorticotrophic hormone (ACTH), gastrin-releasing peptide (GRP), antidiuretic hormone (ADH) and cholecystokinin (CCK). Though they respond well to early chemotherapeutic and radiological treatments, they grow rapidly and metastasize early. Thus, they are often diagnosed late in their progression and are characterized by very poor prognoses.

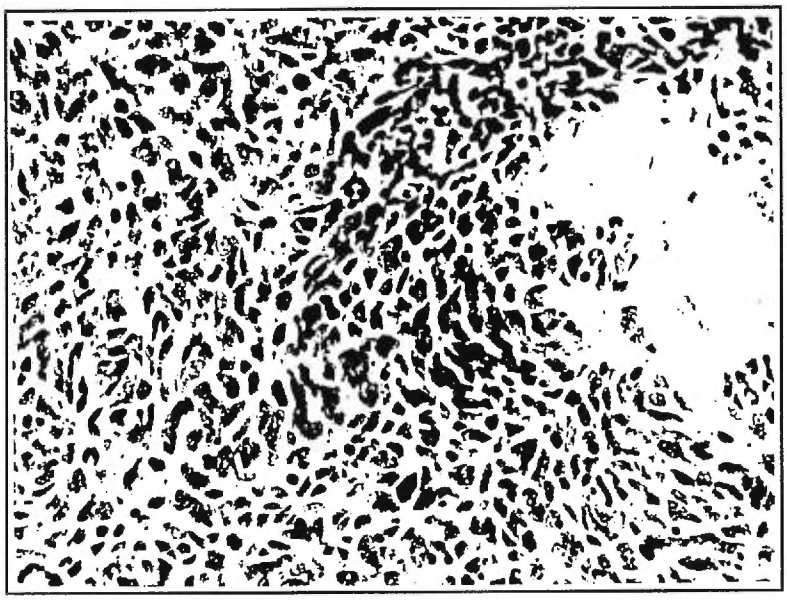

Figure 9: Light micrograph of an oat cell carcinoma. Cross- section: Hematoxylin + Eosin stain, magnification: x360 (taken from WHO, 1981).

4. RECENT CANADIAN STATISTICS ON LUNG CANCER

Bronchogenic carcinoma is the major cause of cancer-related deaths in Western societies, including Canada. Moreover, lung cancer is a particularly aggressive type of cancer with an overall 5-year survival of less than 10-15%.

It has been estimated that in 1997, approximately 130 800 new cases of cancer will have been diagnosed in Canada, population 29.9 million (Canadian Statistics on Cancer, 1997). Of these, 20 300 or 15.5% will have been lung cancer cases. That same year, it is estimated that 60 700 deaths will have been attributable to cancer. Of these, 16 900 or 27.8% will have been due to lung cancer, specifically. Approximately 71% of new cancer cases and 80% of cancer deaths will involve people over 60 years of age.

Cancer type	No. New cases, estimated 1997	No. Deaths, estimated 1997	Ratio deaths/cases, estimated 1997
All			
-Total	130, 800	60, 700	0.46
-Women	60, 600	28, 100	0.46
-Men	70, 200	32, 600	0.46
Lung			
-Total	20, 300	16, 900	0.83
-Women	7, 800	6, 300	0.81
-Men	12, 500	10, 600	0.85
Colorectal			
-Total	16, 400	5, 900	0.36
-Women	7, 500	2, 700	0.36
-Men	8, 900	3, 200	0.36

Table II: Estimated number of new cases and deaths according to cancer type and sex, Canada, 1997 (based on Canadian Statistics on Cancer, 1997).

According to the ratio "mortality/incidence", lung cancer has the third worst prognosis at 0.83. It is preceded only by esophageal cancer, 1.01, and pancreatic cancer, 0.98. (Ratios greater than 1.00 are due to incomplete registration of new cases prior to deaths, or incorrect declaration of numbers of new cases or deaths). Moreover, lung cancer is the leading type of newly diagnosed cancer cases according to the percentage "(new lung cancer cases)/(new cancer cases) x 100%", at 15.5%. It is followed by prostate cancer, 15.1%, and breast cancer, 14.1%. Most strikingly, it is the leading cause of cancer-related deaths according to the percentage of "(lung cancer deaths)/(cancer deaths) x 100%", at 27.8%. This represents 16 900 Canadians in 1997. The second and third leading causes of cancer related deaths are colorectal cancer at 9.7% or 5 900 Canadians, and breast cancer at 8.4% or 5 100 Canadians.

Lung cancer incidence (Figure 10) reached its peak in 1984 for men, nearing 100 new cases per 100 000 inhabitants. However, lung cancer has been steadily increasing at a rapid rate in women, reaching 45 new cases per 100 000 in 1997. In parallel, the number of lung cancer mortalities has decreased steadily for men, but continues to increase for women. It is generally accepted that the main cause behind these trends are the increased consumption of tobacco products since the middle of the 20th century. In fact, the main cause of lung cancer is tobacco smoking, mostly due to the fact that of all sources of carcinogens, it is the most greatly inhaled (Osann *et al.*, 1991). Historically, the two major social causes behind the great increase in incidence of lung cancer are: (1) the automated production of cigarettes, and (2) the grand-scale distribution of these products by the American Red Cross to World War II soldiers.

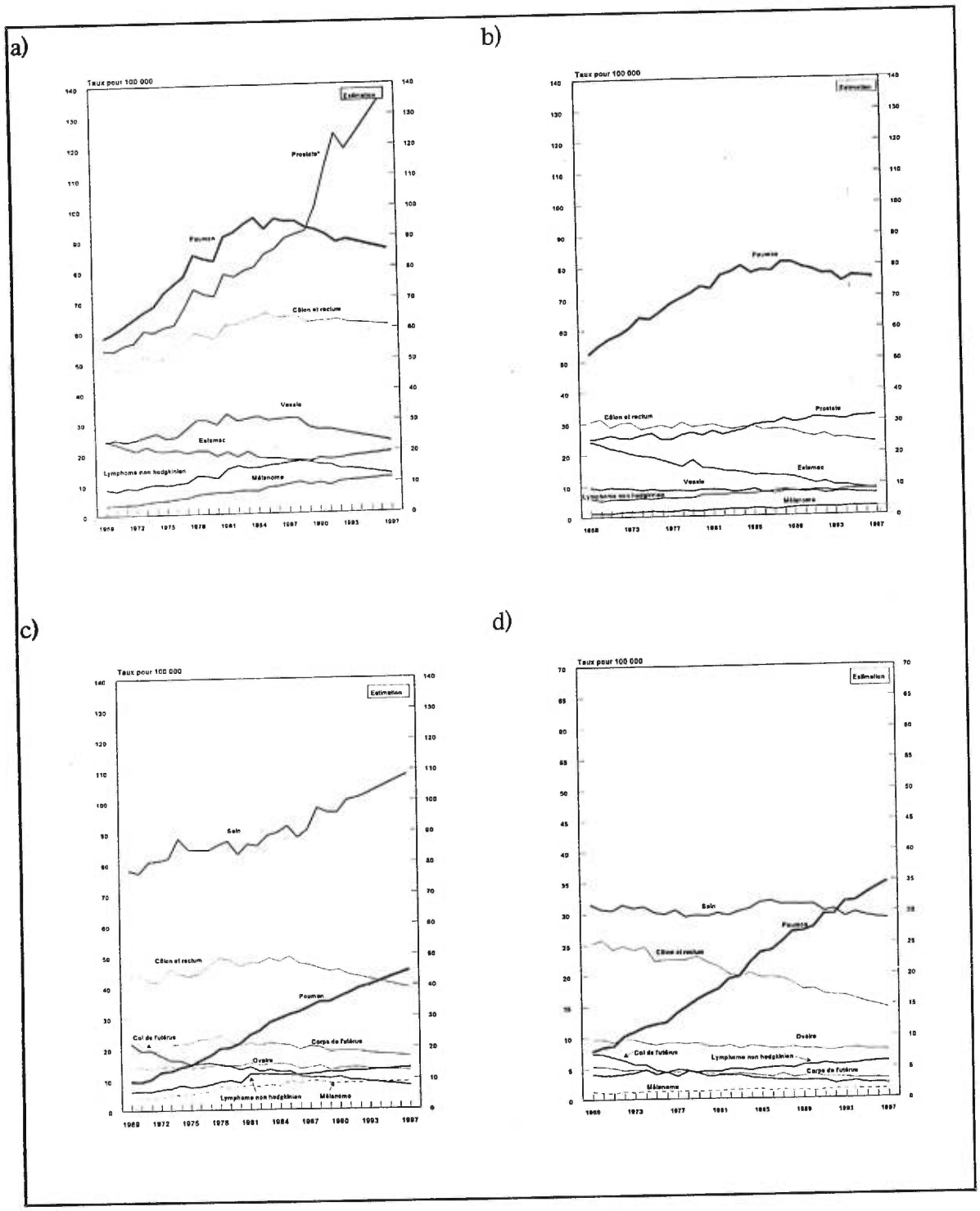

Figure 10: Comparative rates of cancer incidence and mortality according to cancer type and gender, Canada, 1969-1997 (taken from Canadian Statistics on Cancer 1997).
a) Incidence; men; **b)** Mortality, men; **c)** Incidence, women; **d)** Mortality, women.

C. LUNG CANCER RISK FACTORS

1. TOBACCO SMOKING

In the United States, approximately 80% of lung cancer deaths in men and 75% of lung cancer deaths in women are attributable to smoking per year (Ginsberg *et al.*, 1996). This represents 65,000 men and 27,000 women dying from lung cancer due to a highly avoidable risk factor: cigarette smoking. As expected, the number of lung cancer-related deaths progresses at approximately the same rate as the number of smokers (Figure 11), with an intercalation or delay of approximately 20 years (Canadian Statistics on Cancer, 1991). This delay represents the latency or long-term nature of the carcinogenic process.

It is important to note that tobacco smoking is not only associated with lung cancer: it is also associated with many other cancers, such as mouth, pharynx, larynx, esophagus, urinary bladder, pancreas and kidney cancers (Ruddon, 1995). Moreover, it is associated with many other non-cancer diseases, such as chronic bronchitis and emphysema, two major obstructive lung disease (Lipscomb, 1992). The pathogenesis of some of these diseases may increase the risk for lung cancer in smokers and newly non-smokers. For instance, a major effect of the loss of elastic recoil in emphysema, is the decrease of expiratory flow rate. Due to this decrease, exposure of the emphysematous lung to the carcinogens contained within tobacco smoke will increase. In the case of the newly non-smoking emphysema patient, the damage already incurred may increase risk of lung cancer in several ways. For example, the greater numbers of macrophages and neutrophils in the emphysematous lung will lead to abnormally high levels of free radicals such as hydrogen peroxide, a metabolic by-product of these phagocytes. Free radicals are deoxyribonucleic acid (DNA) damaging agents, capable of causing transmissible carcinogenic damage to the hereditary component of the cell (i.e. parent-cell to daughter-cell).

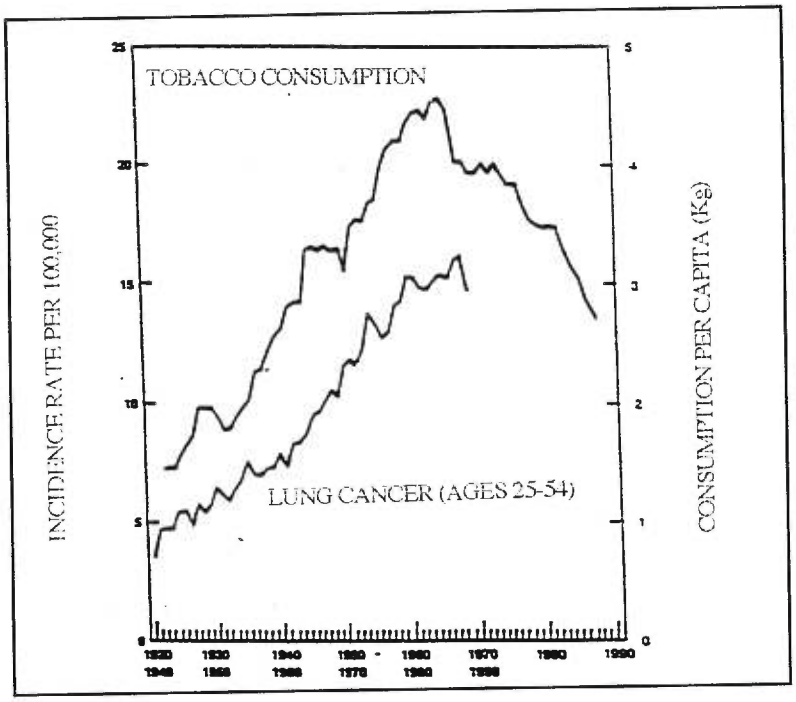

Figure 11: Graph of number of smokers and lung cancer deaths in the United States from 1920 to 1991 (taken from Canadian Statistics on Cancer, 1991).

1.1 Smoking behaviour

The risk of developing lung cancer is positively correlated to several aspects of smoking behavior: (1) the number of cigarettes smoked; (2) the duration of the smoking habit; (3) the early onset or age of commencing the habit; (4) the depth of inhalation; (5) the amount of nicotine contained in the cigarette; and (6) the use of unfiltered cigarettes. Moreover, recent epidemiological studies have suggested that dose for dose, women are more susceptible to cigarette smoke carcinogens than men (Begg *et al.*, 1995; Dwyer *et al.*, 1994; Osann *et al.*, 1993; Risch *et al.*, 1993). Another study has confirmed that the gender differences observed cannot be attributed to differences in base-line exposure, smoking history or body mass, but are likely due to intrinsic differences in the susceptibility to carcinogens (Zang and Wynder, 1996).

1.2 Tobacco smoke

Tobacco combustion (i.e. tobacco smoke) produces two kinds of product phases; (1) the particulate phase and (2) the gaseous phase (Minna *et al*, 1997). The chemical and biochemical analyses of these phases reveal that tobacco combustion yields more than 3 000 different chemical compounds. Of these, approximately a dozen are major DNA damaging agents (Table III).

PHASE	SUBSTANCE
PARTICULATE PHASE	
Neutral fraction	1. Benzo [a] pyrene
	2. Dibenzo [a] pyrene
Basic fraction	3. Nicotine and nicotine derivatives
	-NNN
	-NNK
	4. Tobacco-specific nitrosamines
Acidic fraction	5. Cathecol
	6. Unidentified
Residue	7. Nickel
	8. Cadmium
	9. ^{210}Polonium
GASEOUS PHASE	10. Hydrazine
	11. Vinyl chloride

Table III: Major DNA damaging agents in tobacco smoke (based on Minna *et al.*, 1997 and Castonguay, A., *et al.*, 1985). **NNN** = N'-nitrosonornicotine; **NNK** = 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone.

1.3 DNA damaging agents and the multistep process of carcinogenesis

DNA damaging agents are mutagenic under certain conditions (eg. faulty DNA repair). Heritable genetic mutations are the foundation of the theory of multistep carcinogenesis first proposed by Bert Vogelstein in 1990 (Figure 12). This theory suggests that progressively dysfunctional multi-gene expression (due to the accumulation of a number of various types of genetic alterations over time) causes dysfunctional cellular and sub-cellular organization, a hallmark of tumours. Examples of different levels of organization which may be affected are: growth rate, differentiation and localization.

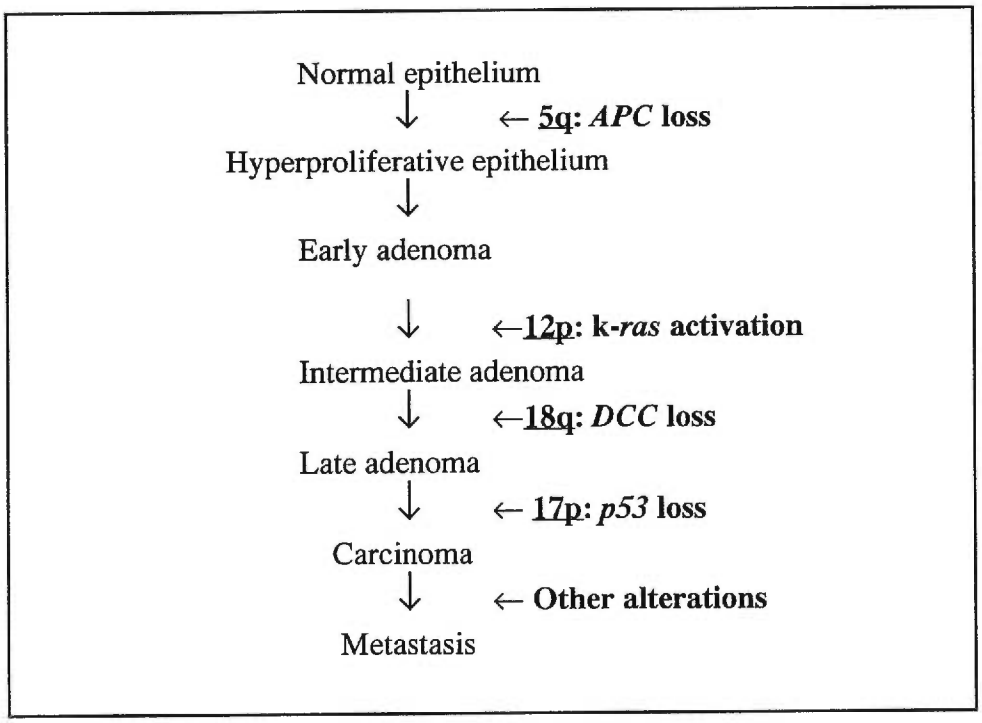

Figure 12: Model for colorectal tumorigenesis highlighting the multistep accumulation of genetic lesions (based on Vogelstein and Kinzler, 1993, and Kinzler and Vogelstein, 1996).
Chromosomal localization underlined, gene in *Italics*.
APC = *Adenomatous polyposis coli* gene; ***DCC*** = *Deleted in colon cancer* gene.

1.4 Tumor initiation and promotion

The theory of tumor initiation and promotion is an attempt to dissect the development of a cancer, from its non-cancerous beginning to its metastatic and invasive end, according to the biological events that characterize each phase of this progression (Figure 13). The mouse model, where skin tumours are induced with carcinogenic products such as nitrosomethyl urea (NMU) or 12-O-tetradecanoyl phorbol-13 acetate (TPA), is the historical foundation of this theory. Rodent lung cancer models commonly use benzo [a] pyrene for this purpose, because it is a potent carcinogenic agent associated with lung cancer.

First, tumour initiation is caused by a short exposure to an initiating carcinogen. The resulting DNA damage is irreversible (Reddel *et al.*, 1989). Moreover, when affecting somatic cells, it is inherited by daughter cells upon division. A typical type of initiating event is the activation of an oncogene. The initiating activity of the carcinogen may be enhanced by factors that increase its effective dose, such as tissue retention of the carcinogen. Tumour initiation requires promotion to result in a tumour. Such promotion occurs when a second event gives a selective advantage to the initiated cells, but not to the un-initiated cells, and results in the clonal expansion of the initiated cells. Unlike tumour initiation, tumour promotion is reversible. Initiating and promoting carcinogens and cocarcinogens have been identified experimentally (eg. benzo [a] pyrene and estrogen, respectively).

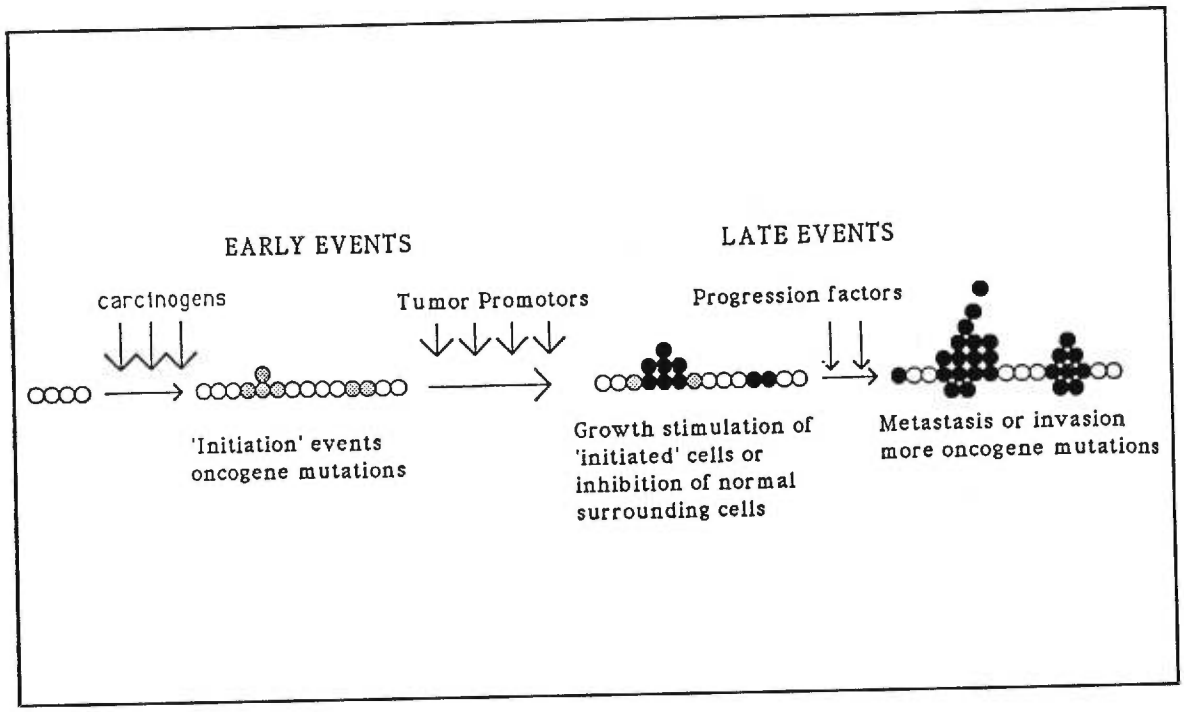

Figure 13: Events in the development of a cancer (taken from Birrer, M.J. *et al.*, 1992).

1.5 Polycyclic aromatic hydrocarbons

The neutral fraction of tobacco smoke contains polycyclic aromatic hydrocarbons (PAHs) such as benzopyrenes (BPs) (Minna *et al.*, 1997). PAHs are thought to have the most tumorigenic activities of all tobacco combustion substances. Some of these compounds require an enzymatically catalyzed metabolism to become activated. The enzymes responsible for such deleterious conversions, such as α-hydroxylases, are normal enzymes contained within various tissues of the human body, including the lung. Once activated, the carcinogen forms covalent adducts with several different types of macromolecules, especially DNA. These adducts are responsible for the mutagenic process. Whether or not these DNA alterations are the precise causes of specific tumours has not yet been shown. A relevant mutation in the study of lung cancer is the guanine to thymidine (G→T) transversion caused by activated benzopyrenes. This type of DNA alteration is a common point mutation of the mutant *ras* gene (Bos, 1989), which is frequently mutated in lung tumours (Slebos and Rodenhuis, 1989).

Tobacco smoke has initiating and promoting activities in lung carcinogenesis (Minna *et al.*, 1997). Smoking duration is directly correlated to risk of lung cancer. However, the carcinogenic activities of tobacco smoke are reversible, since cessation of the smoking habit decreases the risk of lung cancer (Reddel *et al.*, 1989). The period required to return lung cancer risk to the normal risk of a non-smoker is not precisely known, and probably varies significantly from one person to the next.

1.6 Catechols

Catechols block the detoxification process by inhibiting benzo [a] pyrene catabolism. Moreover, catechol metabolic by-products are hydrogen peroxide (H_2O_2) and superoxide anion (O_2^-) (Hoffmann *et al*, 1985). Aldehydes and peroxides have promoting activities (Reddel *et al.*, 1989). Studies have shown that these compounds selectively induce differentiation of normal bronchial epithelial cells and decrease their clonal growth rate (Willey *et al.*, 1984). These same compounds have but minor effects on transformed bronchial epithelial cells.

2. NON-SMOKING LUNG CANCER RISK FACTORS

Though approximately 80% of lung cancers are attributable to cigarette smoking, only 10 to 15% of smokers develop lung cancer (Stanley and Stjernsward, 1989). This is because there are several other factors associated to the risk of developing lung cancer. They are: (1) occupational, professional and medical exposure to various chemicals and ionizing radiation sources; (2) nutritional deficiencies in antioxidant micronutrients; (3) immunocompromising conditions and diseases; (4) genetic predisposition; and (5) increasing age. Compared to smoking, these factors are minor.

2.1 Exposure to carcinogens

Two hundred years ago, diseases of the lung were so scarce, that they prompted medical curiosity. Today, lung cancer alone is a primary concern of American, Canadian and European epidemiological studies. The epidemic proportions of lung cancer are

without a doubt mainly associated with tobacco consumption. However, technological advance, urbanization, modernization and war are also significant contributors to the increased incidence of lung cancer. Many epidemiological and experimental studies have shown that certain substances and radiations are associated with lung cancer (Table IV).

SUBSTANCE	EXPOSURE
Arsenic	Smelters, pesticide manufacturers, medical treatments
Asbestos	Miners, millers, insulation, shipyard, railroad, and cement workers
Chloromethyl ethers	Ion-exchange resin manufacturers
Chromium	Chromate and pigment manufacturers
Hydrocarbons	Coal tar/gas workers, roofers, coke producers; medical treatments
Mustard gas	Poison gas manufacturers
Nickel	Refiners
Radiation	Miners of uranium and other ores; cancer treatments

Table IV: Documented occupational, professional and medical exposure to lung cancer carcinogens (modified from Frank, 1989).

The main principles revealed from epidemiological studies of occupational carcinogens are: (1) the latency between onset of exposure and the overt development of the related cancer (approximately 20 years), and (2) the synergistic effect (multiplicative versus additive) on the risk for lung cancer when combined with smoking.

As Table IV illustrates, most of these carcinogens come in contact with humans through occupational, professional and medical practices. However, others come in contact with humans through more common practices of every day living. For example, radon is a naturally occurring product of uranium and it is present in soil and rocks and may accumulate in the basements of homes in certain areas (eg. the Reading Prong, Eastern United States).

Certain substances are particulate and take up long-term residence within the alveoli of the lung (eg. asbestos and beryllium) (Frank, 1989). Thus, once inhaled, occupational exposure need not be continuous to be carcinogenic. Others are gaseous and

appear to require continuous long term exposure (eg. chloromethyl ethers). Radiation, such as UV light and ionizing radiation, does not necessarily require long term exposure, because it has a dose-dependent effect (Ruddon, 1995). For example, ionizing radiation is frequently used to treat childhood cancers, and has been shown to increase their risk for cancer in adulthood tenfold (DeLaat and Lampkin, 1992).

2.2 Nutritional deficiencies in antioxidants

Antioxidant micronutrients are associated with a protective effect against lung cancer. Retinoids (metabolic derivatives of vitamin A), ascorbic acid (vitamin C), tocopherol (vitamin E) and selenium scavenge and trap free radicals and reactive oxygen molecules, highly unstable molecules capable of causing DNA damage. Moreover, carotenoids and ascorbic acid also stimulate the immune system.

Of particular interest are the retinoids, such as vitamin A. Due to its important effects on proliferation and differentiation of epithelial tissues, including lung tissue, vitamin A has long been studied in the framework of basic and clinical investigations of cancer. Many epidemiological studies have shown that it has anticancer activities. One study showed that dietary intake of vitamin A is inversely proportional to risk for lung cancer (Graham, 1984). The histological types of lung cancer most closely associated are squamous cell carcinoma and small cell carcinoma, the types of lung cancer also most closely associated to tobacco smoke.

2.3 Compromised immunity

The immune system grants immunity to all normal cells within the body by recognizing them as self rather than foreign. This is because they express antigenic determinants on their cell surface known as the major histocompatibility class (MHC) proteins. These are the proteins that allow specific recognition of self. However, tumor cells undergo various biochemical changes, including the production of new non-self antigenic determinants on their cell-surface. These antigenic determinants can be recognized as foreign and can elicit a humoral antibody response and a cellular cytolytic response from the immune system. However, a compromised immune system has a lower

probability of recognizing a tumour cell than does a healthy immune system. Unfortunately, findings both confirm and negate the theory of immune surveillance.

2.4 Genetic predisposition

As we've seen, carcinogen metabolites are the causative agents of some genetic mutations. Because the catabolic pathway of carcinogens is genetically pre-determined, the formation of malignant mutations may be partially due to genetic predisposition. Moreover, faulty DNA repair caused by faulty DNA repair genes, may also increase the probability for cancer by ineffectively repairing mutations.

2.5 Age

Almost all risk factors for lung cancer specifically, as well as other types of cancer, are potentiated by increasing age. First, genetic lesions are accumulated over time, and the ability to maintain and repair these lesions is decreased with aging. Second, the increasing probability of suffering from another ailment or disease which may contribute to lung cancer risk, increases with age. For example, Diabetes Type II, a pathological condition occurring in aging adults, causes among other sequelae, immunosuppression. Thus, decreased immune surveillance increases the probability for neoplastic transformation by increasing the transformed cell's probability for survival. Third, reduced dietary nutritive content and/or absorption in the aging, reduces the amounts of micronutrient antioxidants such as retinoids, proven to be associated with lung cancer prevention.

D. THE MOLECULAR BIOLOGY OF LUNG CANCER

1. ALTERED CELLULAR SIGNALING PATHWAYS

Signal transduction pathways (STPs) comprise a complex assortment of cellular and subcellular interactive molecular messengers. Most signal transduction components are located in specific cellular and subcellular compartments, such as the nucleus, the cytoplasm, or the plasma membrane. Others are secreted. The STPs create a molecular network which interlinks all the stimuli the cell is receiving. This cellular "conscience" allows the cell to respond appropriately to these various signals. It is this molecular signaling which is thought to be the central control of cellular physiology and homeostasis.

Lung cancer cells exhibit four main alterations of cellular signaling pathways:

(1) oncogene activation;
(2) tumor suppressor inactivation;
(3) autocrine and paracrine stimulatory growth loops;
(4) epigenetic changes causing gene expression level changes.

These alterations of signaling components are mainly due to genetic and epigenetic lesions. The ultimate result of the altered signal transduction cascade is the activation of nuclear oncogene products.

1.1 Genetic lesions

Molecular analyses of lung cancers reveal numerous genetic lesions. It is thought that between 10 to 20 genetic events take place before a lung cancer becomes overt. The two most common types of genetic events in lung cancer are, in order: (1) the loss of function of recessive tumor suppressor genes, such as *p53* and *Rb*, and (2) the gain of function of dominant oncogenes, such as *ras*, *myc*, *erbB1* and *erbB2* (Tables V and VI). These genetic events are present in both SCLCs and NSCLCs. Certainly, some genetic

lesions can be used as markers in the diagnosis and prognosis of lung cancer. More importantly, the early detection and molecular and cellular significance of genetic lesions can be used in lung cancer prevention.

1.2 Oncogene activation

The activation or gain of function of dominant oncogenes (Table V) requires that the oncogene be affected by only one activating mutation, because its dominant character will override the other allele. Examples of activating events are point mutations, gene amplifications, and translocations. Consequently, the gene product is constitutively overexpressed or it is abnormally altered, giving rise to a new and deleterious function which constitutes a growth advantage for the cell in question. Thus, activated oncogenes are positive effectors of the transformed phenotype.

DOMINANT ONCOGENE ACTIVATION

GENE	FUNCTION	LOCALE	EVENT	RESULT
ha/ki-*ras*	GDP/GTP binding	Plasma membrane	Point mutation	Decreased activity
N-*myc*	DNA binding	Nucleus	Amplification	Overexpression
erbB1	EGF receptor TK	Transmembrane	Amplification	Overexpression
erbB2	p185 receptor TK	Transmembrane	Amplification	Overexpression

Table V: Common genetic alterations affecting oncogenes found in lung cancer. **TK** = Tyrosine kinase. **GDP** = Guanosine diphosphate. **GTP** = Guanosine triphosphate.

1.3 Tumor suppressor gene inactivation

The inactivation or loss of function of recessive tumor suppressor genes (Table VI) requires two mutational events. This is due to the recessive character of the alleles. The first event is usually a small genetic alteration, such as a somatic point mutation. When only one version of the chromosomal sequence is left, rather than both the maternal and paternal alleles, this phenomenon is known as loss of heterozygosity (LOH). Finally, a second genetic alteration occurs at the other allele. However, this event is usually larger, such as partial deletion or complete allele loss. Since the normal cellular counterpart of a tumor suppressor gene product protects the cell from transformation, tumor suppressor genes are the negative effectors of the transformed phenotype.

RECESSIVE TUMOUR SUPPRESSOR GENE INACTIVATION

GENE	FUNCTION	LOCALE	1st EVENT	RESULT
p53	Cell cycle control	Nuclear	Point mutations Deletions	Sequesters $p53^{wt}$
Rb	Cell cycle control	Nuclear	Point mutations Deletions	Inactivation

Table VI: Common genetic alterations affecting tumor suppressor genes found in lung cancer. **wt** = wild-type.

1.4 Allele loss

Allele loss is a frequent occurrence among lung cancer tumour cells. Cytogenetic analyses have revealed that several types of lesions occur in lung cancer cells, including (1) non-reciprocal translocations; (2) deletions; (3) and numeric abnormalities. Many chromosomal regions appear to be affected (Table VII).

Site of Chromosomal Lesion	Gene
1p	
1q	
3p12-13, 3p14.2	*FHIT*
3p14-23	*RARβ*
3p14-ter	
3p21.3	
3p25	
5q	*APC, MCC*
6p	
6q	
8p	
9p21	*CDKN2/p16*
11p13	
11p15	
13q14	*Rb*
17p13	*p53*
22q	

Table VII: Chromosomal regions frequently involved in cytogenetic abnormalities in lung cancer cells. ***FHIT*** = *Frequently hypomethylated in tumours* gene; ***APC*** = *Adenomatous polyposis coli* gene; ***MCC*** = *Mutated in colon cancer* gene; ***CDKN2/p16*** = *Cyclin-dependent kinase inhibitor 2* gene; ***Rb*** = *Retinoblastoma* gene.

1.5 3p13/14-ter and 3p24-ter deletions

One of the earliest preneoplastic lesions occurring in lung cancer cells is the loss of the short arm of chromosome 3. This region may potentially harbour one or more tumour suppressor genes involved in lung carcinogenesis and is the active interest of many researchers. The putative tumour suppressor gene, RARβ, is among the genes

exhibiting tumour suppressor activity. SCLCs and squamous cell lung carcinomas exhibit many genetic alterations including the loss of function of the putative tumor suppressive gene RARβ. 100% of SCLCs have deletions spanning 3p13/14-ter, and 85-90% of squamous cell lung carcinomas have deletions spanning 3p24-ter (Houle *et al.*, 1991). RARβ is localized in 3p24.

1.6 Other genetic alterations

Other alterations commonly seen in lung tumour cells are: (1) microsatellites; (2) chromosomal deletions; (3) non-reciprocal translocations; (4) double minutes; and (5) hypermethylation.

2. NON-SMALL CELL LUNG CARCINOGENESIS, MODEL

Analogous to Vogelstein's model for colorectal carcinogenesis (Figure 12, p. 20), the preneoplastic changes observed in NSCLC carcinogenesis may also be dissected, and sequential gene alterations used as markers. The precise methodology used in lung cancers is the microdissection of preneoplastic foci, followed by allelotyping. The main difference observed in the earliest lesions of NSCLCs, in comparison with colorectal carcinogenesis, is that rather than large chromosomal deletions, deletions involving chromosome 3p are small (Table VIII).

3p REGION
3p12-13
3p14.2
3p21.3
3p25

Table VIII: List of 3p regions frequently lost in pulmonary preneoplastic foci.

As hyperplasia evolves into dysplasia followed by carcinoma *in situ*, further genetic lesions are incurred and accumulated, including 9p21 LOH (*CDKN2/p16*), aneuploidy, 17p13 LOH (*p53*) and *ras* mutations (Figure 14). It is noteworthy to mention that as the carcinogenesis resumes its invasive stage, more than one region involving 3p is lost (Chung, 1995). Thus genes localized on 3p may act as gate-keepers in lung carcinogenesis.

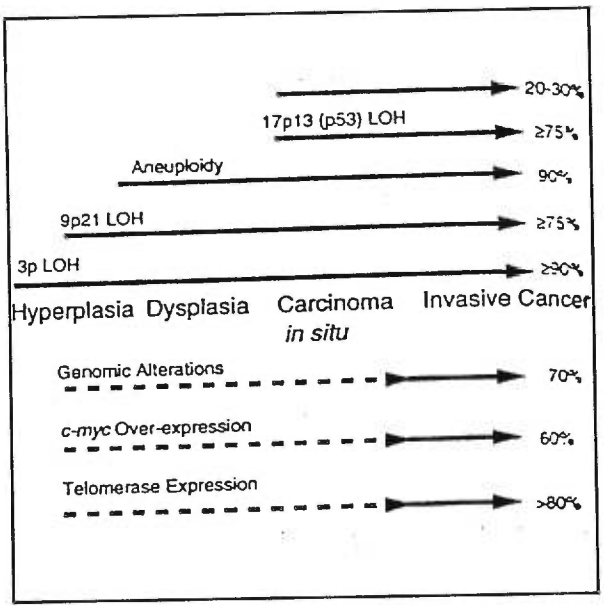

Figure 14: Genetic alterations observed in NSCLC preneoplastic lesions (taken from Minna *et al.*, 1997). **LOH** = Loss of heterozygosity.

3. INACTIVATION OF TUMOUR SUPPRESSOR GENES VIA HYPERMETHYLATION

With respect to loss of function, the silencing of a tumour suppressor gene through epigenetic mechanisms is equivalent to its inactivation through genetic alterations. Methylation is such an epigenetic mechanism. Chemically, it is the covalent modification of the carbon-5 of cytosines within CpG dinucleotides, via addition of a hydrophobic methyl group ($-CH_3$). When clusters of CpG dinucleotides within the 5' untranslated region of genes are hypermethylated, their expression is turned off. It has been shown that the hypermethylation of various housekeeping genes, such as *Rb* (Ohtani-Fujita *et al.*, 1993) and *Von Hippel Lindau* (*VHL*) (Herman *et al.*, 1994) is associated with the silencing of their transcriptional expression. Similarly, hypermethylation of the *CDKN2/p16* gene in NSCLC (Merlo *et al.*, 1995) and hypermethylation of the 3p region in SCLC has been well documented (Suzuki *et al.*, 1994). For more ample detail of tumour suppressor inactivation in the context of DNA methylation, please refer to section G-2.1.2, pp.51-53.

E. RETINOID BIOLOGY

1. RETINOIDS

1.1 Retinoids

Retinoids are vitamin A derivatives and related substances (such as artificially synthesized analogs). Vitamin A, or retinol (ROL) is the metabolic precursor of retinal (RAL), *all-trans* retinoic acid (*at*RA), 3,4,-didehydro retinoic acid and 9-*cis* retinoic acid (9*c*RA) (Figure 15). Retinoids are involved in major cellular processes, including differentiation, proliferation and apoptosis (Roberts *et al.*, 1984). As potent regulators of these processes, they affect development, organogenesis, morphogenesis and cellular homeostasis.

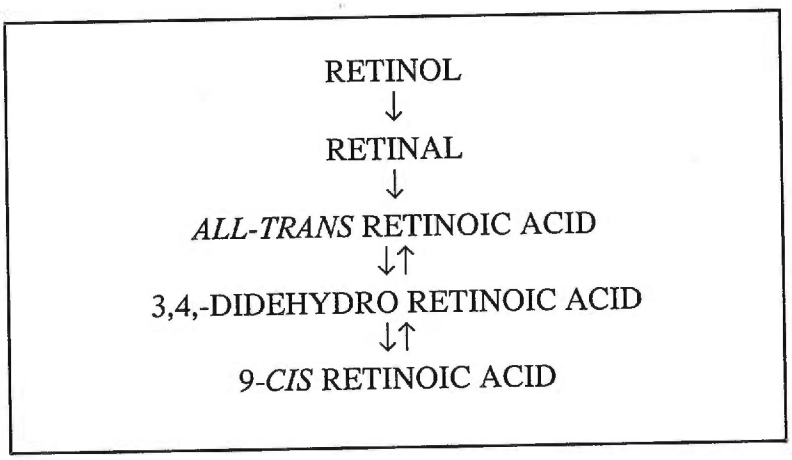

Figure 15: Metabolic pathway of retinol (vitamin A).

Regarding both structure and function, retinoids share many similarities with steroid hormones (such as estrogen and testosterone) and amino acid-derived hormones (such as thyroid hormone and vitamin D_3): (1) they are steroid-like lipophilic molecules (Figure 16, p.35) and thus spontaneously diffuse through the cellular plasma membrane (a lipid bi-layer); (2) upon binding their cognate receptor, they cause it to undergo an allosteric conformational change; (3) the ligand receptor complex translocates into the nucleus; (4) it mediates differential control of gene expression; and (5) its ultimate

function is to regulate development, differentiation and organ physiology in a tempo-spatial-specific way.

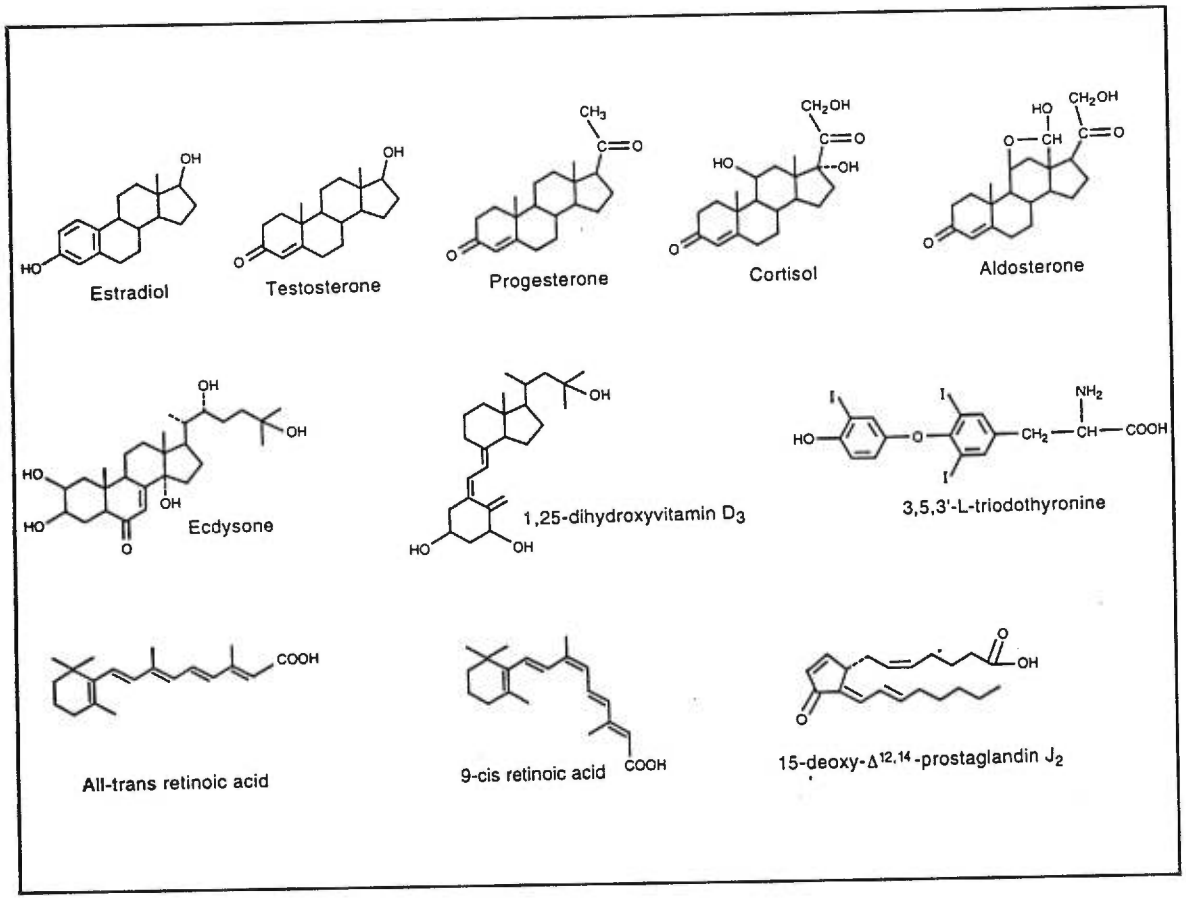

Figure 16: Structures of known steroid-like ligands of nuclear hormone receptors (taken from Mangelsdorf *et al.*, 1995).

1.2 Vitamin A in the diet

Vitamin A is an essential component of the diet: it must be consumed, because neither retinol nor its bio-active metabolic intermediates (*all trans* retinoic acid and 9-*cis* retinoic acid) can be produced from other pre-existing molecular building blocks. Vitamin A requires bile salt production for intestinal absorption and is stocked in the liver (Tortora and Anagnostakos, 1987). It has long been known that vitamin A is essential for vision, growth, tissue maintenance, reproduction and overall survival (Wolbach *et al.*, 1925).

Vitamin A ($C_{20}H_{30}O$), also known as vitamin A_1, is a fat-soluble terpene alcohol (Figure 17). It is mainly obtained from carotenes (provitamins) in green and yellow fruit and vegetable sources; and retinoyl esters in animal fats such as fish oils, milk and butter.

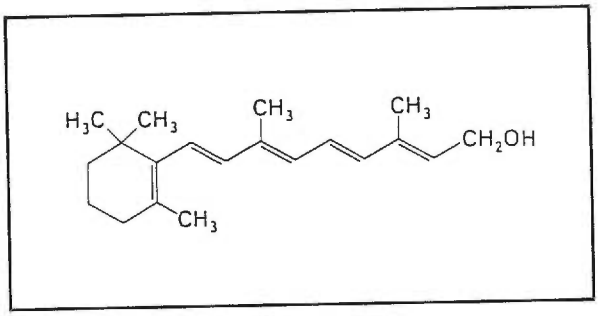

Figure 17: Vitamin A (retinol) (taken from Stryer, 1988).

1.3 Clinical symptoms of vitamin A deficiency

Common signs and symptoms seen in patients suffering from a vitamin A-deficient regime include: (1) dry skin and hair; (2) infections of the ear, sinus, respiratory tract, intestinal tract and genital tract; (3) inability to gain weight; (4) xerophthalmia (drying and ulceration of the cornea); (5) hemeralopia (day blindness); rickets (shunted development of the bones and teeth in the developing years); (6) osteomalacia (demineralization and softening of the bones, accompanied by pain and weakness) (Tortora and Anagnostakos, 1987).

1.4 Experimental vitamin A deprivation

Histologically speaking, the major characteristic of vitamin A deficient (VAD, pronounced "vad") experimental animals is squamous epithelial metaplasia affecting the following systems/organs: (1) respiratory tract; (2) the alimentary tract; (3) the genito-urinary-tract; and (4) the eyes (Wolbach *et al.*, 1925; Sporn *et al.*, 1994; Underwood *et*

al., 1994). Moreover, retinoic acid can reverse all these effects except visual impairment (Wald *et al.*, 1968).

1.5 Retinoids prevent carcinogenesis, experimental evidence

Experimental VAD animals display squamous metaplasia of the bronchial epithelium (Wolbach *et al.*, 1925; Harris *et al.*, 1973). Early experimental evidence that vitamin A specifically plays a role in the prevention of carcinogenesis includes studies that showed that retinoids prevent chemically-induced carcinogenesis (Wolbach *et al.*, 1925; Harris *et al.*, 1973; Chopra and Wilkoff, 1977). More recent studies show that retinoids inhibit cellular proliferation of tumour cell lines (Fazely *et al.*, 1988; Eliason *et al.*, 1993; Maxwell and Mukhopadhyay, 1994). Recent studies confirm that retinoids, such as synthetic retinoid analogs, induce apoptosis of human lung cancer cells *in vitro* and *in vivo* (Lu *et al.*, 1997; Sun *et al.*, 1997, respectively).

Unfortunately, the findings are not consistent. Some earlier studies indicate that retinoids have either no effect on carcinogenesis, or worse, that they may even induce carcinogenesis (Schroder and Black, 1980; National Research Council, 1989).

1.6 Retinoids prevent carcinogenesis, epidemiological evidence

Epidemiological studies have shown that in humans, dietary intake of vitamin A is inversely proportional to risk for lung cancer (Graham, 1984). A vast literature concerning retinoid dietary intake and prevention of cancers (lung, oral cavity, larynx, esophagus and stomach) exists. Though most studies indicate that retinoids may be useful in chemoprevention, results vary between studies. This may be due to several factors: (1) different compilation strategies for dietary data; (2) life-style factors (eg. smoking appears to be associated with destruction of carotenoids and retinol (Handelman, 1996); (3) environmental factors; (4) alcohol consumption (interaction of ethanol and retinol) (Mayne *et al.*, 1991); (5) zinc status (Mayne *et al.*, 1991); (6) other artifactual factors. Two other key factors may be: (1) the bioavailability of retinoic acid to different tissues under different circumstances, such as age and other factors; (2) the levels of cellular retinoic acid binding proteins I & II (CRABP I & II) and retinol binding proteins I & II (CRBP I & II).

Recent epidemiological evidence shows that β-carotene is a marker for some protective factor(s) against lung cancer (Comstock *et al.*, 1997). Moreover, other epidemiological studies indicate that the protection is due to the dietary intake of fruits and vegetables, not the pre-formed and purified oral supplements of vitamin A (Yong, 1997). Unfortunately, once again, finding are inconsistent. A study in 1996 revealed that β-carotene and vitamin A have no benefit or possible harm on lung cancer (Marwick, 1996).

1.7 Retinoids as anticancer agents

Retinoids are currently being tested in a variety of experimental and clinical settings. Experimental studies investigate their effects on the proliferation, differentiation and apoptosis of normal, premalignant and malignant epithelial cells. Preclinical and clinical studies include patients at high risk for lung cancer, such as smokers and asbestos workers, patients with lung cancer, and the general population.

Leading clinical research involving retinoic acid is most interested in its chemopreventive and anticancer potentials in respiratory cancers.

2. RETINOID RECEPTORS

2.1 The nuclear receptor superfamily

The immediate cellular effectors of retinoids comprise the retinoic acid receptors (RARs) and the retinoid X receptors (RXRs). RARs and RXRs belong to the nuclear receptor superfamily, which consists of steroid (cholesterol derivatives), non-steroid, and orphan (unknown ligand) nuclear receptors (Table IX, p.39). Please refer to Figure 16 of page 35 for the chemical structure of cognate ligands. All nuclear receptors bind lipophilic hormones and translocate into the nucleus (as ligand-receptor complexes) where they regulate differential gene expression via activation or repression of transcription.

Receptor superfamily	Receptor family	Major Ligand(s)
Steroid receptors	GR	Glucocorticoids (G)
	MR	Mineralocorticoids (M)
	ER	Estrogens (E)
	PR	Progestins (P)
	AR	Androgens (A)
Non-steroid receptors	RAR	*at*RA* & 9*c*RA
	RXR	9*c*RA
	TR	Thyroid hormone
	VDR	Vitamin D_3
Orphan receptors	SF-1	?
	NGFI-B	?
	HNF4	?
	COUP-TF I & II	?
	PPARα	?

Table IX: Human nuclear receptor superfamily, receptor families and major cognate ligands. **R** = receptor; ***at*RA** = *all trans* retinoic acid; **9*c*RA** = 9-*cis* retinoic acid; **SF-1** = steroidogenic factor-1; **NGFI-B** = nerve growth factor induced receptor-B; **HNF4** = hepatocyte nuclear factor 4; **COUP-TF** = chicken ovalbumin upstream promoter transcription factor; **PPARα** = peroxisome proliferator activator receptor α; **?** = cognate ligand is as of yet unknown.

* *at*RA has been shown to be more active than 9*c*RA (Sporn and Roberts, 1984).

The nuclear receptor superfamily is characterized by its unique structural and functional modularity. It comprises five main regions (Figure 18): (A/B) a variable amino (N)-terminal region; (C) a central conserved DNA binding domain (DBD); (D) a variable hinge region; (E) a conserved ligand binding domain (LBD); and (F) a variable carboxy (C)-terminal region. The A/B region contains an activation factor-1 (AF-1) site. The E region contains an AF-2 site. AF sites differentially activate RA-responsive gene promoters.

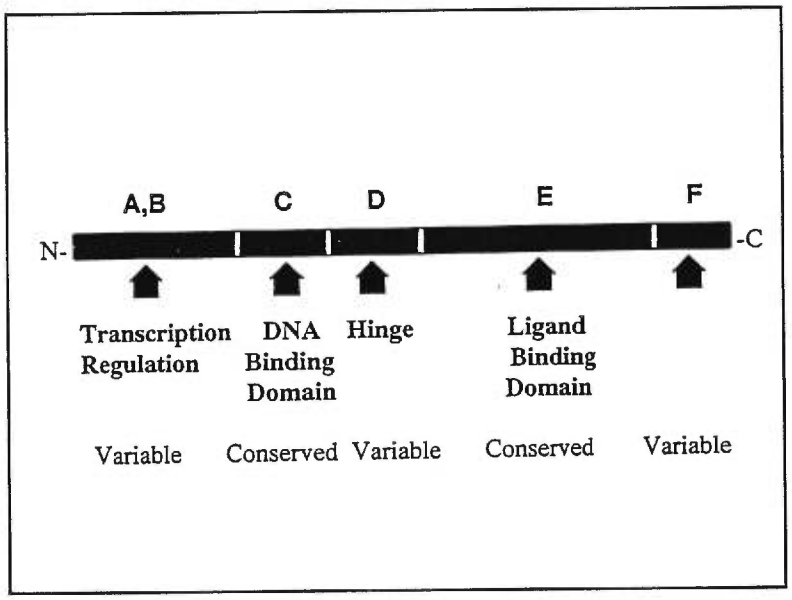

Figure 18: Schematic diagram of a nuclear receptor.
N- = amino terminus; **-C** = carboxy terminus.

The LBD recognizes its cognate hormone and thus ensures the specificity and the appropriateness of the ultimate physiological response. Moreover, the LBD contains a heterodimerization motif. Another heterodimerization motif is contained within the DBD core.

The hinge region is variable and allows up to 180 degrees swivel.

The DBD is composed of 66 amino acids forming two zinc fingers. It recognizes DNA consensus sequences contained within genes that have the appropriate hormone response elements (HRE). The hormone response elements (HREs) specific for RARs are

known as RAREs, or retinoic acid response elements. They consist of two direct repeats separated by one-to-five nucleotides. The direct repeat consists of a minimal core hexad with the consensus sequence PuG(G/T)TCA (non-palindromic)(Pierre Chambon, 1994). RAREs are numerous and polymorphic. They ensure (1) the specificity of the hormone response and (2) the differential control of expression of the hormone receptor's target genes.

The N-terminal region varies considerably both in amino acid sequence and in length. It has transactivating properties on the transcription of the genes containing the DBD consensus sequence.

Thus, the two major characteristics which distinguish nuclear hormone receptors are: (1) their direct permeability to the nucleus (which abrogates the need for membrane receptors and multi-messenger signal transduction cascades common to water-soluble peptide hormones and growth factors); (2) their unique modular structure which permits hormone-specific/multi-target gene-specific transcriptionally-dependent responses. It is essential to remark that hormone nuclear receptors are conserved through species ranging from the invertebrates to the vertebrates. This evolutionary conservation underlies their efficient structure-function design and critical importance to life processes.

2.2 Retinoic acid receptors and retinoid X receptors

RARs were first discovered in 1987 (Petkovitch *et al.*, 1987; Giguère *et al.*, 1987); RXRs were first discovered in 1990 (Mangelsdorf *et al.*, 1990). Each receptor family consists of three different receptor types: α, β, and γ. These receptor types are coded by three different genes (Table X, p.42). Through the use of different promoters and/or alternative splicing (a post-transcriptional modification) at the 5' end, each receptor type exists as several different isoforms. RXRs form homodimers and heterodimerize with various other receptors, including RAR, TR, VDR, and peroxisome proliferator-activator receptor (PPAR). RARs do not appear to form homodimers, but heterodimerize with RXRs.

FAMILY	RECEPTOR TYPES	MAJOR ISOFORMS	CHROMOSOMAL LOCALIZATION
RAR	α	α1, α2	17p21.1
	β	β1, β2,	3p24
	γ	γ1, γ2	12q13
RXR	α	α1, α2	9q34.3
	β	β1, β2, β3, β4	6p21.3
	γ	γ1, γ2	1q22-23

Table X: Human retinoid receptor family, type, isoform and chromosomal localization of corresponding coding gene (based on Chambon, 1996).

2.3 RAR single knockouts

Single null mutant mice RARα1$^{-/-}$, RARγ2$^{-/-}$, RARβ$^{-/-}$ and RARβ2$^{-/-}$ appear normal (Li *et al.*, 1993; Lohnes *et al.*, 1993; Lohnes *et al.*, 1994; Luo *et al.*, 1995, respectively). Single null mutant mice RARα$^{-/-}$, RARγ$^{-/-}$, RXR$^{+/-}$ and RXRβ$^{-/-}$ exhibit poor viability, shunted growth and male sterility, manifestations reminiscent of postnatal VAD syndrome. RA administration can cure or prevent these manifestations, likely because RA induces the expression of certain isoforms: α2, β2, γ (Leroy *et al.*, 1991; Zelent *et al.*, 1991; Lehman *et al.*, 1992). Overall, these limited pathological phenotypes suggest the possibility for functional redundancy among retinoid receptor types and isoforms.

2.4 RAR double knockouts

The majority of double mutants (especially those involving RARβ and α1 or γ) exhibit extremely poor viability (for a review, see Kastner *et al.*, 1995). These findings demonstrate that (1) RAR/RXR heterodimers mediate the physiological effects of vitamin A derivatives and (2) there is some functional redundancy among RAR receptors.

2.5 Combinatorial complexity among retinoid receptors

EXAMPLES OF KNOWN FACTORS CONTRIBUTING TO RETINOIC ACID'S COMBINATORIAL COMPLEXITY AND DIVERSITY OF PHYSIOLOGICAL RESPONSE

BIOLOGICAL LEVEL	EXAMPLES OF DETERMINING FACTORS
Ligands	RAR: *at*RA and 9*c*RA RXR: 9*c*RA
Binding proteins	CRBP I, CRBP II, CRABP I, CRABP II
Distribution and Tissue bioavailability	Systemic circulation and liver storage Bile production/intestinal absorption Ethanol interference Zinc interference Tobacco-smoke-destruction of retinoids
Spatial specificity	Organ-specificity Tissue-specificity Cellular-specificity
Cellular entry	Possible causes of steric hindrance?
Receptor heterodimerizations	36 possibilities: 6 RAR isoforms x 6 RXR isoforms
Receptor expression	Tissue-specificity Temporal-specificity Promoter methylation?
Signaling modifications	A/B region phosphorylation
Response elements	Numerous and polymorphic
Target genes	Numerous
Coactivators Corepressors Cross-talk	AF-1 and AF-2 agonists N-CoR, SMRT, NFκB, AF-1 and AF-2 antagonists Other cellular factors (TFs, TIFs, etc)
Chromatin modifications	Methylation status Acetylation status

Table XI: Non-exhaustive list of factors contributing to the combinatorial complexity of the retinoic acid response in humans. **N-CoR** = Nuclear receptor corepressor; **SMRT** = Silencing mediator for RAR and TR; **NFκB**=nuclear factor kappa B; **AF-1** = Activator function-1; **TF** = transcription factor; **TIF** = transcription intermediary factor; **TR** = thyroid hormone receptor.

F. RARβ EXPRESSION

1. RARβ EXPRESSION IN 40 DIFFERENT CELL LINES

40 different cell lines have been previously assayed for RARβ expression in our laboratory (Houle *Ph.D Thesis*, 1991). RNAase protection analyses were performed on RNA extracted from various cell-lines with or without retinoic acid (3.3×10^{-7} M) stimulation. The cell-lines consisted of those listed in Table XII.

ORGAN TYPE	CLASSIFICATION OF CELL-LINE STUDIED	NUMBER OF LINES STUDIED	RARβ POSITIVE
LUNG	Small cell lung carcinoma (SCLC)	7	4(57%)
	Squamous cell lung carcinoma	5	2(40%)
	Lung adenocarcinoma	5	4(80%)
	Adenosquamous lung carcinoma	2	1(50%)
	Large cell lung carcinoma	2	1(50%)
	Normal tracheo-bronchial epithelium (immortalized)	1	0
COLON	Adenocarcinoma	6	1(17%)
	Intestinal polyps	4	0
OTHER	Cervical squamous cell carcinoma	3	0
	Breast adenocarcinoma	2	0
	Maxillary gland squamous cell carcinoma	1	0
	Pharyngeal squamous cell carcinoma	1	0
	Melanoma	1	0
TOTAL		40	13

Table XII: RARβ expression in various cell-lines (based on Houle *Ph.D Thesis*, 1991).

1.1 Actual data from RARβ expression analyses

Please refer to Table XIII for specific information regarding RARβ expression analyses.

Lignées cellulaires étudiées

Lignées	Origine	Expression au locus RARβ
BEAS-2B	Cellules immortalisées provenant de l'épithélium trachéo-bronchique normal	-
BK-T	Carcinome pulmonaire à petites cellules	+
NCI-H69	Carcinome pulmonaire à petites cellules	+
NCI-H82	Carcinome pulmonaire à petites cellules	- (1)
NCI-H128	Carcinome pulmonaire à petites cellules	n.d.
NCI-H209	Carcinome pulmonaire à petites cellules	-
MM-1	Carcinome pulmonaire à petites cellules	+ (2)
RG-1	Carcinome pulmonaire à petites cellules	+ (2)
A549	Adénocarcinome pulmonaire	+
CALU-3	Adénocarcinome pulmonaire	+
LC-T	Adénocarcinome pulmonaire	+
NCI-H23	Adénocarcinome pulmonaire	+
SKLU-1	Adénocarcinome pulmonaire	n.d.
NCI-H125	Carcinome pulmonaire adénosquameux	+
NCI-H596	Carcinome pulmonaire adénosquameux	-
CALU-1	Carcinome pulmonaire épidermoïde	-
NCI-H157	Carcinome pulmonaire épidermoïde	-/+ (3)
NCI-H520	Carcinome pulmonaire épidermoïde	+
P2	Carcinome pulmonaire épidermoïde	-
SKMES	Carcinome pulmonaire épidermoïde	-
QU-DB	Carcinome pulmonaire à grandes cellules	+
RVH-6849	Carcinome pulmonaire à grandes cellules	-
CACO-2	Adénocarcinome du côlon	+
COLO-201	Adénocarcinome du côlon	-
COLO-205	Adénocarcinome du côlon	-
HCT-15 VII	Adénocarcinome du côlon	-
LS-180	Adénocarcinome du côlon	-
SW-1222	Adénocarcinome du côlon	-
HTB-126	Adénocarcinome du sein	-
MCF-7	Adénocarcinome du sein	-
ME-180	Carcinome épidermoïde du col utérin	-
MS-751	Carcinome épidermoïde du col utérin	-
SiHa	Carcinome épidermoïde du col utérin	-
A-253	Carcinome épidermoïde des glandes maxillaires	-
FaDu	Carcinome épidermoïde du pharynx	-
MeWo	Mélanome	-
235EP	Polype intestinal, passage précoce, non-tumorigènes	-
235LP	Polype intestinal, passage tardif, faiblement tumorigènes	-
411	Xénogreffe d'une tumeur induite par les cellules 235LP, fortement tumorigènes	-
330	Polype intestinal, non-tumorigènes	-

(1) L'expression de RARβ peut être induite par un traitement à l'acide rétinoïque (figure 3A)
(2) Seule la forme β1 est exprimée
(3) L'expression est devenue détectable après maintien prolongé en culture
n.d. Non déterminé

Table XIII: RARβ expression in 40 different cell-lines, determined by RNAase Protection Assays (taken from Houle *Ph.D Thesis*, 1991).

G. DNA METHYLATION

1. BRIEF OVERVIEW

1.1 Biochemistry of DNA methylation

DNA methylation stands for the post-replicational biochemical process by which cytosines (C) or adenines (A) within specific DNA sequences (such as CpG dinucleotides) are modified by the covalent addition of a methyl ($-CH_3$) group (Alberts *et al.,* 1994). This reaction (Figure 19) is normally catalyzed by an enzyme, such as methyl transferase (MeTase) and requires the $-CH_3$ group-donor S-adenosylmethionine (S.A.M.) (Szyf, 1996).

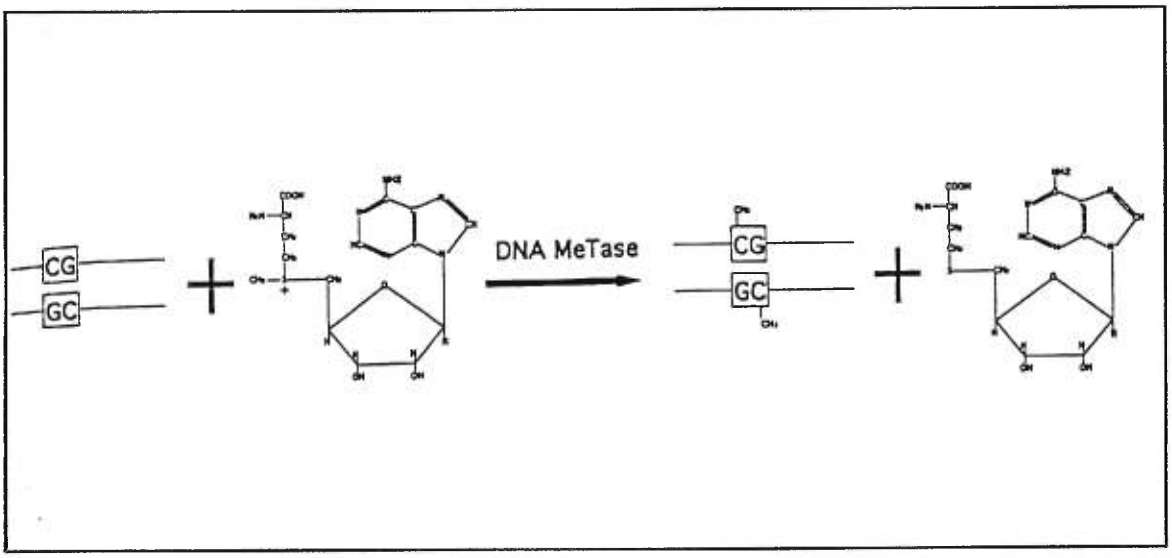

Figure 19: The DNA methylation reaction (taken from Szyf, 1996).

1.2 DNA methylation is not conserved among species

DNA methylation exists in prokaryotes, such as bacteria (eg. *Escherischia coli*), and vertebrate eukaryotes, such as man. Homozygous MeTase mutant mice die at

midstage embryogenesis (Li *et al.*, 1992), suggesting that DNA methylation is essential in mouse development. However, DNA methylation is inexistant in some lower eukaryotes, such as yeast (eg. *Saccharomyces cerevisiae*), the worm (eg. *Caenorhabditis elegans*), and the fruit fly (eg. *Drosophila melanogaster*). This lack of evolutionary conservation connotes the secondary role methylation may play in fundamental life processes, and is one of the many facts which sheds confusion onto the study of this process.

1.3 DNA methylation protects the bacterial genome proper

It is noteworthy to mention that the ancestral function of DNA methylation is thought to have been the protection of the genome against the organism's own mechanisms that neutralize invading DNA (Bestor, 1990), such as that from retro-transposons and proviruses. For example, *Escherischia coli* methylates its As and Cs post-replicationally, within specific sequences (eg. GATC is methylated at its A residue by adenine methylase). The bacterium has endogenous restriction enzymes which hydrolyze specific sequences of 4 to 8 nucleotides. Bacteriophage proviral integrated sequences (which are generally not methylated) are sensitive to these restriction enzymes. The corresponding methylated sequences (belonging to the bacterium's genome) are protected because these enzymes cannot hydrolyze methylated sequences.

1.4 The methyl-directed pathway of mismatch repair

Another role methylation plays in prokaryotes and eukaryotes is parental DNA strand identification to allow daughter strand-specific mismatch repair. More precisely, the *de novo* synthesized strand during DNA replication is not immediately methylated upon polymerization (Kornberg and Baker, 1992). Instead, its GATC sequences remain in the unmethylated state, in order to allow mismatch repair enzymes in the "methyl-directed pathway" (Modrich, P., 1987) to distinguish which strand's bases should be removed upon mismatch recognition. Thus, the unmethylated daughter strand is targeted for repair, whereas the methylated parental strand is not. This conserves the parental strand as the original template throughout the replicative cycles.

1.5 Functions of DNA methylation in eukaryotes

In vertebrates, methylation most frequently involves the CpG dinucleotide (Gruenbaum *et al.*, 1981). DNA methylation in higher eukaryotes is known to play an important role in the normal and pathological differential control of gene expression.

More precisely, it has been shown to regulate the expression of:

(1) tissue-specific genes;
(2) parentally-imprinted allele-specific genes;
(3) positive and negative effectors of carcinogenesis.

Thus, the main role of DNA methylation in higher eukaryotes is the differential control of gene expression. Briefly, every cell contains the same genetic information (the genome) but every cell has different epigenetic information regulating the ultimate expression of that genetic information. DNA methylation may be a form of epigenetic information, in that it is an additional level of information not encoded in the genetic information, but added through covalent modifications by non-replication machinery.

1.6 Precise molecular biological effects of DNA methylation

The exact way(s) by which DNA methylation is capable of regulating gene expression is(are) as of yet not precisely known. However, three specific processes/states are known to be affected by the status of the corresponding sequence's DNA methylation. They are: (1) mutability; (2) alteration of chromatin structure; and (3) level of transcription.

1.6.1 DNA methylation and mutability

DNA methylation facilitates point mutations because cytosines methylated at their 5-carbon position (5m-C) may spontaneously deaminate to thymines (T). More than one third of all DNA mutations implicated in human diseases are C→T transition substitutions within CpG dinucleotides (i.e. CpG to TpG) (Cooper and Krawczak, 1989).

This is a very frequent event when considering the fact that CpG dinucleotides are rare and the fact that there is a DNA repair mechanism specific for this type of problem. Thus, methylation increases the risk of mutation considerably.

1.6.2 DNA methylation and chromatin structure

Chromatin is the eukaryotic structural and functional organization of chromosomal material and small basic nuclear proteins, such as histones. Chromatin can be organized into an open accessible structure (euchromatin) or it can be compactly organized (heterochromatin). The latter is generally associated with transcriptional repression. Some studies have found an association between DNA methylation and structural changes of chromatin. For example, high levels of *de novo* methylation were shown to be associated with altered chromatin structure (Antequera *et al.*, 1990), and nickel, a carcinogen found in the residue phase of tobacco smoke, has been shown to induce methylation and silence genes via specific chromatin condensation (Lee *et al.*, 1995).

1.6.3 Transcription repression is correlated with DNA methylation

The main lines of evidence suggesting that transcription repression is correlated with DNA methylation are:

(1) endogenous gene expression correlated with hypermethylation (please see section 1.6.4, p.50, for more detail);
(2) *in vitro* promoter methylation of exogenous gene constructs correlated with decreased expression or lack of expression (eg. Rhodes *et al.*, 1994);
(3) *in vitro* use of cytidine analogs inhibiting methylation (5-azacytidine and 5-aza-2'deoxycytidine) correlated with activation of gene expression (reviewed by Jones, 1985);
(4) *in vitro* use of antisense *MeTase* cDNA transfection to inhibit methylation thereby abrogating the transcription block (Ramchandani *et al.*, 1992).

Also, inactivating mutations of MeTase in the mouse cause embryonic lethality (Li *et al.*, 1992), suggesting that methylation-regulated gene expression per se may be essential for developmental and other processes.

1.6.4 Endogenous gene repression is correlated with hypermethylation

Numerous studies since the late 1970's have found an association between DNA hypermethylation and repression of transcription. The first suggestion that transcriptional repression was associated with hypermethylation came from Razin and Riggs, in 1980. Some examples of studies that came to follow concern: parentally-imprinted genes (Sapienza *et al.*, 1987; Swain *et al.*, 1987), X-inactivated genes of mammalian females (for a review, Grant and Chapman, 1988; Pfeifer *et al.*, 1990a; Migeon, 1994), such as X-linked phosphoglycerate kinase (Pfeifer, 1990b; Park and Chapman, 1994), and inherited Fragile-X disease (Sutcliffe *et al.*, 1992; Knight *et al.*, 1993), to name a but a few.

1.6.4.1 CpG islands

The vast majority of the mammalian genome is CpG-poor. For example, whereas the frequency of GpC dinuclotides is approximately 0.043 in the human spleen, the frequency of CpG dinucleotides is only 0.010 (Subak-Sharpe *et al.*, 1966). The regions containing the least CpG sites are transcriptionally-inactive, such as introns (Kornberg and Baker, 1992). Most of the genes that have been shown to be repressed by methylation have promoters with a high frequency of CpGs. These promoters are said to be on "CpG islands", regions which are CpG-rich. Between 70 and 90% of genomic CpGs are methylated. However, CpG islands are not normally methylated (Bird *et al.*, 1985; Bird, 1986). The extent of CpG-methylation-associated repression has been shown to be dependent upon CpG density and promoter strength (Boyes and Bird, 1992). CpG-methylation may hinder transcription directly, by altering chromatin structure, or it may hinder transcription indirectly, via Methyl CpG-Binding Proteins. On the one hand, CpG methylation may hinder transcription *directly*, by altering DNA or chromatin structure (Hsieh, 1994; Antequera, *et al.* 1990; Tazi and Bird, 1990). On the other hand, CpG methylation may hinder transcription *indirectly*, via methyl CpG-binding proteins (Nan *et al.*, 1997; Tate *et al.*, 1996; Lewis *et al.*, 1992; Boyes and Bird, 1991).

2. ABERRANT METHYLATION PATTERNS AND CANCER

2.1 Aberrant methylation is associated with cancer

Studies concerning both specific and general methylation patterns in cancer cells are contradictory. Some studies show widespread hypomethylation (Gama-Sosa *et al.*, 1983; Feinberg *et al.*, 1988); other studies reveal hypermethylation of specific chromosomal regions (de Bustros *et al.*, 1988; Nelkin *et al.*, 1991; Makos *et al.*, 1992). Nonetheless, hypomethylation generally affects bulk (or non-CpG island) DNA; hypermethylation generally affects CpG-rich DNA.

2.1.1 Oncogene activation and tumour suppressor inactivation

It is plausible that these two epigenetic statuses, hypo- and hyper-methylation, may involve different types of genes. Since the state of methylation is correlated with gene expression, *positive effectors* of carcinogenesis may be *hypo*methylated, and *negative effectors* of carcinogenesis may be *hyper*methylated. More precisely, oncogene activation may involve aberrant hypomethylation of its promoter, whereas tumour suppressor gene inactivation may involve aberrant hypermethylation of its promoter.

It is not known whether aberrant methylation is causally related to either carcinogenesis nor transcriptional regulation. Methylation may be causal or consequential in both cases.

2.1.2 Tumour suppressor hypermethylation and cancer

Relatively recent studies have shown that the inactivation of several tumour suppressor genes is associated with hypermethylation in tumour cells (Table XIV). As of yet, not many tumour suppressor genes have been studied for promoter regional methylation.

Back in 1992, partial methylation of the CpG island 5' of the *WT*-1 gene was demonstrated in 7% of 29 Wilm's tumour samples tested (Royer-Pokora and Schneider, 1992). In 1993, CpG hypermethylation was shown to affect 16% of 31 unilateral retinoblastoma tumours tested (Ohtani-Fujita *et al.*, 1993). A study in 1994 found that 19% of 26 clear cell renal carcinomas were hypermethylated on the CpG island 5' of the *VHL* gene (Herman *et al.*, 1994). Then in 1995, approximately 26% of 27 NSCLC primary tumours, 25% of 4 head-and-neck squamous cell carcinoma primary tumours and 23% of 13 glioma primary tumours, were shown to be *de novo* methylated on the 5' CpG island of *CDKN2/p16*, whereas normal tissue was not (Merlo *et al.*, 1995). In all cases, hypermethylation correlated with lack of gene expression. Moreover, in all cases, a significant proportion of these tumours were hypermethylated.

GENE	%TUMOURS METHYLATED	CANCER CELLS	REFERENCE
WT-1	7%	Wilm's tumour embryonal kidney	Royer-Pokora and Schneider, 1992
Rb	16%	Retinoblastoma	Ohtani-Fujita *et al.*, 1993
VHL	19%	Clear cell renal carcinoma	Herman *et al.*, 1994
p16/CDKN2	20%	Various primary tumours	Merlo *et al.*, 1995

Table XIV: Summary of known hypermethylation-associated tumour suppressor gene inactivations in cancer cells.

These findings suggest that 5' hypermethylation of tumour suppressor genes may be an important and common promoting event in carcinogenesis. This is because tumour suppressor inactivation is one of the most frequent genetic alterations observed in cancer. Thus, hypermethylation-associated repression may be an epigenetic alteration which affects one or both alleles of tumour suppressor genes.

In light of hypermethylation's *increased potential for mutations* (section G-1.6.1, p.48) and Vogelstein's *multistage* model for colorectal carcinogenesis (section C-1.3, p. 20), it is conceivable that hypermethylation is a preneoplastic event which is causally associated with subsequent neoplastic genetic alterations such as point mutations. In other words, DNA hypermethylation may be a reversible tumour promoting event which *antedates* irreversible carcinogenic genetic alterations.

Evidence supporting this hypothesis might consist of data showing that one or both alleles of a tumour suppressor gene is/are altered via hypermethylation in early stage carcinogenesis, but later stages of carcinogenesis are characterized by hypermethylation affecting one allele at most, and point mutations affecting the other allele(s). Such evidence has yet to be documented, however similar findings have. The study on *VHL* hypermethylation mentioned above (Herman *et al.*, 1994), showed that of the 5 clear cell renal carcinomas which were hypermethylated, 1 had two heavily methylated copies and 4 had LOH. Thus, methylation may inactivate both alleles of a tumour suppressor gene, but it may also lead to a genetic alteration causing permanent inactivation of *one* allele (eg. LOH), or even *both* alleles. Further supporting evidence consists of the fact that over 33% of all mutations associated with human pathology comprise CT transitions within CpG dinucleotides (Cooper and Krawczak, 1989).

Recently, hypermethylation of the *HIC*-1 locus (*hypermethylated in cancer*) was shown in 33% of 51 lung tumours analyzed (Eguchi *et al.*, 1997). However, 31% of corresponding non-tumour tissue samples (at least 4cm distal to the tumour) were also hypermethylated at this locus. Interestingly, hypermethylation was shown to be correlated with poor differentiation grade and smoking history. Since smoking is thought to be the number one risk factor for lung cancer, it is plausible that hypermethylation may be an intermediate event between smoking behaviour and the ultimate development of lung cancer.

2.1.3 Pharmacologically-induced hypomethylation can reverse the cancer process

Numerous studies show that pharmacologically-induced hypomethylation inhibits carcinogenesis. Several *in vitro* studies reveal that the cytidine analog 5-azaC has anti-leukemic activity (Christman *et al.,* 1983; Gambari *et al.*, 1984; Momparler *et al.*, 1984; Motoji *et al.*, 1985; Attadia, 1993). Clinical studies have also shown that 5-azaC can

inhibit leukemia (Jehn, 1989). Other studies have shown that 5-azaC is cytotoxic to human tumour cells *in vitro*, alone or in conjunction with cisplatin (Frost *et al.*, 1990). Intestinal neoplasia is suppressed in APC mutant mice treated with 5-azaC (Laird *et al.*, 1995).

It is not known whether the induced hypomethylation affects tumour suppressor genes *per se* and/or another facet of the carcinogenic process. Moreover, some studies show that 5-azaC can induce transformation *in vitro* (Carr *et al.*, 1984; Walker and Nettesheim, 1989).

2.2 Is *RAR*β methylated in epithelial cancers such as lung and colon cancers?

*RAR*β is a candidate tumour suppressor gene. It is located on 3p24, a chromosomal locus involved in one of the earliest preneoplastic gene alterations to occur in lung cancer. The second alteration leading to its loss of function may be genetic, or, as we've seen, it may be epigenetic. In other words, hypermethylation of the *RAR*β promoter is a possible carcinogenesis-associated alteration, associated with or resulting in its transcriptional repression. Expression of RARβ has been shown to be frequently repressed in lung cancer cell-lines and colon cancer cell-lines (Houle *et al.*, 1991).

The *RAR*β promoter 2, (P2), is located on a CpG island. Thus, it is possibly sensitive to CpG hypermethylation. Moreover, 5-azadeoxyC-treated DLD-1 colon carcinoma cells have recently been shown to regain RARβ expression (Côté and Momparler, 1997). Not only is hypermethylation frequently associated with epidermoid carcinogenesis, but a new study reveals that hypermethylation is also associated with patient smoking history in non-small cell lung carcinoma (Eguchi *et al.*, 1997). The reader will easily recall that smoking is the main risk factor for the development of a lung cancer, especially of the epidermoid type (eg. squamous cell lung carcinoma, an NSCLC).

H. PROJECT OBJECTIVES

1. OBJECTIVE 1

1.1 Specific description of objective 1

The first aim of this project was to determine the CpG methylation status of the *RAR*β2 P2 promoter in lung cancer cell lines, colon adenocarcinoma cell-lines, and lung tumour cancer cells.

1.2 Rationalization

The *RAR*β gene codes for the putative tumour suppressor gene product, RARβ2, under the control of the P2 promoter. *RAR*βis located at the chromosomal locus 3p24. One of the earliest preneoplastic lesions occurring in lung cancer cells involves this locus. 100% of SCLCs have deletions spanning 3p13/14-ter, and 85-90% of squamous cell lung carcinomas have deletions spanning 3p24-ter (Houle *et al.*, 1991). Moreover, lack of RARβ expression has been shown in numerous lung and colon cancer cell-lines (Houle *Ph.D Thesis*, 1991).

Hypermethylation of CpG islands within gene promoters is correlated with a decrease or lack of gene expression. Moreover, hypermethylation is frequently associated with epidermoid carcinogenesis. Recently, hypermethylation has also been shown to be associated with cigarette smoking, the main risk factor for lung cancer, in NSCLCs (Eguchi *et al.*, 1997). Also, RARβ expression has been shown to be re-activated in DLD-1 colon carcinoma cells by inhibiting methylation with 5-aza-2'-deoxycytidine, a cytidine analogue (Côté and Momparler, 1997).

1.3 Hypothesis

Lack of RARβ2 expression correlates with hypermethylation of the CpG island within the RARβ P2 promoter in lung and colon cancer cells. Thus, hypermethylation may constitute an epigenetic event involved in RARβ2 loss of function in RARβ2-*non*-expressing cancer cells.

2. OBJECTIVE 2

2.1 Specific description of objective 2

The second aim of this project was to improve the *Bisulphite Genomic Sequencing Protocol* (Clark *et al.*, 1994). The specific subordinate objectives were: **(1)** to increase the amount of recoverable bisulphite-treated DNA; and **(2)** to increase the conversion efficiency of non-CpG-cytosines (to their bisulphite-reacted products, uracils) to the established threshold of 98.6% or greater.

2.2 Rationalization

The *Bisulphite Genomic Sequencing Protocol* is a highly sensitive qualitative analysis of the methylation status of all individual CpG-cytosines within a target sequence. Bisulphite converts unmethylated cytosines to uracils, whereas methylated cytosines are unreactive and remain unconverted. Subsequent sequencing of subcloned PCR-amplified bisulphite-treated DNA positively displays *5-methylcytosines* as *cytosines* (i.e. **5m-C = C**), and negatively displays unmethylated cytosines as thymines (i.e. **C = T**).

First, for obvious reasons bisulphite-treated DNA must not be completely degraded by the harsh chemical conditions of the bisulphite conversion reaction (see *Materials and Methods*, pp.62-66). Enough DNA must be recoverable for further PCR amplification. However, DNA is hypersensitive to the reaction conditions, causing a significant amount of DNA to be degraded. Material loss during preliminary experimentation was frequently so severe, that bisulphite-treated DNA was not PCR-amplifiable. Second, cytosines not within CpG dinucleotides should consistently be converted to uracils, since they are not normally methylated. However, large portions of the treated molecules remained unconverted or partially converted, primarily because the reaction is highly single-strand specific.

2.3 Criteria

The criteria were to increase DNA recovery and circumvent inefficient conversion. If attained, this objective of improving an already highly sensitive protocol could significantly increase its potential use by decreasing the concomitant workload and cost.

II. MATERIALS AND METHODS

A. CELL CULTURE

1. DESCRIPTION OF CELL LINES

Fifteen different cell lines, belonging to four main histological classifications, were successfully examined in this project: (1) NSCLCs; (2) SCLCs; (3) normal bronchial epithelium; and (4) colon adenocarcinomas. Morphologically, all of the cell lines used are epithelial-like, except COLO-201 which is fibroblast-like.

1.1 Non small cell lung carcinoma cell lines

CALU-1, an epidermoid cell line, was obtained from the American Type Culture Collection (ATCC; Rockville, MD). CALU-1 clones, C-19 and C-59, are RARβ transfectants (Houle *et al.*, 1993). The latter expresses functional RARβ. NCI-H157, NCI-H520 and NCI-H596 were kindly supplied by Dr. Adi Gazdar, (NCI, NIH, Bethesda, MD). NCI-H157 and NCI-H520 are epidermoid; NCI-H596 is adenosquamous.

1.2 Small cell lung carcinoma cell lines

NCI-H82 was kindly supplied by Dr. Adi Gazdar (NCI, NIH, Bethesda, MD). MM-1 and RG-1 were established in our laboratory, from pleural and pericardial effusions, respectively (Dobrovic *et al.*, 1987).

1.3 Normal bronchial epithelium cell line

NBE-E_6E_7 is an immortalized normal bronchial epithelium cell line, kindly provided by Dr. Jean Viallet (Viallet *et al.*, 1994).

1.4 Colon adenocarcinoma cell lines

CACO-2, COLO-201, COLO-205, HCT-15, LS-180 and SW1222 are cell lines derived from colon adenocarcinomas. They were obtained from the American Type Culture Collection (ATCC; Rockville, MD), except SW1222 which was kindly provided by Dr. Clifford Stanners (McGill University).

2. TISSUE CULTURE CONDITIONS

2.1 Media

Cell lines derived from NSCLCs were grown in α medium, except H157 and H-520, which were grown in RPMI-1640 (GibcoBRL). Cell lines derived from SCLCs were grown in RPMI-1640 medium, supplemented with 10% fetal calf serum (FCS) (Wisent, Inc.). NBE-E_6E_7 was grown in Keratinocyte-Serum Free medium (Keratinocyte-SFM) (GibcoBRL), supplemented with 50 µg/mL bovine pituitary extract (BPE) and 5ng/mL recombinant human epidermal growth factor (rEGF) (GibcoBRL). CACO-2, LS-180 and SW1222 were grown in α medium, supplemented with 10% FCS. COLO-201, COLO-205 and HCT-15 were grown in RPMI-1640 medium, supplemented with 10% FCS. Before use, FCS was incubated at 55°C for one hour, to inactivate Complement factors, which are heat-labile. When appropriate, medium was supplemented with a 1% mixture of streptomycin/penicillin (GibcoBRL). HCT-15 medium was supplemented with 0.1% fungisone (GibcoBRL).

2.2 Incubator conditions

All cell lines were grown in 15 cm tissue culture dishes (Falcon) in humidified incubators, at 37°C, and 5% CO_2.

2.3 Cell-passage

Adherent cells were passaged approximately every 2 to 4 days, upon reaching 80% confluence, at a dilution of 1:3 to 1:4. Cell passage comprised two rinses with phosphate buffered saline (PBS), pre-warmed at 37°C, and disassociation with trypsin-EDTA (0.05% trypsin-0.53mM EDTA-4Na) (Mediatech), for 2 to 3 minutes at 37°C. The medium was changed the following day. NBE-E_6E_7 cells were passaged with trypsin-EDTA in the same manner, and then incubated with an equal volume of 0.1% soybean trypsin inhibitor (GibcoBRL) and 0.1% bovine serum albumin (BSA) (GibcoBRL) in Keratinocyte-SFM for another 2 to 3 minutes. NBE-E_6E_7 cells were rinsed twice with pre-warmed PBS before being re-plated. Cells in suspension (COLO-201 and COLO-205) were passaged approximately every 3 days.

B. GENOMIC DNA EXTRACTION

Upon reaching approximately 90% confluence, cells were harvested, using the same conditions as for cell passage (section 2.3 of this page). Cold conditions were maintained to reduce DNAse activity. Approximately 5-10 x 10^6 cells were rinsed three times with ice-cold PBS, and resuspended gently in 500µL ice-cold resuspension buffer (10mM Tris, pH 7.9; 2mM EDTA; 400mM NaCl), in 1.5mL eppendorf tubes. Then, 20µL of a 20mg/mL solution of proteinase K (Pharmacia) and 25µL of a 10% sodium dodecyl sulphate (SDS) (GibcoBRL) solution were added, and tubes were carefully inverted five or six times. The tubes were incubated at 37°C overnight (O/N). The next day, the lysates were phenol: chloroform: isoamyl alcohol (25:24:1) extracted, and the aqueous phase was chloroform : isoamyl alcohol extracted (24:1). The genomic DNA was precipitated using 1mL of ice-cold 100% ethanol (EtOH), and spooled-out using a cold glass rod. The DNA was rinsed twice with ice-cold 70% EtOH, and once with ice-cold 100% EtOH. Finally, the amorphous pellet of DNA was resuspended in 100µL to 1mL ddH_2O, final concentration between 0.5 to 1 µg/µL. Some samples were resuspended in TE buffer (10mM Tris; 1mM EDTA) pH 8.1.

DNA concentration was determined via spectrophotometry and/or estimated via visualization using ethidium bromide agarose gel electrophoresis (see section G.1, p.69 for more details on gel electrophoresis).

C. LUNG TUMOUR DNA

Lung tumour tissue samples were extracted for DNA by previous members of this laboratory upon being freshly resected in the operating room at Notre-Dame Hospital. DNA was extracted by a standard procedure of SDS and proteinase K treatments followed by phenol:chloroform extractions. The DNA was stored at -20°C till it was used. Two samples were extracted from NSCLCs: NSC-T47 and NSC-Auc; one was extracted from an SCLC: SC-Harc.

D. GENOMIC DNA LINEARIZATION

The restriction enzymes *Hin*d III or *Pst* I (Pharmacia) were used to linearize the DNA. They were chosen because they do not have restriction sites within the sequence of interest. For every μg of DNA digested, approximately 5 units of enzyme were used. The total reaction volume was always at least 10x the volume of restriction enzyme used, to properly dilute the glycerol contained within the enzyme storage buffer. For example: 20 μL (10 μg) of DNA was digested with 3 μL *Hin*d III (17 units/μL), in 10 μL 10x concentrated ONE-PHOR-ALL buffer (Pharmacia) and 67 μL ddH2O (total reaction volume = 100 μL). Enzymatic reactions were incubated in a 37°C water-bath O/N.

The following day, enzymes were heat-inactivated at 80°C for 30 minutes, and the digestions were allowed to cool at room temperature (RT) for approximately 30 minutes. The digested DNA was then phenol : chloroform: isoamyl alcohol extracted (25:24:1).

The DNA was precipitated by adding 1/10th volume 3M sodium acetate pH 5.2, and 2.5 to 3 volumes of ice-cold 100% EtOH. The tube was inverted 6 to 7 times, placed on wet ice for 10 minutes and centrifuged at 13 000 revolutions per minute (RPM) for 30 minutes, in a bench-top centrifuge (Canlab). The supernatant was discarded, and the pellet was rinsed once with ice-cold 70% EtOH and left to air-dry for 3 to 5 minutes. The DNA (5 to10 μg) was then resuspended in 54 μL of ddH_2O. *Nota bene*: Some authors suggest that as little as 20pg of DNA (from fewer than 100 cells) is as yielding as 2μg (Clark, *et al.*, 1994). In this study, 2ng, 20ng, 200ng and 1μg consistently failed to give results, due to the harsh conditions of the chemical processing. A minimum amount of 2μg was necessary for a successful experiment.

E. SODIUM BISULPHITE CONVERSION REACTION

Principle:

The sodium bisulphite conversion reaction is a chemical process under very rigorous conditions which ultimately converts non-methylated cytosines, within an oligonucleotide, to uracils (Figure 20). Methylated cytosines remain unmodified. Thus, when modified DNA is cloned and sequenced, methylated Cs are read as Cs and unmethylated Cs are read as Ts. *Nota bene*: C-to-U conversion is highly single-strand-specific.

The original technique is known as the *Bisulphite Genomic Sequencing Protocol*, by Susan J. Clark, Janet Harrison, Cheryl L. Paul and Marianne Frommer, University of Sydney Australia (Clark *et al.*, 1994).

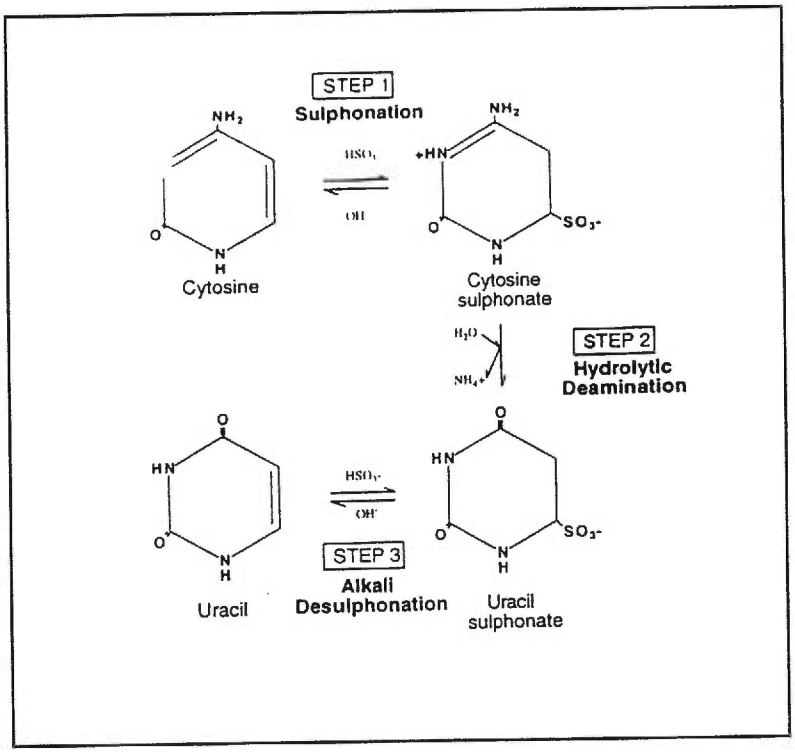

Figure 20: Schematic diagram of the bisulphite conversion reaction (taken from Clark *et al.*, 1994).

REAGENTS:

3M NaOH: FW = 40.00	6.00g of NaOH pellets (J.T.Baker) were added to enough ddH_20 for a final volume of 50 mL. The tube was inverted until all the pellets were dissolved.
3.6M $NaHSO_3$: FW = 104.06	In the dark, 3.75g of sodium bisulphite (SIGMA) were added to 8mL of ddH_20. The tube was inverted gently to avoid aeration. The pH was adjusted to pH 5.0 with 10N NaOH. The final volume was adjusted to 10 mL.
10mM hydroquinone: FW = 110.11	In the dark, a 100 mM solution was prepared by dissolving 60mg of hydroquinone crystals (Biopharm) in 50 mL ddH_20. The tube was inverted gently to avoid aeration. The 10 mM solution was prepared by diluting 5 mL of this solution in 45 mL ddH_20, dilution factor = 10.
10M Ammonium acetate: FW = 77.08	38.54g of ammonium acetate crystals (BDH) were dissolved in 50mL of warm ddH_20.

Table XV: Reagents used in the bisulphite conversion reaction.

Nota bene: All reagents must be freshly prepared. The sodium bisulphite and the hydroquinone solutions are light-sensitive: they must be prepared in the dark.

1. DENATURING THE DNA

Principle:

Bisulphite can convert cytosines as: (1) free bases, (2) nucleosides, (3) nucleotides, or (4) oligonucleotides (Hayatsu *et al.*, 1970; Wang *et al.*, 1980; Shapiro *et al.*, 1974). It does not react with double stranded DNA. Heating the DNA at melting temperature (95°C) is a modification of the original protocol.

Method:

The resuspended DNA linearization product (54 μL) was incubated at 95°C for 5 minutes and then placed on ice for 3 to 5 minutes. 6 μL of a freshly made solution of 3M NaOH was then added, for a final concentration of 0.3M NaOH. The mixture was incubated at 37°C for 20 to 25 minutes.

2. SULPHONATION REACTION AND HYDROLYTIC DEAMINATION

Principle of sulphonation:

Bisulphite reacts with cytosine (within the single-stranded oligonucleotide) to make cytosine sulphonate (Figure 20, Step 1). *Nota bene*: The forward reaction, the nucleophilic attack of an HSO4 molecule on the carbon-6 of a cytosine base, is favoured by a pH lesser than 7. The reverse reaction is favoured by a pH greater than 7. The reaction reaches equilibrium only after several hours. The pH is kept at 5.0 throughout, to ensure that the third chemical step, alkaline desulphonation, does not occur before the second step, hydrolytic deamination.

Principle of hydrolytic deamination:

Hydrolytic deamination, is the chemical conversion of the cytosine sulphone into the uracil sulphonate (Figure 20, Step 2). This reaction is catalyzed by bisulphite. *Nota bene*: The pH must remain approximately 5.0.

Method:

In the dark, 250 μL of a freshly made 3.6M solution of sodium bisulphite, pH 5.0, (SIGMA) was added to the single stranded DNA solution, final concentration 2.9M sodium bisulphite. Then, 15 μL of a freshly made 10mM solution of hydroquinone was added, final concentration 0.46mM, and the tubes were very carefully inverted 10 times. The mixture was overlaid with mineral oil (SIGMA) to capacity, to avoid oxidation, and incubated at approximately 57°C to 59°C O/N, in the dark. This temperature is a modification of the original protocol. For the samples in this investigation, 0°C (as suggested by Feil *et al.*, 1994) was attempted to maintain the DNA in a single stranded state and to reduce the extent of DNA degradation. However, conversion efficiency at this temperature was almost nil. Instead, particularly difficult samples were incubated at 95°C for 3 to 5 minutes every 3 hours for 16 hours (as suggested by Tasheva and Roufa, 1994).

3. DESALINATION

Method:

First, most of the mineral oil was carefully removed and discarded. Then, the reaction mixture was removed from the leftover mineral oil, by plunging the pipetman tip through the oil, aspirating the mixture, and depositing it onto a piece of clean paraffin paper (American National Can), to absorb residual oil. Finally, the mixture was placed in a fresh eppendorf tube without carrying over any oil. *Nota bene*: Even minute traces of mineral oil can severely hamper DNA yield upon desalination and precipitation.

The DNA was desalinated using Wizard Magic Miniprep DNA Purification Resin (Promega). 1mL of the resin solution was added to the sulphonated DNA reaction mix and the tube was inverted 6 to 7 times. The mixture was then transferred to the column by attaching a 3CC disposable sterile syringe to the luer-lock extension of the minicolumn, and pushing the slurry into the column. The DNA was washed with 2mL of the wash solution. It was finally eluted in 120 μL ddH_2O, pre-warmed at 65°C. The eluate was then speed-vacuumed to a final volume of 100 μL.

4. DESULPHONATION

Principle:

Desulphonation is the chemical reaction whereby the uracil sulphonate, under alkaline conditions, loses the sulphone adduct to become uracil (Figure 20, Step 3). *Nota bene*: Desulphonation occurs at very high pH.

Method:

10 μL of 3M NaOH was added to the DNA eluate, final concentration 0.3M NaOH. The reaction was incubated at 37°C for 20 minutes, and then cooled on ice.

5. NEUTRALIZATION, PRECIPITATION AND RESUSPENSION

0.43 volume (47 μL) of a freshly made 10M ammonium acetate solution, pH 7.0, was added to the desulphonated DNA mixture, to neutralize and precipitate it. The liquids were carefully mixed together by tapping the tube gently several times. Then, 3 volumes (471 μL) of ice-cold 100% EtOH was added, and the tube was inverted 6 to 7 times, and placed on wet ice for 10 minutes. The precipitate was then centrifuged for at least 30 minutes at 13 000 RPM. The supernatant was discarded and the pellet was washed once with 70% ice-cold EtOH. After 3 to 5 minutes of air-drying, the pellet was resuspended in 100 μL TE buffer (10mM Tris; 1mM EDTA) pH 8.1. Subsequent PCR amplifications frequently failed to yield product if the treated DNA was resuspended in ddH_2O.

F. PCR AMPLIFICATION OF THE BISULPHITE-TREATED DNA

The polymerase chain reaction (PCR) amplification of the target bisulphite-modified sequence was always carried out using the semi-nested PCR approach. A primary PCR amplification provided the template for a second PCR amplification. This second, or semi-nested, PCR amplification used the same upper oligonucleotide primer (oligo) as the first PCR amplification, but a different internal lower oligo. Different sets

of oligos were used to: (1) amplify different regions of the target sequence; and (2) increase specificity for completely bisulphite-modified DNA, as opposed to partially bisulphite-modified DNA.

1. PCR CONDITIONS

1.1 Reaction mixtures

The majority of reactions were carried out in 100µl volumes. The amount of primary PCR template, the bisulphite-treated DNA, varied between 25 and 250ng, depending on the degree of degradation and loss due to processing. The volume of nested PCR template was always 5µl of the primary PCR product.

The reaction volume consisted of the following: 5µl of the bisulphite-treated DNA; 0.5µM each primer; 40µM dNTPs (Pharmacia); 1.5 - 3mM $MgCl_2$; and 2.5 units of *Taq* polymerase (Pharmacia). Pharmacia 10X *Taq* polymerase buffer was used for all PCR amplifications requiring 1.5mM $MgCl_2$. However, when 3mM $MgCl_2$ was required, home-made Taq buffer was used (500mM KCl; 30mM $MgCl_2$; 100mM Tris-HCl, pH 8.5).

The template of the negative control for the primary PCR reaction was 5µl of water. The templates of the negative controls for the semi-nested PCR reaction were: (A) 5µl of the primary PCR negative control and (B) 5µl of water.

1.2 Thermal cycling

The majority of reactions were carried out in Hybaid DNA thermal cyclers, under the following conditions: 94°C/4 minutes x 1 cycle; 94°C/1min., 61°C/2min., 72°C/2min. x 5 cycles; 94°C/1min., 61°C/1.5min., 72°C/1.5-2min. x 25 cycles; 72°C/5min. x 1 cycle.

1.3 Primer characteristics and criteria

One of the many crucial steps in the *Bisulphite Genomic Sequencing Protocol* is the design of oligonucleotide primers. Clark has suggested that 30bp oligos which cover equally C-rich regions are two key factors in optimal amplification (Clark *et al.*, 1994). The oligos used in this project followed several guidelines. The characteristics of the "HOU" oligos were:

(1) they were all approximately 30-mers;
(2) they did not contain CpG dinucleotides;
(3) they had limited inter-oligo and intra-oligo complimentarity;
(4) they had few mismatches to the bisulphite-converted sequence;
(5) the semi-nested oligos amplified relatively long products (539 to 548 bp).

The criteria for the "BIS" oligos included characteristics 1 through 4, mentioned above, and:

(5) upper oligos all had a T (in place of a C) at their very 3' extremity, in order to increase uracil-specific extension (i.e. preferential binding and extension of the polymerase enzyme to converted DNA);
(6) they all had a relatively high C content (approximately 30%), to optimize bisulphite-converted DNA annealing; and
(7) the semi-nested oligos amplified regions just under 400bp long, for facilitation of subsequent sequencing.

1.4 Primer sequences

1.4.1 HOU oligos

Primary PCR: (1) upper oligo = HOU-1: *RAR*β position 381 (GGA GTG GAA AAA TAT ATA AGT TAT AAG GAA); (2) lower oligo = HOU-4: *RAR*β position 1250 (AAA AAA ATC CAC CCA ACT CCA TCA AAC TCT). *Semi-nested PCR*: (1) upper oligo = HOU-2: *RAR*β position 701 (GGT TGT TGG GAG TTT TTA AGT TTT GTG AGA); (2) lower oligo = HOU-4. Alternatively, (1) upper oligo = HOU-1; (2) lower oligo = HOU-3: *RAR*β position 920 (AAA ATT CTA ATC CCC CCT TTA ACA AAA AAT).

*Based on Shen et al, 1991. For positions of CpG sites #1-38, please refer to appendix.

These oligos were designed by Dr. Benoit Houle (Institut du cancer de Montréal, Hôpital Notre-Dame), and purchased at Integrated DNA Technologies, Inc. and GibcoBRL.

1.4.2 BIS oligos

Primary PCR: upper oligo BIS-1: *RAR*β position 1 (TTA GAA TAT ATA GTT GGT AAG TGG TAG ATT) with lower oligo BIS-3: *RAR*β position 377 (TAC AAA AAC AAA CAA CCA AAA AAA CAA ACA A). Semi-nested PCR: upper oligo BIS-1 with lower oligo BIS-2: *RAR*β position 320 (GTT TTT TTT AAG GGG TAG TTA TTT TTT GTT T) or HOU-1 (see section 5.1.4.1). These oligos were designed by the author, and purchased at Alpha DNA.

G. GEL EXTRACTION OF PCR PRODUCTS

1. GEL ELECTROPHORESIS AND ETHIDIUM BROMIDE STAINING

20µl of the PCR product was mixed with 4µl 6X loading buffer (0.25% bromophenol blue; 0.25% xylene cyanol FF; 50% glycerol in water) and loaded onto a 2% agarose gel (2.0g agarose in 100mL TAE (0.04M Tris-Acetate; 1mM EDTA); 0.5µg ethidium bromide/mL. 0.5 to 1 µg of the 1 kB molecular weight ladder (MW) (GibcoBRL) was loaded alongside the PCR samples in order to estimate their MW. The gel was electrophoresed in TAE electrophoresis buffer for 1 to 2 hours at approximately 80V.

2. GEL EXTRACTION

The PCR product, migrating at the expected molecular weight, was visualized using UV light, and precisely excised with a new scalpel blade. The DNA was extracted from the agarose gel, using the Sephaglas™ BandPrep Kit (Pharmacia). Three volumes of Gel Solubilizer solution were added to the agarose in a 1.5mL eppendorf tube. The tube was incubated at 50°C for 10 minutes or until the agarose had dissolved, with periodic vortexing. The solution was then cooled to RT and 10 to 20µl of Sephaglas milk was added, and the mixture was inverted 10 times. The mixture was incubated at RT for 10

minutes, with periodic agitation, and centrifuged at 13 000 RPM for 1 minute. The supernatant was discarded and the pellet was washed three times with ice-cold Wash Buffer. The DNA was eluted in 20 to 40µl of Elution Buffer for 10 minutes at 37°C.

H. MULTIPLE RESTRICTION ENZYME DIGEST

1. MRED

Principle:

Semi-nested PCR products were digested with several restriction enzymes, which had restriction sites within unmodified DNA, but not within bisulphite-modified DNA (Tables XVI and XVII). This step was an alternative plan of approach to the original protocol. It was devised to increase the probability of cloning a fully modified DNA molecule, rather than a partially modified or unmodified molecule. *The careful choice of restriction enzymes must not be biased for methylation status of CpG dinucleotides*. In other words, restriction sites should optimally not contain CpG dinucleotides. However, if they do, they should also contain another C, not within a CpG. The latter C, ultimately converted to a T upon PCR amplification, shall cause the restriction site to no longer be recognized by the restriction enzyme.

Nota bene: To ensure non-CpG cytosine conversion throughout the fragment, the restriction sites were more or less *equally distributed* along the target sequence. Lack of single standedness, the major cause of lack of conversion, may be due to (1) incomplete denaturation, (2) re-annealing of complementary strands, or (3) formation of secondary structures between complementary segments within the same strand. These three possibilities may be involve regional portions of the sequence only. Thus, the same final product may be fully converted in certain regions, and unconverted in others. For an example, please refer to Figure 23, p.82.

IMPORTANT: All of the following criteria **must** be adhered to when choosing enzymes for the preferential digestion of incompletely modified products:

(1) The restriction enzyme used does not have any sites in the bisulphite-modified sequence;

(2) The restriction site contains at least one C that is not within a CpG dinucleotide, because if it is methylated, it will not react with bisulphite and thus not convert to a uracil. Hence the restriction site will remain intact and thus be cleaved;

(3) The restriction site does not contain a single C which is also at the very 3' end, if it is adjacent to a G in the downstream sequence;

(4) The restriction site does not contain a G which is also at the very 5' end, if it is immediately preceded by a C in the upstream sequence and no other non-CpG-C.

Method:

Some of the restriction enzymes used require 2X buffer, others require 1X buffer. Thus, enzymatic reactions began in 200µl at 2X buffer and were then diluted twofold with ddH_2O to reduce the buffer concentration to 1X, upon addition of the other enzymes. Approximately 2µg of a specific semi-nested PCR product was digested in a total reaction volume of 200µl (for 2X buffer) or 400µl (for 1X buffer). The reaction mixture consisted of the following: 80µl PCR product (approximately 3µg DNA, maximum); 10-20 units each of several different restriction enzymes (please refer to Tables XVI and XVII); 10X ONE-PHOR-ALL buffer (Pharmacia), completed to 200µl or 400µl with ddH_2O. The digestions were carried out in a 37°C water-bath for 2.5 to 4 hours. Some samples were digested O/N.

HOU-1/HOU-3 MRED

RESTRICTION ENZYME	RESTRICTION SITE	#CONVERTIBLE Cs (not in CpGs) (a)	#SITES WITHIN SEQUENCE (b)	TOTAL (a) x (b)
Alu I	AGCT	1	3	3
Ava I	CCCGGG	2	1	2
Dde I	CTNAG	1	1	1
Hha I	GCGC	1	3	3
Hpa II	CCGG	1	3	3
TOTAL				**12**

Table XVI: Characteristics of restriction enzyme sites used in the Multiple Restriction Enzyme Digest (MRED) of HOU-1/HOU-3 PCR products. CpGs are underlined. Bisulphite-convertible Cs are in **boldface.** These Cs are Ts in the 100% bisulphite converted DNA PCR-amplified product, and are thus not recognized by the restriction enzymes in question.

HOU-2/HOU-4 MRED

RESTRICTION ENZYME	RESTRICTION SITE	#CONVERTIBLE Cs (not in CpGs) (a)	#SITES WITHIN SEQUENCE (b)	TOTAL (a) x (b)
Ava I	CCCGGG	2	1	2
Bam H I	GGATCC	2	1	2
Eco R I	GAATTC	1	1	1
Hae III	GGCC	2	1	2
Mbo I	GATC	1	5	5
TOTAL				12 -2* **10**

Table XVII: Characteristics of restriction enzyme sites used in the Multiple Restriction Enzyme Digest (MRED) of HOU-2/HOU-4 PCR products. CpGs are underlined. Bisulphite-convertible Cs are in **boldface**. These Cs are Ts in the 100% bisulphite converted DNA PCR-amplified product, and are thus not recognized by the restriction enzymes in question.

* 2 is subtracted from the total because the *Bam* HI site contains the *Mbo* I site.

2. MRED PRECIPITATION

1/10th volume of 3M sodium acetate pH 5.3 was added to the restriction digestion reaction volume, followed by 3 volumes of ice-cold 100% EtOH. When there was less than 0.5μg of DNA, 1μg of carrier tRNA (transfer ribonucleic acid) was also added. The mixture was mixed, quick-spun, and placed on wet ice for 10 minutes, followed by centrifugation at 13 000 RPM for 30 to 60 minutes. The supernatant was carefully removed, and the pellet was carefully rinsed with ice-cold 70% EtOH, then with ice-cold 100% EtOH. The pellet was then left to air-dry for three to five minutes, and resuspended in 20μl TE buffer, pH 8.1.

3. GEL ELECTROPHORESIS AND ETHIDIUM BROMIDE STAINING

The MRED product was mixed with 4μL transparent 6X loading buffer (50% glycerol in water) and loaded onto a 3% agarose gel (3.0g agarose in 100mL TAE (0.04M Tris-Acetate; 1mM EDTA). The gel was electrophoresed in TAE electrophoresis buffer for 2 to 3 hours at 70 - 80V. The gel was soaked in a solution of 0.5μg ethidium bromide/mL TAE buffer for approximately 30 minutes.

4. GEL EXTRACTION OF MRED

Undigested product, migrating at the expected molecular weight, was visualized using UV light, and precisely excised with a new scalpel blade. The DNA was extracted from the agarose gel, using the Sephaglas™ BandPrep Kit (Pharmacia), as previously described (Section G-2, p.69). The DNA was eluted in 20 to 40μL of Elution Buffer for 10 minutes at 37°C.

I. SUB-CLONING

1. LIGATION OF THE SEMI-NESTED PCR PRODUCT TO TA-VECTOR

Approximately 50 to 150 ng of gel-extracted MRED PCR product was ligated to home-made TA-vector (150ng). The reaction volume consisted of the following: 6μl of gel extracted PCR product; 1μL TA vector; 1μL T4 DNA ligase (1unit) (GibcoBRL); and 2μL 5X T4 DNA ligase buffer. The mixture was incubated at 15 to 16°C for 24 or 48 hours in a 0.5mL eppendorf tube.

2. HOME-MADE TA-VECTOR

Home-made TA-vector was made by blunt-cutting the plasmid pBS KS (2, 958 kB) (Promega) with the restriction enzyme *Eco* RV (Pharmacia), and then extending the 3' ends with Ts. The extension reaction conditions consisted of the following: 30μg DNA, 30Units *Taq* polymerase (Pharmacia), 2mM dTTPs and 1x buffer in a total reaction volume of 600μL. The reaction was incubated at 70°C for 2 hours. The

principle is that since PCR products frequently have additional As on their 3' ends, they can be ligated to any vector with T overhangs.

3. COMPETENT *E.coli* TRANSFORMATION

DH5 *E.coli* (GibcoBRL), rendered competent using the polyethylene glycol/dimethyl sulfonyl oxide/$MgCl_2$ method, (Chung and Miller, 1988), were transformed with 5 or 10μL of the ligation reaction volume. This strain of bacteria allows for the Blue/White colour selection of insert-containing clones. 0.1mL of competent bacteria (10^9/mL) were thawed on ice for 5 minutes. 5 or 10μL of the ligation mixture was added and gently stirred-in. The bacteria/DNA mixture was incubated on ice for 30 minutes. Then, the bacteria were heat-shocked at 42°C for 90 seconds, and cooled on ice for 2 minutes. They were then diluted in 10 volumes (1mL) of Luria Bertani (LB) medium (10g bacto-tryptone; 5g bacto-yeast extract; 10g NaCl in 1L H_2O). The culture was incubated at 37°C in a shaking incubator (225 RPM) for 1 hour. The bacteria were then plated on a LB agar petri dish containing 40μg/mL X-Gal (5-Bromo-4-chloro-3-indolyl-β-D-galactoside) and 100μg/mL ampicillin. The petri dishes were then incubated at 37°C O/N for a maximum of 16 hours.

4. PICKING COLONIES

The following day, the petri dishes were refrigerated for 3 or more hours to allow optimal colour development. All white colonies (usually 1 to 12 per ligation) were inoculated into 10 mL of LB medium (100μg/mL ampicillin), and grown till end-log phase at 37°C in a shaking incubator (225 RPM) for 15-16 hours.

5. MINIPREPS

The bacteria of 3mL of the cultures were harvested by centrifuging at 12 000 RPM for 1 minute. The pellet was resuspended in 100μL of ice-cold resuspension buffer (50mM dextrose; 10mM EDTA; 25mM Tris, pH 8.0; 100ug/mL RNAase A1). The bacteria were then alkaline lysed with 200μL of lysis buffer (0.2N NaOH; 1% SDS) on ice for 10 minutes. Then, 100μL of 3M potassium acetate (3M KOAcetate; 1.8M formic

acid) was added to precipitate cellular debris, and incubated on ice for 10 minutes. The mixture was centrifuged at 13 000 RPM for 10 minutes. The supernatant was recovered, phenol : chloroform : isoamyl alcohol (25:24:1) extracted, and precipitated with 2.5 volumes of ice-cold 100% EtOH. After 10 minutes on ice, it was centrifuged again for 20 minutes. The pellet was rinsed once with 70% EtOH and resuspended in 40μL TE, pH 8.1.

The clones were restriction enzyme digested with an appropriate enzyme, usually *Pvu* II (Pharmacia), to select clones containing the insert of the desired length. Clones containing the insert were selected for sequencing.

Insert-positive clones were sometimes digested with an enzyme previously chosen for MRED with many restriction sites within the insert, such as *Mbo* I. This is because if the MRED digestion was incomplete, some clones may be false positives in that they contain the correct length sequence, but the sequence is incompletely bisulphite-modified. Restriction enzyme digestions were put on agarose gels and their digestion pattern was compared visually to a known true-positive. These clones were selected for sequencing.

J. SEQUENCING

All sequencings were achieved through automated sequencing. Plasmid DNA was purified using Qiagen tip-20s (Qiagen), for sequencing-grade purity. Then, 10μg of plasmid DNA was extended with fluorescent T7 primer as per protocol, AutoRead ™ Sequencing Kit (Pharmacia). Half of the sequencings were executed in our laboratory, using AutoRead ™ Sequencing Kit (Pharmacia) and the Automated Laser Fluorescent A.L.F.™ Sequencer (Pharmacia LKB). The other half of the sequencings were done at Bio S&T, Inc. (Lachine, Québec), using the same methods of sequencing.

Nota bene:
The term ***"sequence clone"*** signifies the sequencing results obtained from a single subcloned PCR-amplified bisulphite-modified fragment.

K. THE HUMAN RARβ PROMOTER 2 REGION

Figure 21 illustrates the approximate location of sequences targeted for sequencing using the bisulphite genomic sequencing protocol.

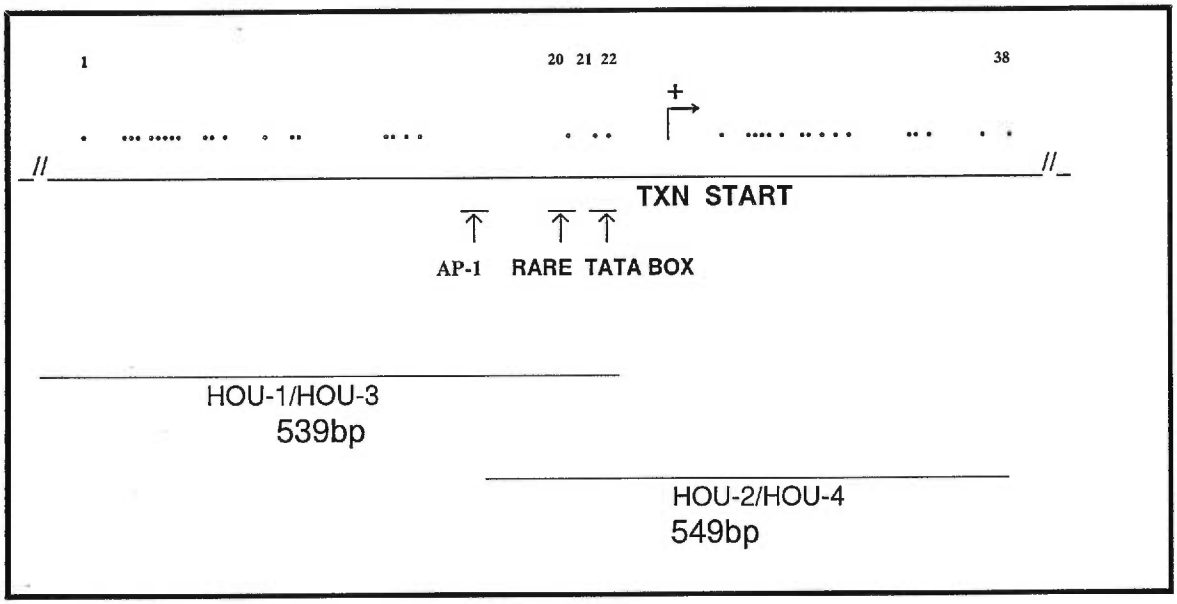

Figure 21: Schematized diagram of the human RARβ promoter 2 region targeted for bisulphite sequencing. Dots (.) represent approximate location of individual CpG dinucleotides #1 through #38; CpGs #1, 20, 21, 22 and 38 are numbered. Approximate regions for semi-nested PCR amplifications using the HOU oligos are indicated. Note that CpG #21 is within the RARE (Retinoic Acid Response Element) and that CpG #22 is located immediately upstream from the TATA Box.

III. RESULTS

A. BRIEF RECAPITULATION OF THE *BISULPHITE GENOMIC SEQUENCING* PRINCIPLE: PRELIMINARY RESULTS

1. BISULPHITE REACTS WITH UNMETHYLATED CYTOSINES WITHIN SINGLE STRANDED DNA TO FORM URACIL

The *bisulphite genomic sequencing protocol* (Clark *et al.*, 1994) is a technique which is capable of detecting *all* individual methylated cytosines within a target sequence of choice. The bisulphite conversion reaction, at optimal conditions, converts *un*methylated cytosines to uracils. Upon sequencing, *un*methylated cytosines are read as thymines, whereas 5-methylcytosines are read as cytosines.

2. THRESHOLD OF CONFIDENCE

The threshold of confidence was established at 98.6% conversion efficiency. If this threshold was not attained, then positive displays of Cs within CpG dinucleotides may have been due to the absence of conversion rather than the presence of a methyl group on carbon-5 of the cytosines in question.

EXAMPLE

Of the 74 non-CpG-cytosines within a target sequence (eg. amplified by HOU-2/HOU-4 oligos), at least 73 should be converted (to uracils in the bisulphite reaction; thymines upon amplification and sequencing) to reach the established threshold of conversion efficiency of 98.6%.

SAMPLE CALCULATION:
Sequence-clone SW1222-102

$$\frac{\text{\#converted (non-CpG) cytosines}}{\text{\#convertible (non-CpG) cytosines}} \times 100\% = \text{Conversion efficiency}$$

$$= \frac{73}{74} \times 100\% = 98.6\% \text{ Conversion efficiency}$$

3. TECHNICALITIES

Though straightforward, the protocol has important technicalities, discussed in detail in section IV. *DISCUSSION*. Briefly, the first challenge encountered was the consistent and total (or near total) loss of DNA during the bisulphite reaction and upon desalination. The second challenge was discovered when sequencings revealed multiple molecular species of various degrees of conversion.

This section begins with an overview of preliminary sequencing results, and resumes with the results obtained when the protocol was enhanced for increased conversion efficiency using the *Multiple Restriction Enzyme Digest*, or *MRED*.

B. PCR AMPLIFICATIONS OF THE TARGET SEQUENCE

Prior to using MRED, PCR products were directly subcloned and sequenced - as per original protocol (Clark *et al.*, 1994). Figure 22 is an example of PCR-amplifications of the target sequence from several different samples of bisulphite-treated DNA.

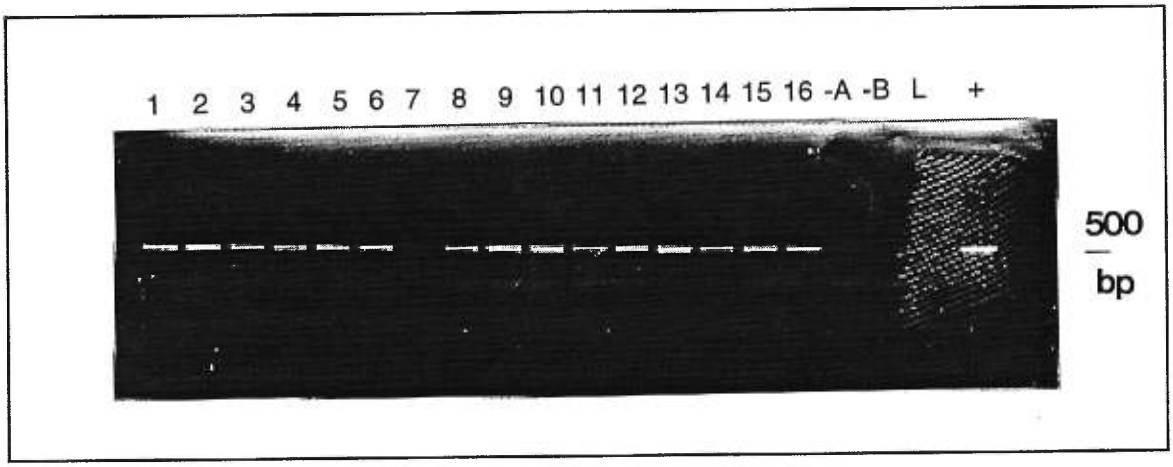

Figure 22: HOU-2/HOU-4 PCR products. 2% agarose gel electrophoresis of the semi-nested PCR products HOU-2/HOU-4, amplified from 16 different samples of bisulphite-treated cell line DNA as template. 20µl (1/5th total reaction volume). Mol. wt. of desired product = 549bp. Note the absence of non-specific bands. Sample 7 was negative. **1**, CALU-1; **2**, C-19; **3**, C-59; **4**, H-82; **5**, H-157; **6**, H-520; **7**, H-596; **8**, MM-1; **9**, RG-1; **10**, NBE-E_6E_7; **11**, CACO-2; **12**, COLO-201; **13**, COLO-205; **14**, HCT-15; **15**, LS-180; **16**, SW1222. **-A** = Negative control of primary PCR reaction, re-amplified in semi-nested PCR conditions; **-B** = Negative control of second (semi-nested) PCR reaction, i.e. template = water; **L** = 1kB Mol. wt. ladder, 500ng (GibcoBRL); + = Positive control.

C. CONVERSION EFFICIENCIES PRIOR TO MRED

As Table XVIII indicates, approximately 90% of preliminary clones chosen for sequencing were partially converted or wholly unconverted.

Efficiency of conversion	Number of clones (%)	
Unconverted:		
0%	15	(50%)
Partially converted:		
1 - 9.1%	8	(27%)
40%	1	(3%)
84 - 97%	3	(10%)
Fully converted:		
100%	3	(10%)
Total =	30	(100%)

Table XVIII: Bisulphite conversion efficiencies of 30 different sequencing clones.

Figures 23 through 25 (pages 82-84) are examples of sequencing results which illustrate the range of conversion efficiencies observed from PCR amplifications which were directly subcloned and sequenced, and not treated to MRED.

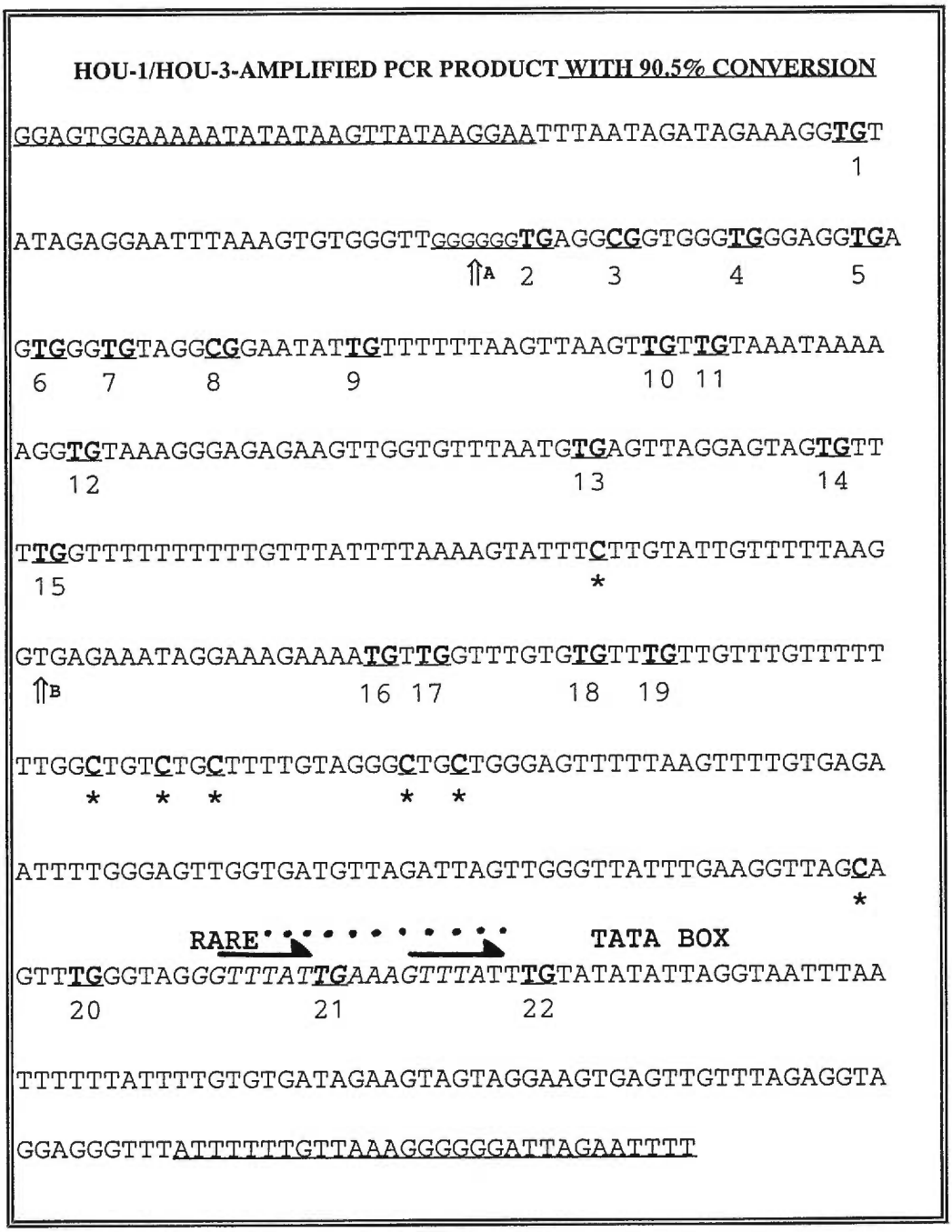

Figure 23: Example of sequencing results with partial (90.5%) conversion:

MM-1 sequencing clone 094. Oligos are underlined. 22 CpG dinucleotides are in **boldface**, underlined, and numbered; location of RARE is in *Italics*; * = unconverted C; ⇑A = 2 extra Gs in a stretch normally containing 4 Gs (insertion?); ⇑B = a T in place of an A (transversion?). Note: Of 96 cytosines, 67 of the 74 convertible - or non-CpG - cytosines are converted (90.5%); and 2 of the 22 CpG-cytosines are unconverted. Note that the unconverted Cs are within a stretch of DNA, approx. 160bp long, rather than being equally distributed throughout the sequence. The 2 apparently methylated CpGs (#3, and 8) are in a region approximately 120 bp 5' from the unconverted Cs.

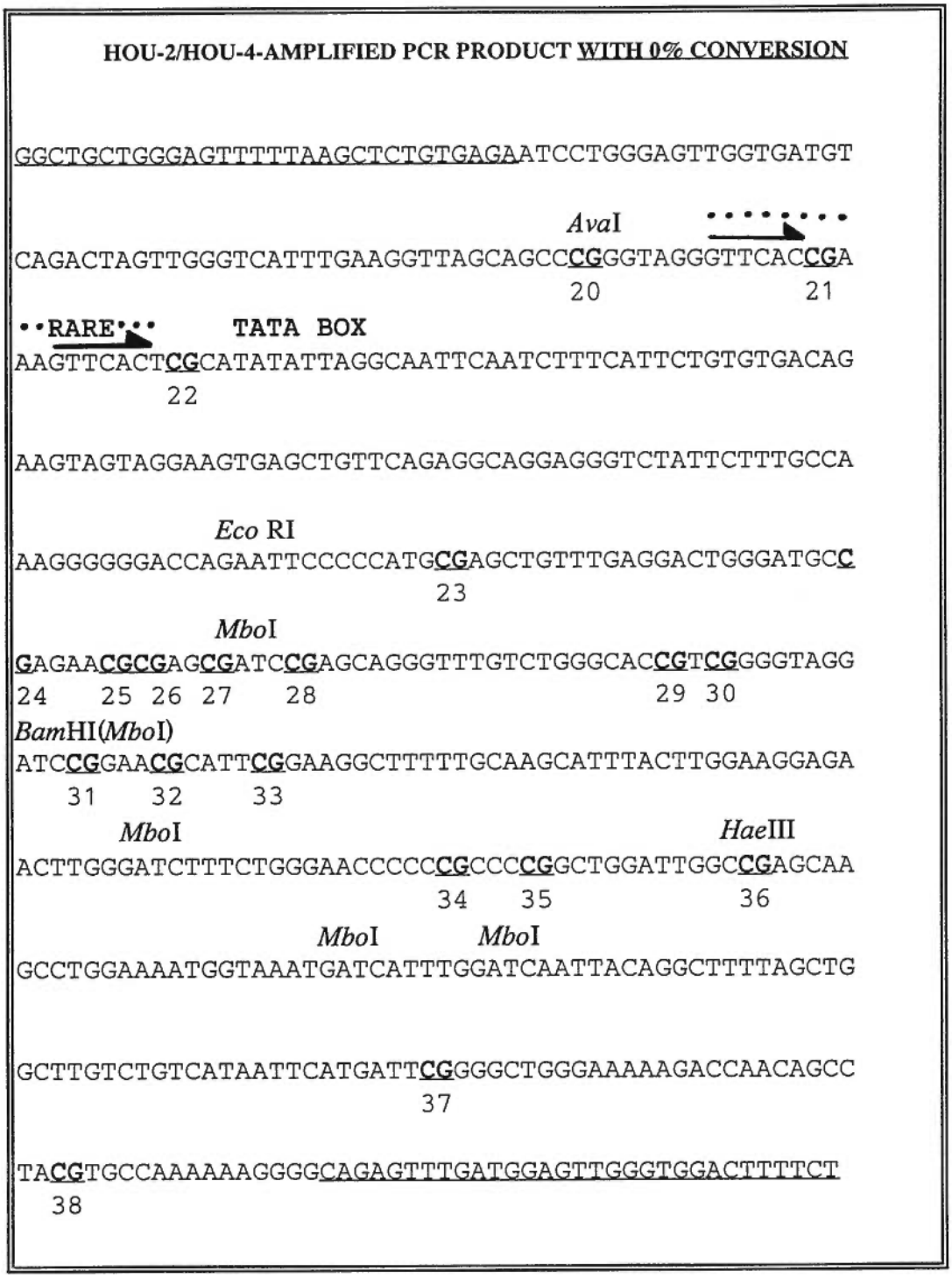

Figure 24: Example of sequencing results with 0% conversion: H-82 clone 030. Oligos are underlined. 19 CpG dinucleotides #20 through #38 are in **boldface**, underlined, and numbered. Of 99 cytosines, none are converted. Also indicated are nine restriction enzyme sites used in MRED.

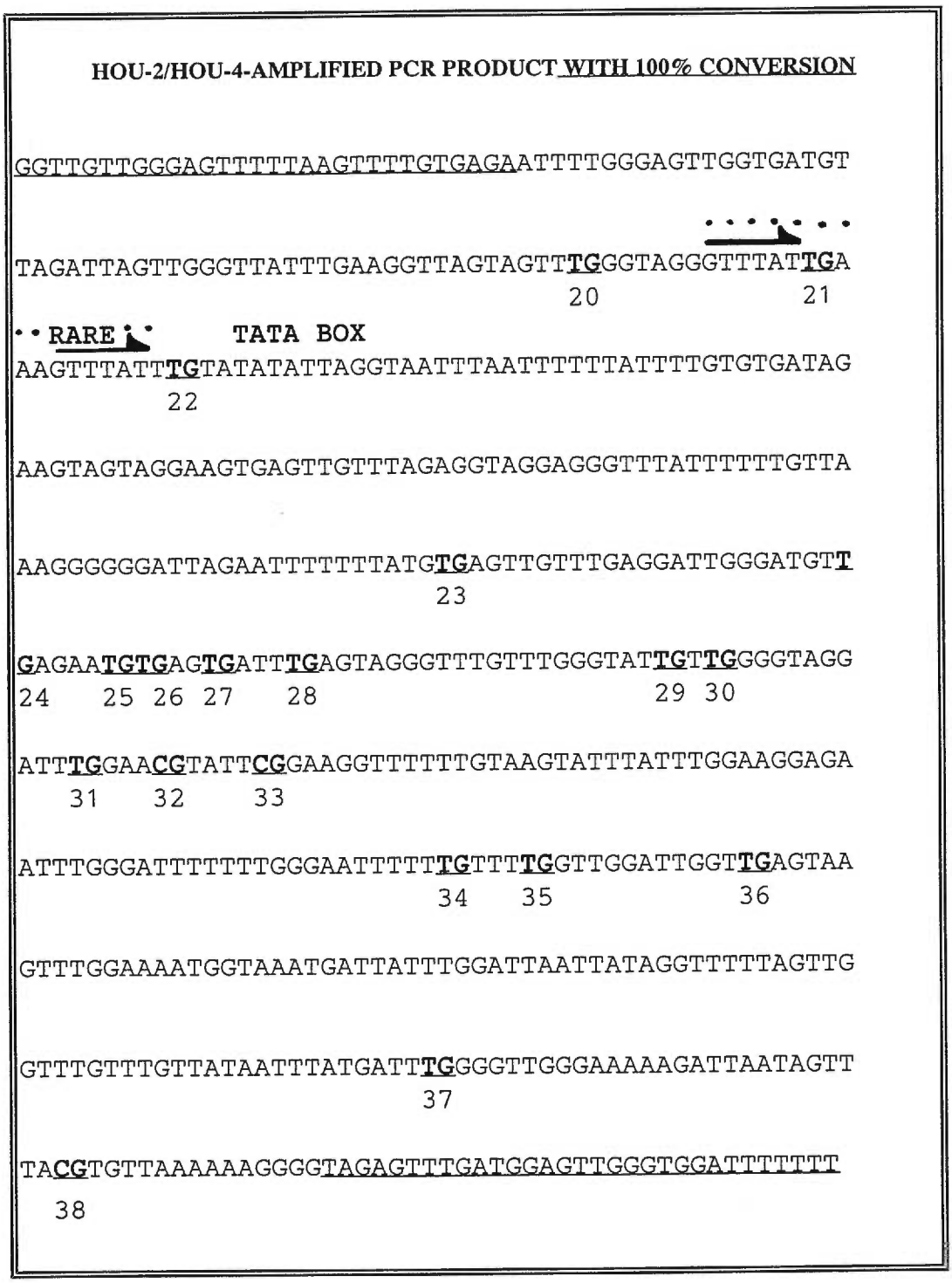

Figure 25: Example of sequencing results with 100% conversion: NBE-E_6E_7 clone 066. Note: 3 methylated CpGs (#32, 33, and 38). Oligos are underlined. 19 CpG dinucleotides #20 through #38 are in **boldface**, underlined, and numbered. Of 99 cytosines, 80 of the 80 convertible (non-CpG-cytosines) are converted (100%); and 3 of the 19 CpG cytosines are unconverted, thus methylated.

D. *MRED*-IMPROVED CONVERSION EFFICIENCIES

Rather than subcloning PCR products directly, PCR products were multiple restriction enzyme digested. Digestion patterns reveal multiplicity of species of various degrees of conversion within the bisulphite-treated DNA (Figure 26).

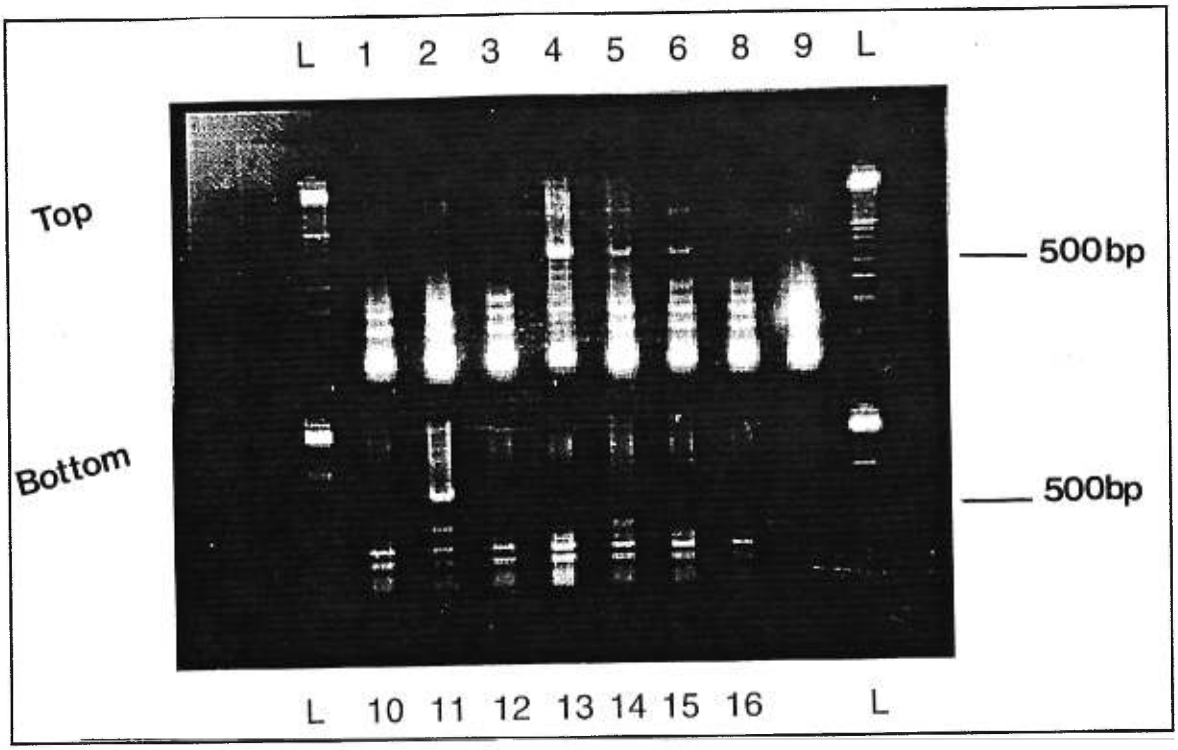

Figure 26: Multiple Restriction Enzyme Digest of the HOU-2/HOU-4 PCR products. 3% agarose gel electrophoresis of the multiple restriction enzyme digest (MRED) of 80µl (4/5 total reaction volume) of the PCR products shown in Figure 22, excluding sample 7. Note the apparent complete digestion of the desired band (549bp) for samples 1-3, 8, 9, 12-16. Samples 4, 5, 6, 10 and 11 remain positive for the desired Mol. wt. band. Note the variable proportion of digestion in positive samples. **L** = 100bp Mol. wt. ladder, 1µg (GibcoBRL).

Of the samples treated with the MRED- Improved Bisulphite Genomic Sequencing Protocol, 87% of sequencing clones were efficiently converted (Table XIX). That is more than the 10% observed without MRED.

Efficiency of conversion	Number of clones	(%)
Incorrect sequence:		
	2	(4%)
Unconverted:		
0%	0	(0%)
Partially converted:		
<10%	1	(2%)
90.5 - 95.9%	3	(7%)
Significantly converted:		
98.6%	9	(20%)
100%	31	(67%)
Total =	46	(100%)

Table XIX: Bisulphite conversion efficiencies of 46 different sequencing clones from the MRED-Improved Bisulphite Genomic Sequencing Protocol.

E. METHYLATION ANALYSES

1. METHYLATION ANALYSIS OF KNOWN RARβ-EXPRESSING LUNG CANCER CELL LINES

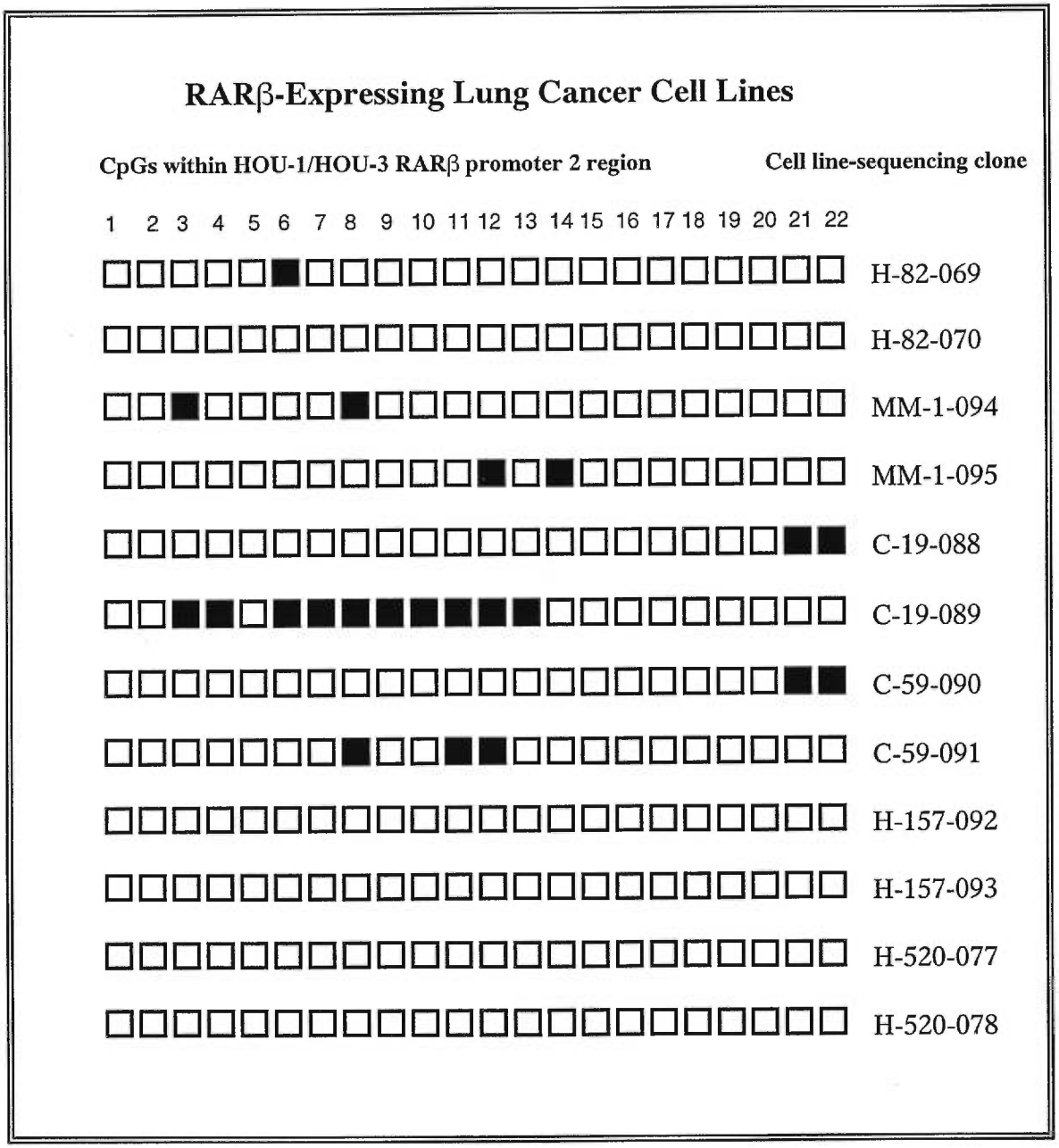

Figure 27: Schematic diagram of the methylation status of CpGs #1 through #22 (within the HOU-1/HOU-3 region) of the RARβ promoter 2 region in RARβ-expressing lung cancer cell lines. ■=methylated; □= unmethylated.

2. METHYLATION ANALYSIS OF KNOWN RARβ-*NON*-EXPRESSING LUNG CANCER CELL LINES

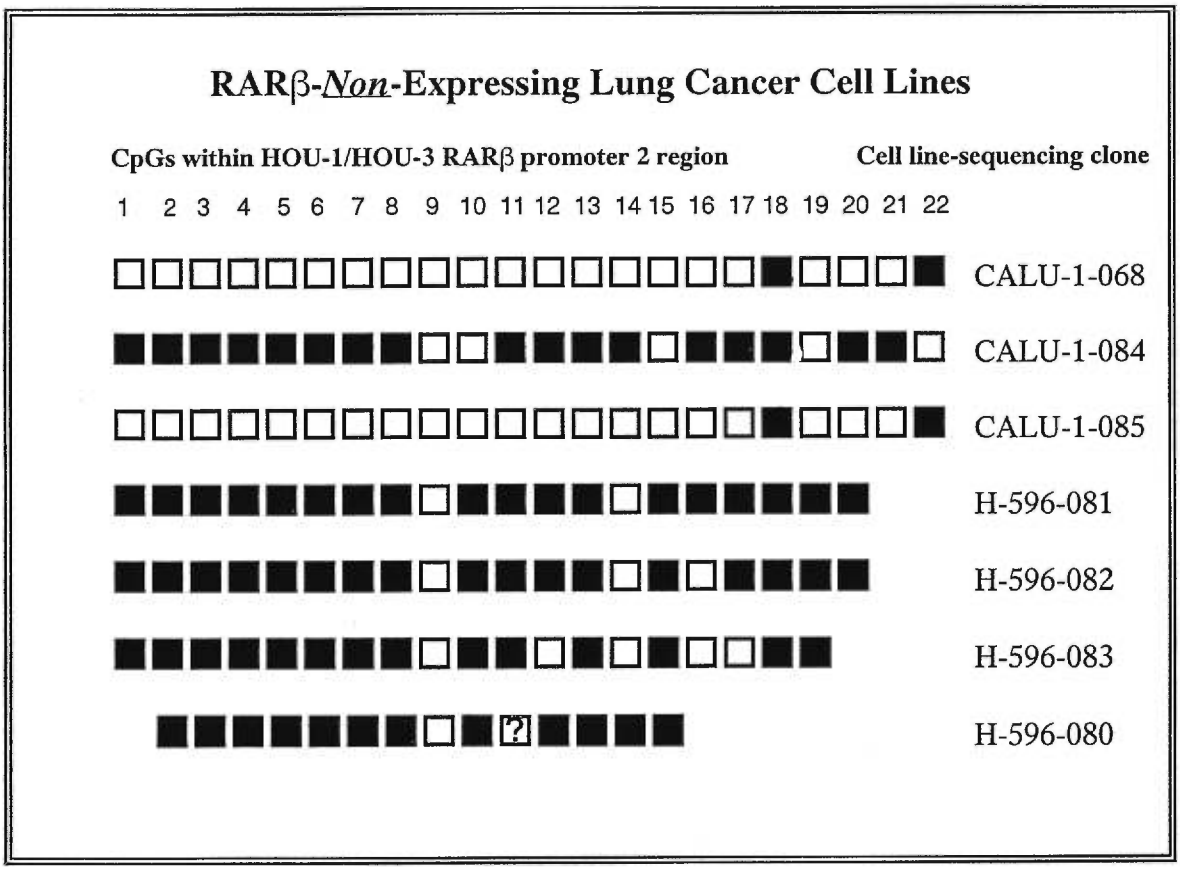

Figure 28: Schematic diagram of the methylation status of CpGs #1 through #22 (within the HOU-1/HOU-3 region) of the RARβ promoter 2 region in RARβ-*non*-expressing lung cancer cell lines. ■=methylated; □= unmethylated; ?= undetermined status due to automatic sequencer ambiguity. Lack of a box means that this region was not sequenced.

3. METHYLATION ANALYSIS OF A KNOWN RARβ-EXPRESSING COLON CANCER CELL LINE

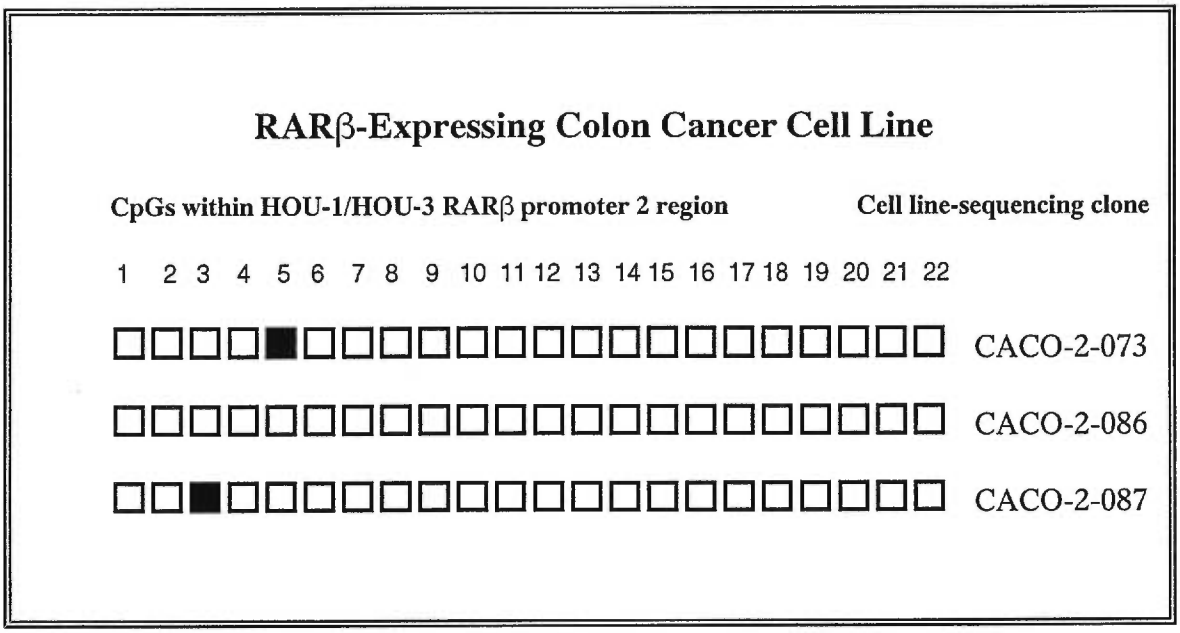

Figure 29: Schematic diagram of the methylation status of CpGs #1 through #22 (within the HOU-1/HOU-3 region) of the RARβ promoter 2 region in the RARβ-expressing colon cancer cell lines CACO-2. ■=methylated; □= unmethylated.

4. METHYLATION ANALYSIS OF KNOWN RARβ-*NON*-EXPRESSING COLON CANCER CELL LINES

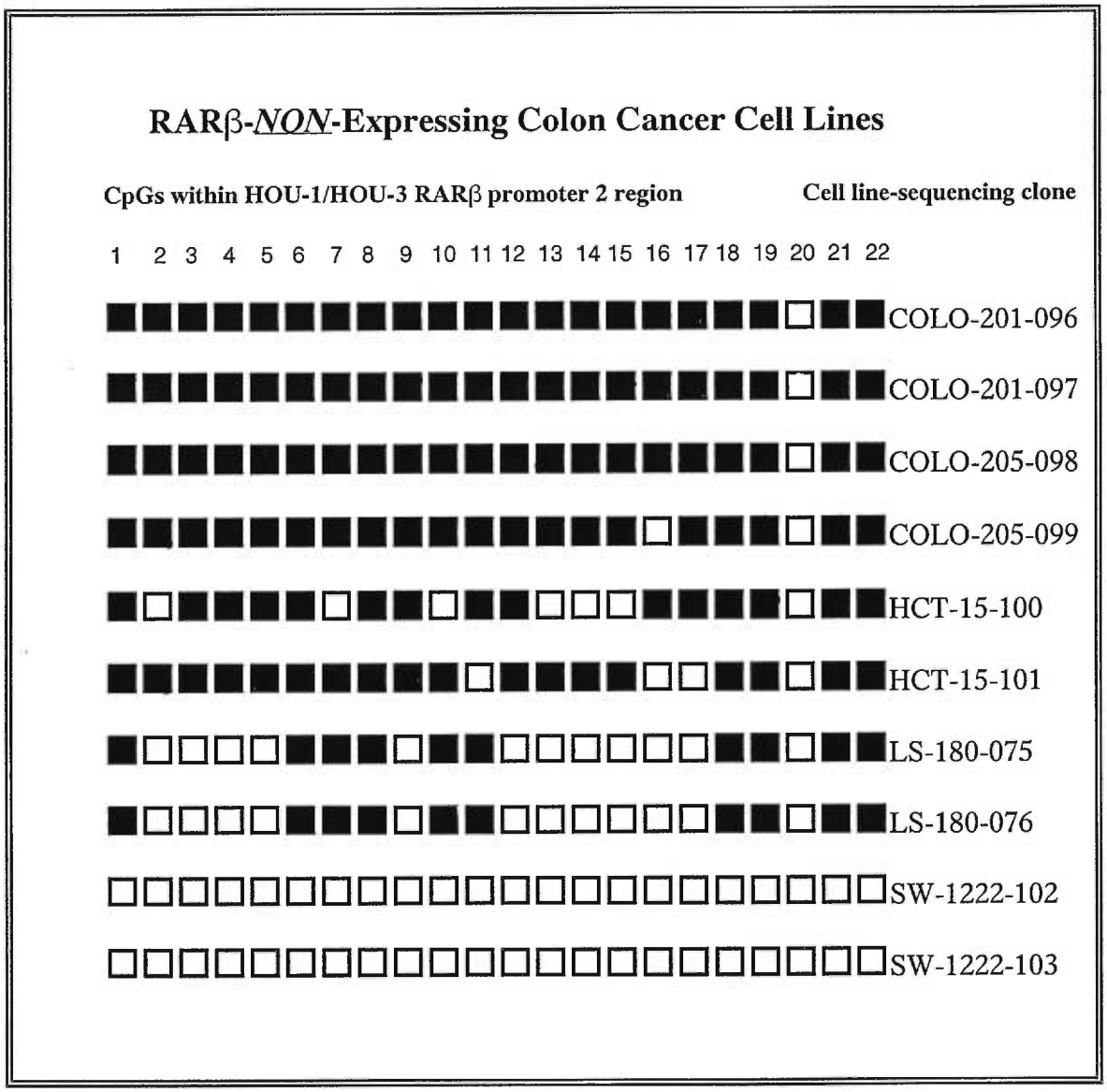

Figure 30: Schematic diagram of the methylation status of CpGs #1 through #22 (within the HOU-1/HOU-3 region) of the RARβ promoter 2 region in RARβ-*non*-expressing colon cancer cell lines. ■=methylated; □= unmethylated.

5. DOWNSTREAM METHYLATION ANALYSIS OF VARIOUS CELL LINES

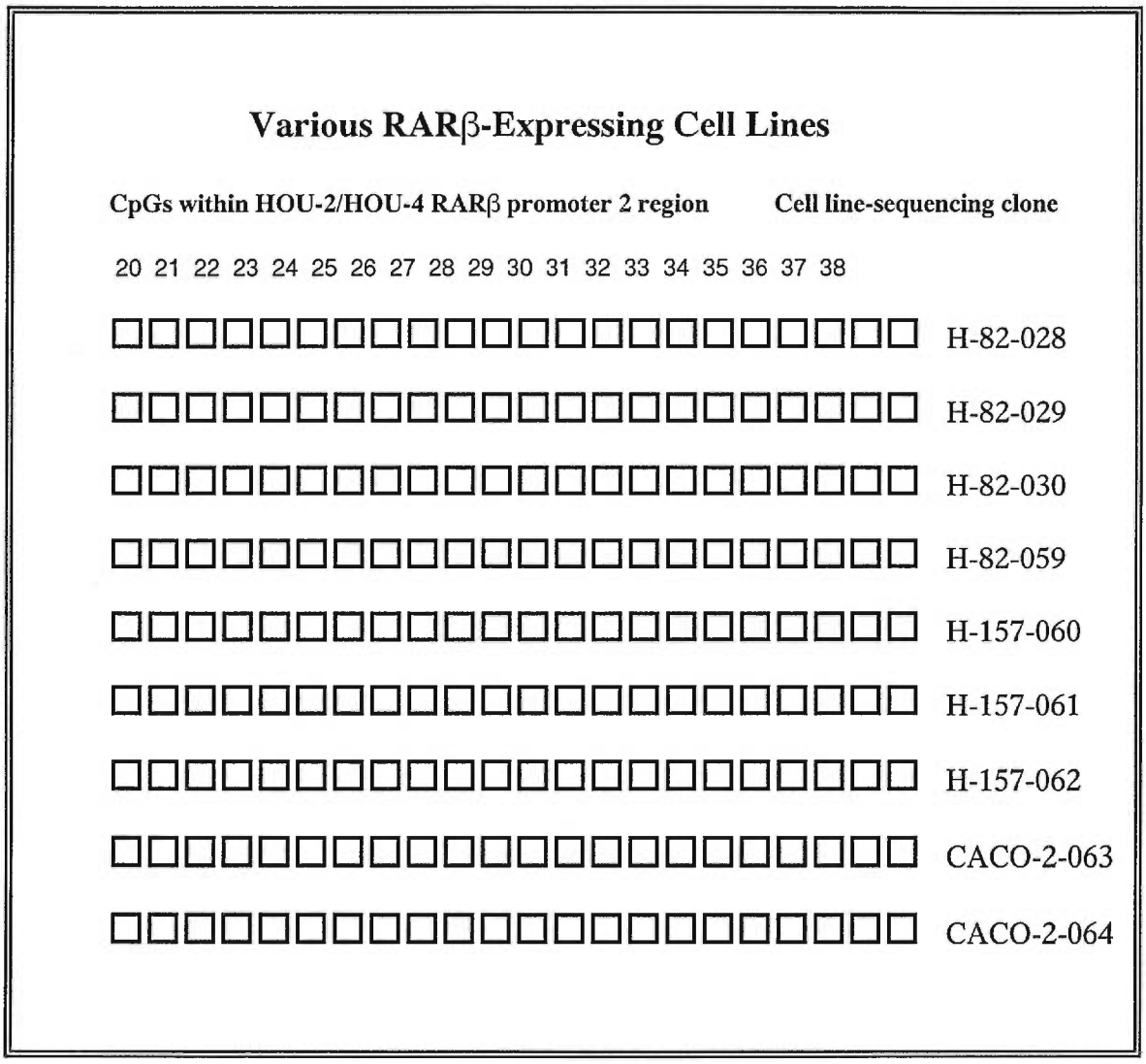

Figure 31: Schematic diagram of the methylation status of CpGs #20 through #38 (within the HOU-2/HOU-4 region) of the RARβ promoter 2 region in various RARβ-expressing cell lines. ■=methylated; □= unmethylated.

6. METHYLATION ANALYSIS OF LUNG TUMOURS

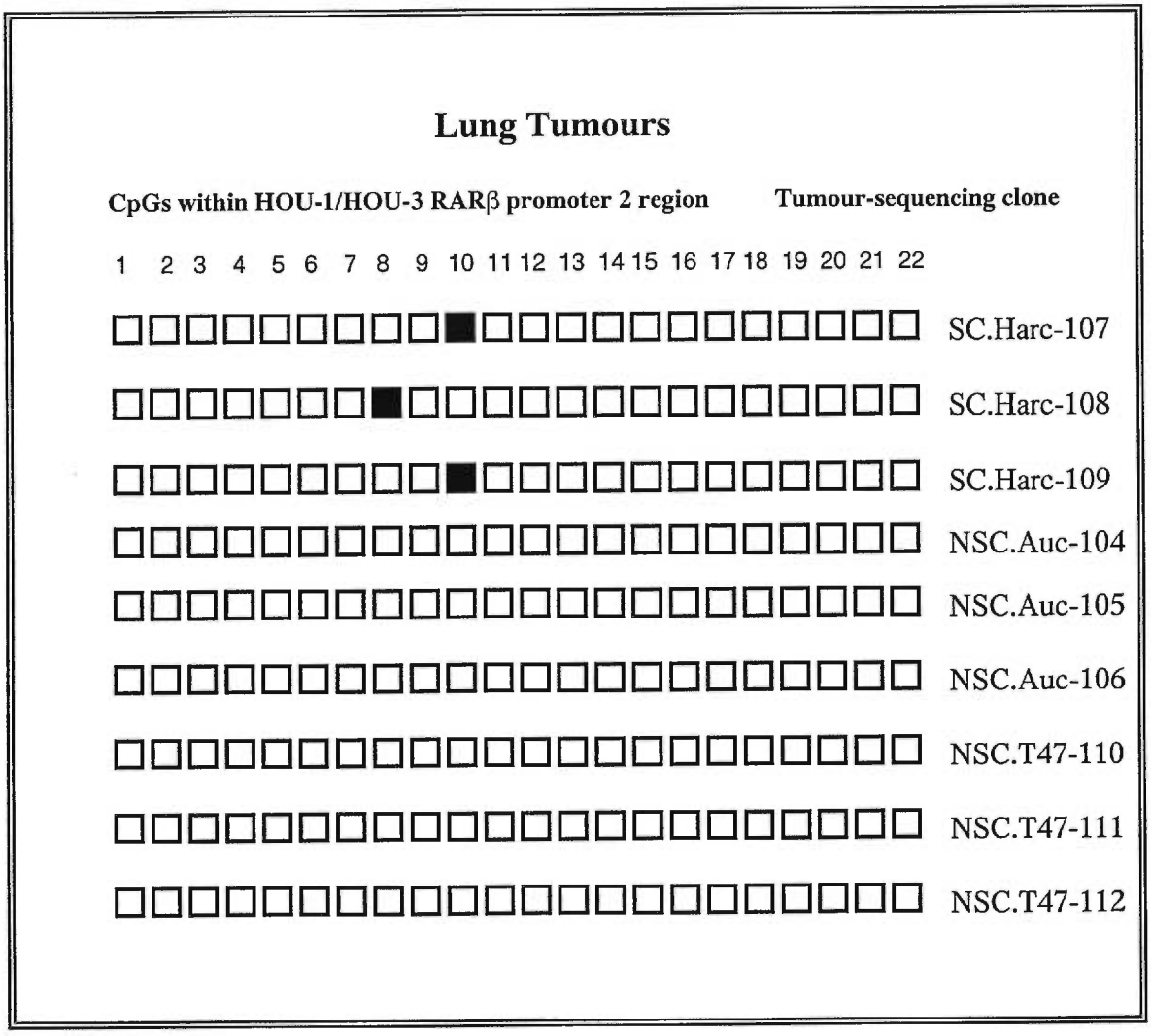

Figure 32: Schematic diagram of the methylation status of CpGs #1 through #22 (within the HOU-1/HOU-3 region) of the RARβ promoter 2 region in various epithelial lung tumours. **NSC** = Non-Small Cell Lung Carcinoma; **SC** = Small Cell lung Carcinoma; ■=methylated; □= unmethylated.

F. DIFFERENCES BETWEEN EXPECTED RARβ SEQUENCE AND OBSERVED SEQUENCING RESULTS

A variety of unexpected differences were observed in the sequencing results when compared to the expected sequence of RARβ. Table XX is a compilation of these differences regardless of position, as "difference-types". Table XXI lists the precise positions of these differences according to the positions indicated on Figure 23, p.82.

DIFFERENCE-TYPE	Occ.	f (%)
2 EXTRA Gs at 75	40	(100%)
A→T at 252	40	(100%)
A→G	20	(50%)
T→C	8	(20%)
T→G	3	(8%)
EXTRA C	2	(5%)
MISSING T	2	(5%)
G→T	1	(3%)
G→A	1	(3%)
T→A	1	(3%)
A→T	1	(3%)
G→C	1	(3%)

Table XX: Observed frequency of a difference-type, (i.e. a difference between the expected sequence and the observed sequencing results regardless of position) among 40 different sequencing clones. The two extra Gs *circa* position 75 and the A→T at position 252 were observed in all sample sources (see Figure 23, p.82). All other differences, when considering position, were observed in different sample sources. **Occ.** = occurrence; **f** = frequency of occurrence.

37 POSITION-DIFFERENCES OBSERVED IN 40 SEQUENCE-CLONES OBTAINED FROM 17 DIFFERENT DNA SAMPLES

NO.	POSITION	DIFFERENCE	OCCURENCE	f (%)	SAMPLE-SEQUENCE-CLONE
1	31	T→C	1	(3)	
2	32	T→A	1	(3)	
3	*34*	*A→G*	*1*	*(3)*	
4	41	A→G	1	(3)	
5	55	A→G	1	(3)	
6	**75**	**2 EXTRA Gs**	**40**	**(100)**	**ALL SAMPLES**
7	105	G→A	1	(3)	
8	145	A→G	1	(3)	
9	*157*	*A→G*	*1*	*(3)*	
10	158	A→G	1	(3)	
11	180	A→G	1	(3)	
12	189	A→G	1	(3)	
13	195	A→G	1	(3)	
14	198	T→G	1	(3)	
15	**214**	**MISSING T**	**2**	**(5)**	**COLO-201-096; HCT-15-100**
16	**214**	**T→G**	**2**	**(5)**	**LS-180-075; HCT-15-101**
17	*219*	*A→G*	*1*	*(3)*	
18	224	A→G	1	(3)	
19	*225*	*A→G*	*1*	*(3)*	
20	**252**	**A→T**	**40**	**(100)**	**ALL SAMPLES**
21	258	A→G	1	(3)	
22	*260*	*A→G*	*1*	*(3)*	
23	301	A→G	1	(3)	
24	342	T→C	1	(3)	
25	*356*	*G→C*	*1*	*(3)*	
26	359	A→G	1	(3)	
27	388	T→C	1	(3)	
28	428	EXTRA C	1	(3)	
29	440	EXTRA C	1	(3)	
30	440	G→t	1	(3)	
31	**453**	**T→C**	**2**	**(5)**	**CALU-1-085; CACO-2-086**
32	**455**	**T→C**	**2**	**(5)**	**CALU-1-085; COLO205-098**
33	472	A→G	1	(3)	
34	475	A→G	1	(3)	
35	482	A→G	1	(3)	
36	484	T→C	1	(3)	
37	494	A→G	1	(3)	

Table XXI: Observed position-differences between expected RARβ sequence and observed sequencing results. Differences observed in tumour samples are in *Italics*; differences observed in cell lines are not. In **boldface** are differences observed more than once. ***Nota bene***: when a difference was observed once, sample-sequence-clone is not indicated. **f** = frequency of occurrence, estimated as "number of sequence clones with a particular position difference/total number of sequence-clones". t = sequencing ambiguity most likely a T.

IV. DISCUSSION

A. METHYLATION ANALYSES

Genetic and epigenetic alterations involving the short arm of chromosome 3 are frequent molecular events in lung cancer. As discussed in the introduction, it has been shown that 85 to 90% of squamous cell lung carcinomas have deletions spanning 3p24-ter, and 100% of SCLCs have deletions spanning 3p13/14-ter (Houle *et al.*, 1991). These regions include 3p24, the locus of the putative tumour suppressor gene, *RAR*β. Moreover, lack of RARβ expression has been shown in numerous lung and colon cancer cell lines (Houle *Ph.D Thesis*, 1991), and may contribute to the cancer process.

Hypermethylation of CpG islands within gene promoters is frequently correlated with a decrease or lack of gene expression. In fact, it has been shown to be associated with the lack of expression of *four* known tumour suppressor genes: *CDKN2/p16*, *VHL*, *Rb* and *WT*-1 (Merlo *et al.*, 1995; Herman *et al.*, 1994; Ohtani-Fujita *et al.*, 1993; and Royer-Pokora and Schneider, 1992, respectively), and with the putative tumour suppressor gene *HIC-1* (Eguchi *et al.*, 1997). DNA hypermethylation is an epigenetic event which may contribute to the multistage carcinogenic process. Since two inactivating events are required for tumour suppressor loss of function, RARβ allele loss through chromosomal deletion may comprise *one* event, and RARβ2 promoter hypermethylation may constitute the *other* event.

If indeed hypermethylation negatively affects *RAR*β2 expression in the cancer process, then its reversibility may be targeted pharmacologically in both chemoprevention and anticancer therapy. Methylation inhibitors, such as methyl transferase antisense oligodeoxynucleotides, or non-methylatable analogs, such as 5-aza-2'-deoxycytidine, have been shown to be associated with reversal of the transcription block associated with promoter hypermethylation (eg. Ramchandani *et al.*, 1997; Merlo *et al.*, 1995, respectively).

Thus, determining the role of hypermethylation of the RARβ2 P2 promoter in the inactivation of RARβ2 in cancer cell lines not only adds to the molecular biological understanding of carcinogenesis, with a particular focus on lung carcinogenesis, the original subject of this project, but also provides key information that may be exploited in chemoprevention, cancer prognoses, and anti-cancer therapy.

1. OBJECTIVE 1

The first aim of this project was to determine the status of CpG methylation in the P2 promoter of the *RAR*β gene in lung cancer cell lines, colon cancer cell lines, and lung tumour cancer cells. This objective was attained for 17 samples: eight of the nine lung cancer cell lines investigated, all six colon cancer cell lines investigated, and all three tumour DNA samples investigated.

The samples were mainly analyzed for methylation at 22 individual CpG sites located in the RARβ P2 promoter, 5' of the transcription start (Figure 21, p.77). The technique which was used was a modified version of the *Bisulphite Genomic Sequencing Protocol* (Clark *et al.*, 1994), here called the *MRED-Improved Bisulphite Genomic Sequencing Protocol* (MRED = Multiple Restriction Enzyme Digest).

The 17 samples successfully sequenced were: *NSCLC cell lines*: (1) CALU-1, (2) C-19, (3) C-59, (4) NCI-H157, (5) NCI-H520, (6) NCI-H596; *SCLC cell lines*: (7) NCI-H82, (8) MM-1 ; *Colon adenocarcinoma cell lines*: (9) CACO-2, (10) COLO-201, (11) COLO-205, (12) HCT-15, (13) LS-180, (14) SW1222; *lung tumour tissue samples*: (15) NSC-Auc, (16) SC-Harc and (17) NSC-T47. The two cell lines which were not successfully sequenced (see below paragraph) were (1) NBE-E_6E_7, a normal bronchial epithelial cell line (Viallet *et al.*, 1994) and (2) RG-1, an SCLC cell line derived from pericardial effusion (Dobrovic *et al.*, 1987).

The lack of results representing normal tissue, such as NBE-E_6E_7, means that results from cancer cell lines cannot be compared to a normal cell line control. Numerous attempts were made to bisulphite-sequence the *RAR*β promoter in NBE-E_6E_7 cells, but all failed. In a last attempt, BEAS-2B normal lung cells (kindly provided by Dr. C. Harris) were also bisulphite-treated, still without success. It is not known why some samples do not yield PCR-amplifiable bisulphite-modified DNA, however it is one of the challenges which is sometimes encountered (Clark *et al.*, 1994; personal communication, Dr. Benoit Houle, Hôpital Notre-Dame, Montréal, Québec).

It is noteworthy to mention that of the cell lines successfully sequenced, many can be used as negative controls for methylation, because as seen in the previous section (II. *Results*) and discussed in detail below, several cell lines (including lung cancer and colon cancer cell lines) *express RAR*β and *are hypomethylated*. It is reasonable to hypothesize

that normal bronchial epithelium is hypomethylated at the RARβ P2 promoter, because (a) normal lung cells generally express RARβ (Houle *et al.*, 1991) and (b) hypomethylation is generally accepted to be associated with gene expression (Szyf, 1996). Furthermore, methylation of promoter-associated CpG island DNA outside of the inactive X chromosome and imprinted domains is not recognized as a normal state, even for unexpressed genes. Still, further experimentation with normal tissue samples must be completed before supporting this hypothesis. Nonetheless, since both hypomethylation and hypermethylation were observed among the treated samples, as discussed in the following section, the significance of methylation status regarding expression can still be ascertained.

2. INTERPRETATION OF RESULTS

Nota bene:
The term ***"sequence clone"*** signifies the sequencing results obtained from a single subcloned PCR-amplified bisulphite-modified fragment. Thus, a sequence clone is taken to represent an *RAR*β allele.

2.1 RARβ-*non*-expressing lung cancer cell lines

NCI-H596 was found to be heavily methylated (up to 90% of CpGs #1-#20) in all four clones sequenced. Interestingly, though the patterns of methylation differed slightly, CpG #9 was unmethylated in all four clones, and CpG #14 was unmethylated in three of the four clones, suggesting the possibility for a site-specific-methyl-CpG role in transcription block, and/or a CpG-site-specific process of enzymatic methylation and/or demethylation. Though CpG density and promoter strength have been shown to play a role in gene repression (Boyes and Bird, 1992), studies regarding the role of sequence-specific CpG-cytosine-methylation have not yet been published/executed (as far as the author is aware). However, one study revealed that hypomethylation at specific CpG positions within the rat prolactin gene (r*PRL*) was correlated with elevated gene expression (Ngo *et al.*, 1996). Presently, methylation status-specific functions, such as lack of gene expression, appear to involve CpG *methylation*, rather than lack thereof. For example, the methyl CpG binding protein 1 (MeCP1) binds DNA sequences which

contain at least 12 methylated CpGs (Meehan *et al.*, 1989). MeCP2 binds a single CpG base pair which is methylated on both strands (Lewis *et al.*, 1992). MeCP1 and 2 are transcription repressors (Boyes and Bird, 1991; Nan *et al.*, 1997, respectively).

It can be argued that some clonal inserts may be unintentionally PCR-amplified from the *same* template molecule of bisulphite-treated DNA. Thus, it is possible that different sequence clones may not actually represent different DNA fragments. The NCI-H596 sequence clones mentioned above, were sufficiently different from each other (i.e. a minimum difference in the methylation status of one CpG, from one clone to the other), that for this sample, that is not the case. For this reason and the fact that all samples were amplified and subcloned using the same methodology, it was generally accepted that similar or identical patterns of methylation observed for other sample sources also represented different DNA fragments. However, the possibility that two or more identical sequence clones represent the same molecule remains, and must be kept in mind.

CALU-1 was hypermethylated (77% of CpGs #1-#22) in *one* of three clones, but only 10% methylated in the other *two* clones. First, the pattern of methylation was identical in the latter clones, in that CpGs #18 and #22 were methylated. This again suggests the possibility for site-specific CpG methylation. Second, the results obtained are unexpected. This is because the degree of methylation is logically anticipated to be elevated *in all clones* analyzed, since (a) CALU-1 cells do *not* express RARβ and (b) one clone was indeed hypermethylated. These results suggest that there are mechanisms other than chromosomal deletion and P2 methylation involved in RARβ2 repression.

Relatively recent findings suggest that certain lung cancer cells, such as CALU-1, may have *two different mechanisms* of RARβ turn-off. First, evidence suggests that a *trans*-acting transcriptional defect affecting the retinoic acid response element (RARE) exists in most lung cancer cell lines (Moghal and Neel, 1995). Such a *trans* defect should by definition affect *both* alleles, so the above-mentioned observation needs an explanation. It is unlikely that these results are simply a laboratory error (eg. mistakenly labeled samples), because of the results obtained for C-19 (see below). Second, our laboratory has shown that CALU-1 is heterozygous for the *thyroid hormone receptor* gene, also located in 3p24 (Manon Pelletier, unpublished results). This suggests that one *RARβ* allele may remain genetically unmodified (as opposed to *epi*genetically modified). Thus, hypermethylation may be one of the mechanisms involved in RARβ turn-off, and

the *trans*-acting transcriptional defect affecting the RARE may be the other. Moreover, hypermethylation may precede the *trans*-acting defect.

If this is the case, an interesting avenue of approach would be to identify which event precedes the other, and if they are associated. In light of the multistage process of carcinogenesis proposed by Vogelstein in 1990 (see Figure 12, p.20), precancerous and progressively cancerous lesions could be biopsied and analyzed for RARβ hypermethylation and other altering events, to determine at which precise stage if any, hypermethylation normally takes place, and, if it antedates other lesions at this locus. Furthermore, though hypermethylation may be considered to be a tumour *promoter*, is it not possible that promoter methylation predates tumour promotion and perhaps even tumour initiation? Its high mutability alone may facilitate genetic alterations such as point mutations, let alone expression. A recent study has shown that hypermethylation at the *D17S5* locus (*HIC-1*, candidate tumour suppressor gene) in NSCLCs correlated with the differentiation grade of tumours and was associated with smoking history (Eguchi *et al.*, 1997). If the inhalation of smoke combustion is causally-associated with hypermethylation, and hypermethylation is a powerful (but reversible) tumour promoter, then lung carcinogenesis may be prevented or reversed.

2.2 RARβ-expressing lung cancer cell lines

NCI-H520 and NCI-H157 were found to be completely unmethylated at CpGs #1 through #22. NCI-H520 cells express RARβ, but at very low levels (Houle *Ph. D Thesis*, 1991; Johane Morin unpublished results). Thus, for NCI-H520, the complete absence of methylation within the target sequence shows that reduced expression is not necessarily associated with methylation, even at a low degree (i.e. it is not a *linear* inverse association). The NCI-H157 cell line is inducible for RARβ expression and has been shown to express RARβ after being in culture for a lengthy period of time, i.e. weeks (Houle *et al.*, 1991). The cells used for genomic DNA extraction in this study were harvested as soon as they reached approximately 90% confluence (i.e. after approximately two to three weeks in culture). Thus, it is likely that they expressed RARβ, but RARβ expression was not determined at that time. Nevertheless, hypomethylation may reflect RARβ expression inducibility.

NCI-H82 and MM-1, the two SCLC cell lines studied, were found to be hypomethylated (0%-9% of CpGs #1-#22). In each case, both clones analyzed per sample differed in their pattern of methylation and thus did not suggest that any particular CpG site predominated for methylation over any other. It is important to note that NCI-H82 is inducible for RARβ expression upon retinoic acid (RA) treatment (Houle *Ph. D Thesis*, 1991). Hypomethylation therefore may again reflect this cell line's inducibility for RARβ expression, as previously seen for NCI-H157. This makes sense because transcription factors presumably have to have access to the promoter sequences immediately upon RA administration.

C-59 was found to be up to 14% methylated at most, and displayed different patterns of methylation among the two clones analyzed.

C-19, recently shown to have lost RARβ expression (personal communication, Johane Morin), presented very different patterns of methylation. More precisely, one clone was lightly methylated at (9%) and the other was moderately methylated (45%). This conspicuous dissimilarity among sequence clones obtained from a single cell line may represent several types of phenomena.

First, recent loss of RARβ expression (determined around the time of harvest) may not have affected all cells within the culture. In other words, some cells may have belonged to a clonal population derived from RARβ-expressing C-19 cells, and others may have been derived from RARβ-*non*-expressing C-19 cells. Biochemical heterogeneity is an important characteristic of cancer cells. Thus, the sequencing clones may represent the heterogeneity of RARβ expression at the early times immediately after the loss of expression began. If frozen harvests of C-19 cells, prior to loss of RARβ expression remained, then the first hypothesis of heterogeneity could be tested by repeating the bisulphite protocol. Unfortunately, this is not the case.

Second, the fact that the C-19 cell line is a derivative of CALU-1, supports the hypothesis mentioned above, that *two mechanisms* may play a role in RARβ turn-off, one of them being hypermethylation, the other being a *trans*-acting transcriptional defect affecting the RARE. Indeed, the actual loss of RARβ observed in these cells may be a direct result of the hypothesis of two mechanisms.

2.3 Conclusion 1: Lung cancer cell lines

In most cases, *hyper*methylation of the RARβ promoter 2 was associated with *lack* of RARβ expression, and the reverse was also true, *hypo*methylation was associated with *expression*. Overall, 12 of 13 RARβ-expressing lung cancer cell line sequence clones were hypomethylated (0-14%). The exception was one of two clones derived from C-19 which was 45% methylated, but as mentioned above, this may be associated with the recent loss of RARβ expression in this cell line, and/or the existence of a second mechanism involved in the inactivation of RARβ. Of the RARβ-*non*-expressing lung cancer cell lines, 5 of 7 sequence clones were hypermethylated (77-90%). However two of the three sequence clones derived from CALU-1, an RARβ-*non*-expressing cell line, were only 9% methylated. The hypothesis that promoter methylation is correlated with absence of gene expression is thus supported for all the lung cancer cell lines investigated, with the exceptions concerning CALU-1 and its C-19 transfection derivative, mentioned above.

It is natural to postulate that if there are two mechanisms affecting RARβ expression (one *cis* and one *trans*) in both CALU-1 and C-19 cells, then C-59 cells may exhibit the same mechanisms. The reader is asked to recall that C-59 is a CALU-1 derivative similar to C-19, in that it is the endogenous RARβ2 which is turned on, but its RARβ2 expression level is higher. Bisulphite sequencing more C-59 clones may possibly reveal hypermethylation among some sequence clones. This would be interpreted as hypermethylation affecting one allele, and another alteration (possibly in *trans*) affecting the other allele, and would support the hypothesis of two mechanisms.

Finally, another possible reason for the inconsistent results concerning CALU-1 and C-19 which cannot be overlooked, is cross-contamination. In other words, the PCR reactions may have been accidentally contaminated with bisulphite-treated DNA from other samples. However, this is unlikely because the patterns of methylation observed in these sequence clones were not even remotely similar to any other sequence clone.

3. COLON ADENOCARCINOMA CELL LINES

3.1 RARβ-*non*-expressing colon adenocarcinoma cell lines

COLO-201 and COLO-205 were found to be very heavily methylated (91 to 95% of the 22 CpGs analyzed) in each of two clones sequenced per sample. HCT-15 was found to be heavily methylated (68 to 82%), and LS-180 was found to be partially methylated (45%). SW1222 was completely unmethylated (0%). The striking difference between SW1222 and the other *non*-expressing colon adenocarcinoma cell lines, may be factual or spurious. The genomic sequencing protocol should be repeated on an SW1222 sample to confirm these results. Completely unmethylated sequences were previously observed (NCI-H157) and may have cross-contaminated the SW1222 samples.

A striking finding was that CpG #20 was always unmethylated in all hypermethylated and partially methylated colon adenocarcinoma sequence clones. Though a conclusion can not be formulated regarding positional effects of CpG methylation status, one recurring pattern of methylation among colon cancer cell lines may be highlighted. Hypermethylation always included methylation of CpGs #21 and #22 and lack of methylation of CpG #20. That lack of methylation at CpG #20 may be associated with RAR binding, is unlikely, because CpG #21 which is normally methylated in hypermethylated cell lines, is *within* the RARE, whereas CpG #20 is *several bp upstream* of the RARE. However, CpG #20, which is in the center of a *Sma* I site (CCCGGG), may be a strategic position for an as of yet undetermined reason.

Interestingly, the only CpG sites within the P2 promoter (upstream of the transcription start) which are conserved between the mouse and human are CpGs numbers 2, 6, 11, 12, 13, 15, 17, 20, 21, and 22. Conservation across species is normally thought to represent the importance of the function dictated by the sequence. Thus, climbing down the hierarchical rungs of sequence-specific functions regarding the sequence in question, first in importance may be the actual gene, then the retinoic acid response element and TATA Box, and finally, the actual CpG sites. However, one might postulate a different order of importance, depending on which function was first adopted by the prototypical cell during the course of evolution. Was it methylation-associated gene expression? If so, then the conservation of *position-specific* CpG sites (eg. X bp upstream of the RARE) as opposed to sequence-specific CpG sites, may be relevant in itself. Was it the gene, the TATA Box and RARE or RARE-like sequences? If so, then the observed conservation does not necessarily reflect CpG site-specific function(s), but rather the sequence itself, disregarding methylation sites.

3.2 RARβ-expressing colon adenocarcinoma cell line

CACO-2 was found to be hypomethylated (5% of CpGs #1-#22). Each of two clones sequenced presented a single but different methylated CpG: #3 or #5.

3.3 Conclusion 2: Colon adenocarcinoma cell lines

Overall, 6 of 10 RARβ-*non*-expressing colon cancer cell line sequence clones were heavily methylated (68-95%), 2 were moderately methylated (45%) and 2 were not methylated. Of the 3 RARβ-*expressing* colon cancer cell line CACO-2 sequence clones, all 3 were hypomethylated. The hypothesis that promoter methylation is correlated with absence of gene expression is thus supported for CACO-2, COLO-201, COLO-205 and HCT-15. However, methylation analysis should be repeated for SW1222, to confirm the present findings.

The methylation analyses of colon adenocarcinoma cell lines presented less ambiguous results than the lung cancer cell lines. In other words, the degree of CpG methylation was either heavily methylated, moderately methylated, or unmethylated in all clones derived from a same sample. Also, CpG #20 was *always* unmethylated.

4. LUNG TUMOUR DNA SAMPLES

All 6 NSCLC sequence clones, 3 derived from NSC-Auc and 3 derived from NSC-T47 were completely unmethylated (0% of the 21 to 22 CpGs analyzed, respectively). Similarly, the 3 sequence clones derived from SC-Harc were hypomethylated (up to 5% methylation, maximum). Though it is expected that these tumour samples did not express RARβ, at the time of tumour processing by previous members of this lab, RARβ expression was not analyzed, nor was RNA extracted. Thus, several possibilities exist.

The first possibility is that the tumours *did not* express RARβ. This is suggested by the high frequency of occurrence of chromosomal deletions involving the 3p24 locus among SCLCs and squamous cell lung carcinomas, coupled by the fact that most lung

cancer cell lines, established from clinically resected tumours, do not express RARβ (Houle *et al.*, 1991). If this is the case, then the hypothesis that hypomethylation is correlated to expression cannot be supported for these tumour samples. However, if another mechanism, other than chromosomal deletion and hypermethylation, were to be involved in RARβ turn-off, than lack of methylation would not refute the hypothesis.

The second possibility is that the tumours *did* express RARβ and the hypothesis could have been supported if RNA analysis of RARβ2 transcription had been done (such as reverse transcript-PCR, Northern blot or RNAase Protection Analysis). But the samples were not processed for RNA by the lab members at the time of DNA extraction. In the immediate future, tumour samples and samples of bronchioalveolar lavage (BAL) will be accessible and further studies would include the analysis of RARβ expression.

If confirmed, early detection of RARβ hypermethylation in patients at risk for lung cancer, such as smokers, may provide a powerful tool in chemoprevention and anti-cancer therapy. As previously mentioned, a combination of 5-aza-2'-deoxycytidine and RA could theoretically be administered to ensure hypomethylation and induction of RARβ, respectively.

4.1 Conclusion 3: Lung tumour cells

The results of the limited number of tumour samples did not provide support for the hypothesis, but only one sample (NSC-T47) was known to be epidermoid. Unfortunately, information on RARβ expression was lacking. At the time of tissue processing, only DNA was extracted. Nevertheless, the high probability that RARβ expression was turned off in this epidermoid tumour suggests the possibility that the RARβ locus is altered, but not by methylation. Suggestions for further studies include: (a) the bisulphite genomic sequencing methylation analyses of many more tumour samples, combined with RARβ expression analyses, evidently, and (b) the identification of other inactivating mechanisms involved in RARβ gene turn-off, perhaps similar to CALU-1 and NCI-H520. If indeed another mechanism is at play in the turning-off of RARβ, then the observed *lack of hypermethylation* does not go against or refute the hypothesis (hypermethylation at the RARβ P2 promoter is associated with lack of RARβ2 expression).

5. P2 HYPERMETHYLATION IS CORRELATED WITH COMPLETE LACK OF RARβ EXPRESSION

Methylation analyses of the RARβ P2 promoter show a correlation between hypermethylation and lack of RARβ expression in lung cancer and colon adenocarcinoma cell lines (Figure 33). This correlation **was observed** for 11 cell lines: NSCLC cell lines: (1) C-59, (2) NCI-H157, (3) NCI-H520, (4) NCI-H596 ; SCLC cell lines: (5) NCI-H82, (6) MM-1; Colon adenocarcinoma cell lines: (7) CACO-2, (8) COLO-201, (9) COLO-205, (10) HCT-15, (11) LS-180; but **was *not* observed** for 3 cell lines: NSCLC cell lines: (1) CALU-1, (2) C-19; Colon adenocarcinoma cell line: (3) SW1222. Finally, the hypothesis could **not be tested** for lung tumour tissue samples: (1) NSC-Auc, (2) SC-Harc and (3) T47, because though hypomethylation was determined, expression level wasn't.

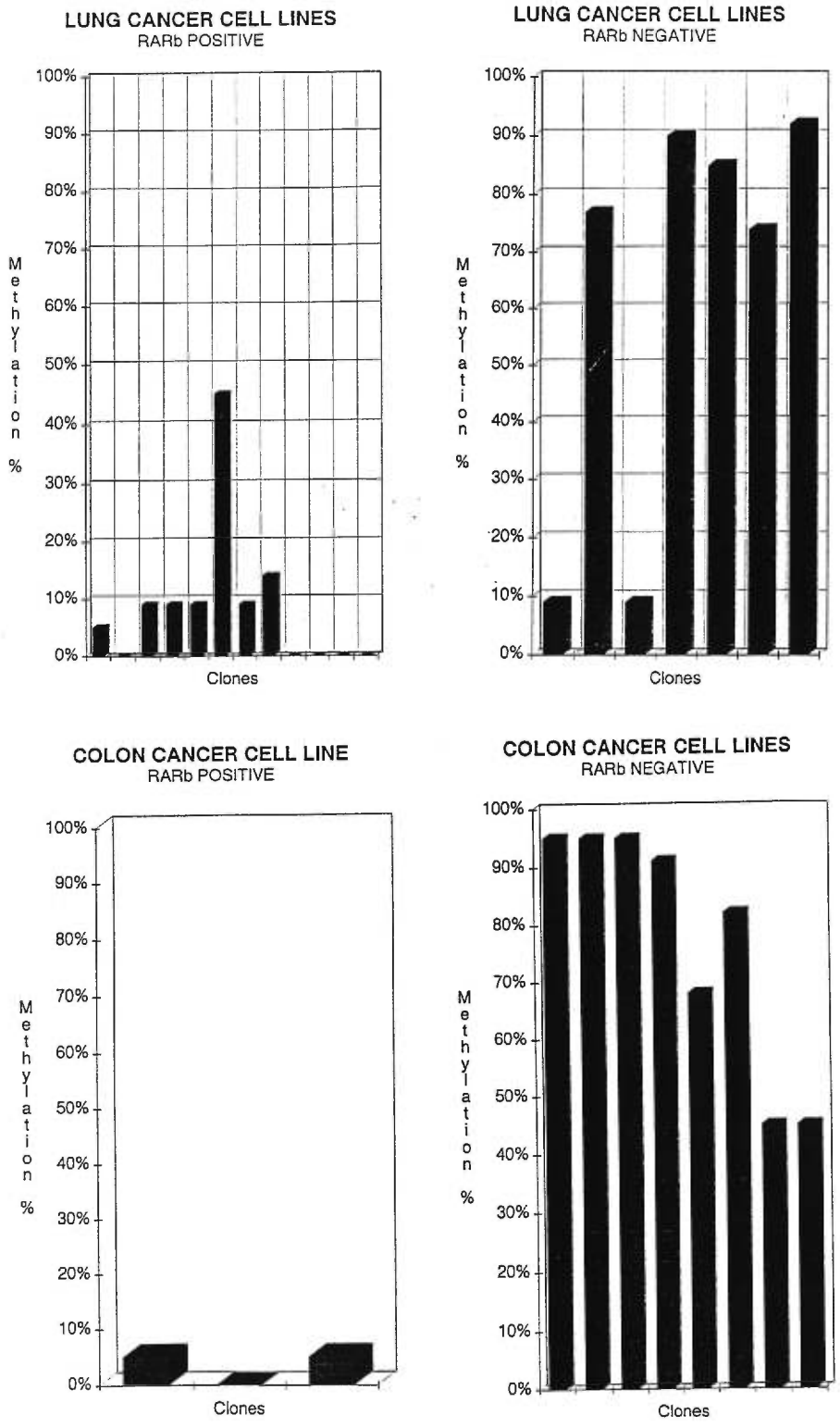

Figure 33: Bar graphs representing degree of CpG methylation within the RARβ P2 promoter (CpGs #1 through 22) of various RARβ-expressing and RARβ-*non*-expressing lung cancer and colon cancer cell line MRED-Improved Bisulphite genomic sequence clones.

The correlation between the lack of RARβ expression and P2 hypermethylation does not reveal whether hypermethylation antedates or post-dates the irreversible genetic events which are characteristic of lung cancer. However, since 5-methylcytosines have increased mutability (Stryer, 1988), and since one third of all mutations observed in human disease involve C → T transitions within CpG dinucleotides (Cooper and Krawczak, 1989), the possibility that hypermethylation is causally-associated with genetic alterations in the carcinogenic process is a likely possibility.

6. METHYLATION PATTERNS

RARβ-*non*-expressing sequence clones exhibited a repetitiveness in the methylation status of *specific* CpG sites. Among 13 hypermethylated sequence clones, individual CpG sites from #1 to #22 were categorized as infrequently methylated (0-50%), frequently methylated (50-99%), or always methylated (100%). The frequency of methylation "f_m" (%) was calculated as the "(number of occurrences of methylation at a specific CpG site)/(number of sequence clones where methylation status at that site was determined) x 100%". Figure 34 charts the frequency of methylation according to CpG site as a bar graph.

The most striking observation was that CpG sites #9 and 20 were *infrequently* methylated (46% and 27%, respectively) and CpG sites #6, 8, 10, and 21 were *always* methylated.

As suggested on several occasions thus far, the significance of the frequency of methylation at a given CpG site may be crucial in understanding the role of methylation in the inactivation of RARβ. It is possible that the *location* of certain CpG sites, such as those which were determined to be methylated at a frequency of 100%, is the *determining factor* for methylation-associated gene repression. Of these CpG sites, for instance, CpG #21 is located *within* the RARE. Therefore, it is plausible to hypothesize that due to methylation at this site, retinoic acid receptor (RAR) binding may be inhibited. Since retinoic acid (RA) is known to induce RAR expression, methylation at CpG #21 may reflect lack of inducibility. As for the role of methylation at CpG sites #6, #8 and #18, several possibilities exist. For one, methylation may block transcription by affecting transcription factor binding. Though the TATA Box is distal to these sites, DNA folding

may result in the juxtaposition of a CpG site with the TATA Box. Another possibility is that these CpG sites are located within transcription factor binding sequences which have not yet been recognized as such.

Finally, when methylation was low (i.e. 5%, 9% or 14%), the CpG sites which were methylated were consistently within CpGs #3 and #14, approximately 175 bp 5' of the region where the AP-1 site, the RARE and the TATA Box are located. When there was hypermethylation, CpG #21 (within the RARE) and #22 (near the TATA Box) were *always* methylated, and CpG #20 (7bp upstream of the RARE) was *never* methylated (except for NCI-H596). These patterns of methylation lead us to hypothesize that methylation of CpGs #21 and #22 (within the RARE and immediately upstream of the TATA Box, respectively) is correlated with lack of expression. Downstream methylation analyses of RARβ-expressing cell lines NCI-H82, NCI-H157 and CACO-2 showed 0% methylation of CpGs #20 through #38, further supporting this hypothesis. The fact that a low degree of methylation does not appear to affect these CpGs, but CpGs which are further upstream, further supports this suggestion.

The lack of methylation at a potentially crucial CpG site, such as CpG #20, may be causal and/or consequential. It may be *causal*, if for instance the spread of methylation is in the 5' to 3' direction, but blocked at these sites via a mechanism as of yet to be discovered. Possibly, this mechanism involves different sequence-specific factors capable of blocking methylation at key regulatory sites and recognizing methylation status at these key regulatory sites. It may be *consequential* because RARβ-expressing cell lines are expected to have nuclear receptors bound at the RARE, therefore these receptor-ligand complexes may cause MeTase access to be sterically hindered in and around these sites. This is an example of a structural change affecting the DNA proper. However, access may denied through chromatin changes as well. It has been shown that RA induces chromatin structural changes in the P2 promoter (Bhattacharyya *et al.*, 1997). Thus, the precise structural changes may inhibit methylation at these two sites specifically. What is striking then, is the fact that 5 of the 7 CALU-1 and CALU-1 transfectant clones (C-19 and C-59), whether they are hypo- or hypermethylated, are always methylated at CpG #21 or #22, or both. This supports the theory of two mechanisms, with hypermethylation or methylation at crucial CpGs comprising one of the two inactivating mechanisms. Thus, methylation at one or both of these CpG sites may be correlated with lack of expression.

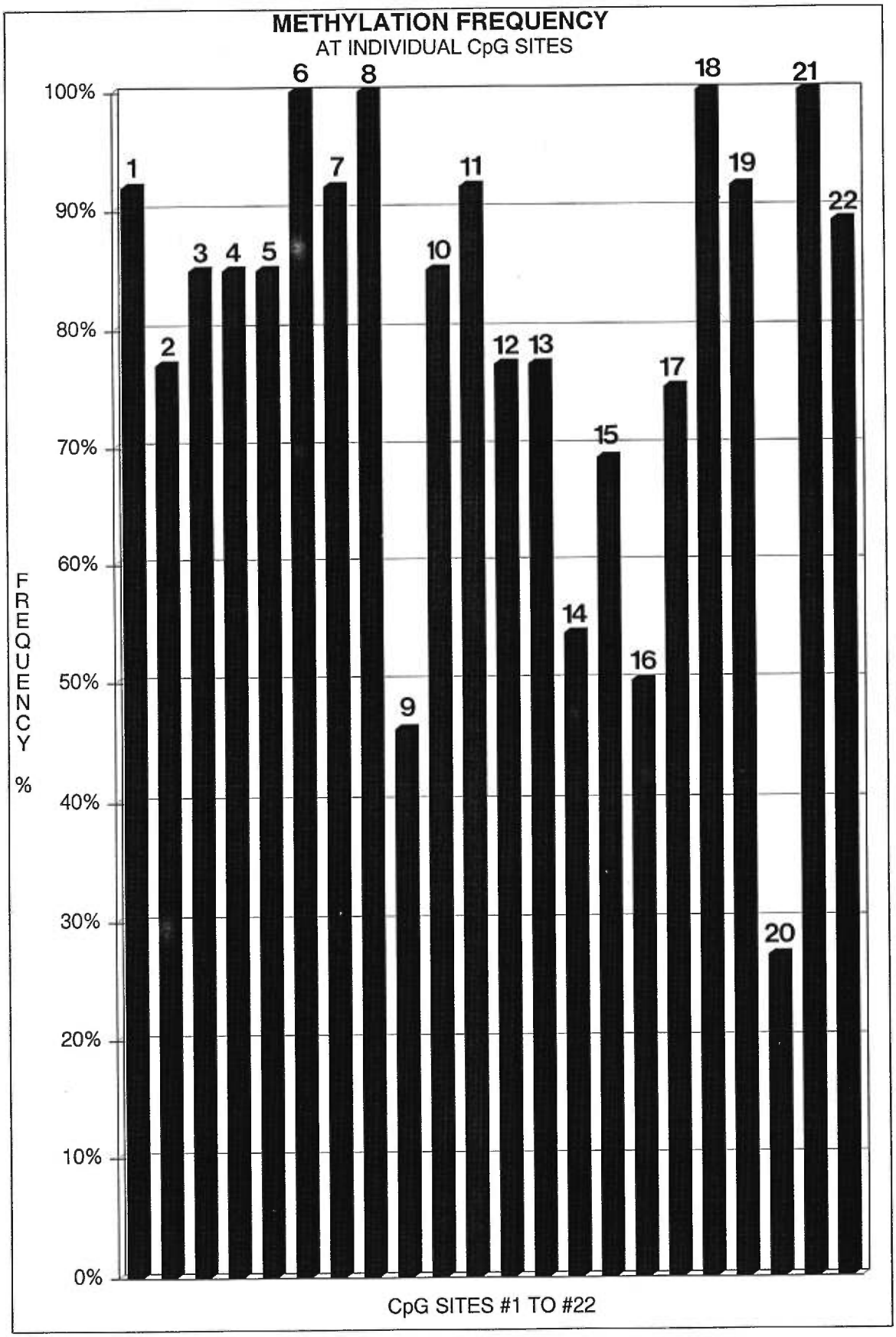

Figure 34: Bar graph representing the frequency of methylation at each of the 22 CpG sites within the P2 promoter 5' of the transcription start among 14 different RARβ-*non*-expressing hypermethylated sequence clones. Note that CpG #20 is 7 bp upstream of the RARE, CpG #21 is within the RARE and CpG #22 is near the TATA Box.

7. SHORT-TERM PERSPECTIVES

A priority for establishing the biological relevance of these results is to assay more lung tumour samples - processed both for DNA and RNA - for *RAR*β promoter 2 methylation and RARβ2 expression, respectively. It is relevant to mention that the cell lines which presented very high degrees of methylation (i.e. approximately 90%), such as NCI-H596, COLO-201 and COLO-205, are also the cell lines which have very tight turn-off of expression. It would be interesting to determine whether increasing levels of methylation of the RARβ P2 promoter, in certain cell lines, is correlated with decreasing levels of expression. This question could be answered by transfecting exogenously methylated P2 promoter (at different levels of methylation) with a reporter gene into RARβ2-*non*-expressing cells, and subsequently determining RARβ2 expression.

To ensure that endogenous methylation does not take place, the P2 region of the construct could be PCR-synthesized with some 5-aza-2'-deoxycytidine in the dNTP mix. Though 5-aza-2'-deoxycytosine bases would be incorporated at random, their positions could be determined later. After exogenous methylation of the expression vector using Sss I methylase (a CG methylase) for example, the position of unmethylated CpGs could be determined using the bisulphite genomic sequencing protocol.

To ensure that demethylation does not take place is a little trickier. The existence of a DNA demethylase is highly controversial. However, CpG-specific demethylase activity has recently been isolated in the nuclear extract of a human lung cancer cell line, A549 (personal communication, Dr. Moshe Szyf). Until demethylase inhibitors are available, the best way to circumvent eventual demethylation would be to analyze *transient* transfectants.

Finally, treating RARβ-expressing cell lines with methylation inhibitors such as 5-aza-2'-deoxycytidine (eg. Côté and Momparler, 1997) or methyl transferase antisense oligodeoxynucleotides (Szyf, 1996), are experimental designs presently being used to determine whether or not inhibiting methylation can reverse the inhibition of transcription. Indeed they do. The RARβ-*non*-expressing cell lines determined to exhibit hypermethylation of the RARβ P2 promoter in this project could be subjected to this type of study.

Further studies including (1) the role of specific CpG-methylation in gene inactivation, and (2) the threshold or density of CpG-methylation required for gene inactivation, are necessary to determine the general requirements of methylation-specific gene turn-off. Such studies may be approached in two ways: (1) correlation studies between CpG-cytosine-specific methylation and level of expression; and (2) transfectant cell line studies using artificial expression vector constructs with known patterns of CpG methylation.

8. LONG TERM PERSPECTIVES

Hypermethylation has been shown to be correlated with lack of *RARβ expression* in several cancer cell lines in this project. The most clinically-relevant aspect of DNA hypermethylation is that through its reversibility or inhibition, it may be targeted in chemoprevention and anticancer therapy using gene therapy and/or pharmacological compounds. First, methylation status may eventually become an epigenetic prognostic of tumour grade and/or suitability for hypomethylation therapy. Second, pharmacologically inhibiting hypermethylation in patients at risk for lung cancer may constitute a possible chemopreventive strategy. Third, combining this with therapeutic RA administration, which represents a non-toxic method of further increasing RARβ expression, may prove to be an effective cancer chemopreventive and/or anticancer therapeutic strategy. A study combining the sequential treatments of 5-aza-2'-deoxycytidine and *all trans* retinoic acid (*at*RA) in DLD-1 colon carcinoma cells *in vitro* was shown to have synergistic effects on reducing their proliferative potential (Côté and Momparler, 1997).

Methyl transferase antisense oligodeoxynucleotides, 5-azacytidine and 5-aza-2'-deoxycytidine have been mentioned on several occasions in this document. However, they have important disadvantages. Firstly, 5-azacytidine and 5-aza-2'-deoxycytidine have been shown to be carcinogenic in numerous studies (eg. Carr *et al*, 1984; Carr *et al.*, 1988, respectively), Secondly, their hypomethylating actions have not yet been biotechnologically engineered to be gene-specific, and therein lies the main problem, since hypomethylation may turn on oncogene expression. In contrast with hypermethylation, endogenous hypomethylation appears to affect DNA *globally* (Feinberg *et al.,* 1988). The use of exogenous global hypomethylating drugs or gene therapy approaches are potentially carcinogenic in themselves. This paradox is reminiscent of radiation therapy and chemotherapy in the treatment of cancer, since these

approaches are mutagenic and affect DNA globally also. Clinical perspectives concerning hypermethylation comprise the fine-tuning of potential DNA demethylating agents, so that their actions be specific for pre-cancerous or cancerous cells versus normal cells, and/or repressed housekeeping genes such as tumour suppressor genes versus other genes, such as proto-oncogenes.

If hypermethylation is found to be an important tumour promoter in lung cancer, and cigarette smoking if found to be positively correlated with hypermethylation, then the avenues of research this opens up are abundant. So are some of the social ramifications. First, are other lung cancer risk factors, such as exposure to asbestos, causally-associated with pre-neoplastic bronchial epithelial DNA hypermethylation in otherwise healthy non-smokers? If so, what are the immediate physiological mediators (if there are any) which respond to these factors by stimulating methylation? Does the cell have an early protective methylation-associated advantage which then becomes deleterious (as we've seen in the case of RARβ) in the long term? Second, what is/are the common factors and the mechanism(s) by which these factors cause hypermethylation? Can this/these factors be removed from cigarette material? Third, do protective factors, such as vitamin A, delay hypermethylation somehow? For instance, does increased binding of RAREs, due to increased absorption of nuclear receptor cognate ligands (such as retinoids) sterically hinder methylation by inhibiting access to the methyl transferase enzyme? In fact, this might be the case since RA has been shown to induce chromatin changes through increased DNase I hypersensitivity in the P2 promoter (Bhattachatyya *et al.*, 1997).

9. COLON CANCER CELL LINES: AN *ASIDE* PURSUED

The original subject of this project was lung cancer. Thus, lung cancer cell lines and tumour cells were analyzed for methylation status. Later, colon cancer cell lines were introduced as an *aside subject*. However, interesting results were obtained and the subject was pursued. The results provide information on the association between RARβ P2 promoter hypermethylation and lack of RARβ2 expression in cells *other* than lung cancer cells. Moreover, they also supply a nice counterpoint in that hypermethylation was observed in *all* colon cancer cell lines analyzed, unlike lung cancer cell lines. Therefore, it appears that no other (as of yet determined) inactivating mechanism is involved in the loss of expression of RARβ in colon cancer cell lines, whereas other inactivating mechanisms may be involved in lung cancer cell lines (eg. a *trans*-acting defect).

B. *MRED-IMPROVED BISULPHITE GENOMIC SEQUENCING PROTOCOL*

The *bisulphite genomic sequencing protocol* (Clark *et al.*, 1994) is a technique which is capable of detecting *all* individual methylated cytosines (CpG-cytosines, or other) within a target sequence of choice. The bisulphite conversion reaction, at optimal conditions, converts *un*methylated cytosines to uracils, but does not affect 5-methylcytosines. Upon sequencing, *un*methylated cytosines are read as thymines, whereas 5-methylcytosines remain as cytosines. The bisulphite conversion reaction is *highly single strand-specific*. Lack of single strandedness leads to blocks of unreacted DNA. Thus, the pool of bisulphite-treated DNA may contain multiple molecular species of various degrees of conversion. Also, the harsh reaction conditions cause DNA degradation.

1. OBJECTIVE 2

The second objective of this project was to improve the *Bisulphite Genomic Sequencing Protocol* by increasing the amount of recoverable bisulphite-modified DNA (Objective 2.1) and increasing the conversion efficiency to the established threshold of confidence of 98.6%, or greater (Objective 2.2).

Objectives 2.1 and 2.2 were attained by modifying the original protocol in the following ways:

2.1
(1) Increasing the amount of DNA to 5 to 10μg;
(2) Deactivating and precipitating the linearization reaction;
(3) Using a new batch of sodium bisulphite (SIGMA);
(4) Using Wizard Magic Miniprep™ Resin (Promega);

2.2
(5) Denaturing the DNA at 95°C for 5 min. (prior to the conversion reaction);
(6) Digesting the PCR-amplified products with multiple restriction enzymes specific for unconverted DNA (*MRED*).

2. *MULTIPLE RESTRICTION ENZYME DIGEST (MRED)*

Since it was determined that there may be *some*, albeit few, fully-converted molecular species (regarding non-CpG-cytosines) within the bisulphite-treated DNA molecular population, a method needed to be devised whereby *those* molecular species could be preferentially selected.

The strategy which consistently resulted in 98.6% to 100% conversion consisted in digesting the PCR product with several restriction enzymes that would differentially digest unmodified DNA, without selecting methylated or unmethylated CpG sites. The strategy devised was named the *Multiple Restriction Enzyme Digest*, or "*MRED*", of the semi-nested PCR product (for a detailed description, see pp.70-74).

In case only certain regions of the DNA fragment were unconverted, the chosen restriction sites were more or less equally distributed along the target sequence. In order to ensure that the restriction enzymes did not have a selective advantage for a particular methylation status, restriction enzyme sites were very carefully chosen. Briefly, four main criteria served as the guidelines for choosing the enzymes.

The restriction site was:

(1) present in *un*converted DNA, but *not* in *converted* DNA;
(2) did not contain a CpG site unless it also contained a non-CpG-cytosine;
(3) did not contain a unique cytosine which was also at the very 3' end if it was followed immediately upstream by a guanine;
(4) did not contain a guanine at the very 5' end if it was immediately preceded by a cytosine in the upstream sequence, unless it also contained a non-CpG-cytosine.

3. CONCLUSION

Using the *MRED-Improved Bisulphite Genomic Sequencing Protocol*, developed in this project, resulted in sequence clones with 98.6% or 100% conversion efficiency, almost nine times out of ten. This is *ninefold more* than the

frequency of clones with at least 98.6% conversion efficiency obtained *without MRED* (Figure 35).

Thus, the *MRED-Improved Bisulphite Genomic Sequencing Protocol* is a highly efficient and reproducible method for the methylation analysis of *all* cytosines within a target sequence.

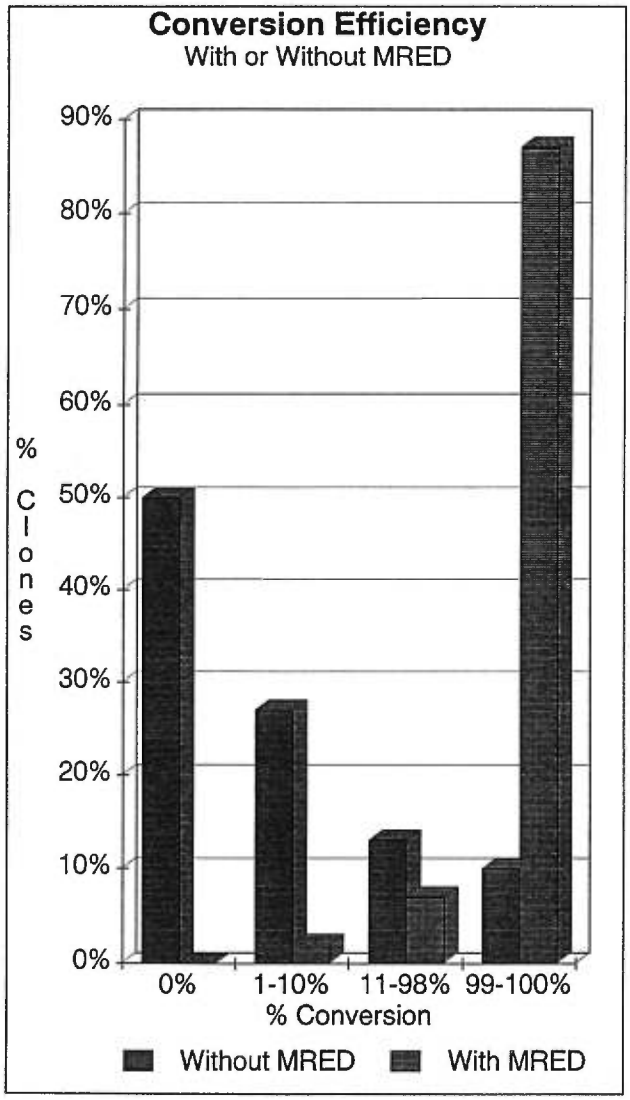

Figure 35: Bar graph charting proportion of sequencing clones (%) versus efficiency of conversion (%) with or without MRED. Conversion efficiency = "number of converted non-CpG-cytosines/number of convertible (non-CpG) cytosines x 100%". **MRED** = Multiple Restriction Enzyme Digest.

C. SEQUENCING DIFFERENCES

There were numerous differences between the observed HOU-1/HOU-3 sequencing results and the expected RARβ sequence. Differences were considered to represent (1) actual differences present in the genomic DNA (such as polymorphisms, point mutations, insertions, and the like), or (2) *Taq* polymerase extension errors, due to base pair mismatching. The high proportion of thymines within an efficiently converted sequence may contribute significantly to such errors by altering the sequences' secondary structure and/or by increasing the number of thymines in tandem and causing slippage. Thymine content within the *un*converted sequence is approximately 24%; thymine content within the *fully converted* sequence, excluding CpG-cytosines, is 38%.

To begin with, differences were categorized according to their *type* (eg. A→G transition-like difference versus T→G transversion-like difference) (Table XX, p.93). This was done in order to determine whether or not particular types of differences (mutations?) may predominate among the samples analyzed (disregarding their position). Later, differences were categorized according to their *position* (eg. position 75 versus position 252) (Table XXI, p.94). This was done to determine whether a particular position was a hot-spot for mutations or *Taq* polymerase errors (of the same or differing type).

It is unfortunate that C→T transitions within CpG sites (i.e. CpG→TpG) cannot be determined through bisulphite sequencing results since conversion is taken to mean lack of methylation.

The most striking finding was that <u>all</u> clones analyzed (100% of 40 clones) had two identical position-differences: (1) 2 extra Gs, *circa* position 75; and (2) G→A. At position 252 (positions refer to HOU-1/HOU-3 539bp PCR product, Figure 23, p.82). At first, the fact that the extra Gs were located near or within a string of four Gs (expected sequence) lead us to think that they were artifactual (due to *Taq* polymerase slippage). However, the fact that they are present in all 40 sequencings highly suggests that they are factual. Their ubiquitousness among <u>all</u> the samples analyzed, suggests that the expected sequence used to compare obtained sequencing results, (Shen *et al.*, 1991) contains two corresponding errors.

Excluding the above-mentioned differences, the number of differences per sequencing ranged from 0 to 4, with an average mean of 0.7 difference per sequence (approximately 540bp/sequencing). Since the *Taq* polymerase enzyme makes an average of 1 mismatch per 500 to 1000 bp (Kornberg and Baker, 1992), 1 to 2 differences per sequence clone were considered to be such mistakes. Moreover, 31 of the 35 position-differences (excluding the two which were present in all clones) were unique *vis-à-vis* sequence clones. In other words, they occurred not more than once among all 40 clones.

Of the other 4 occurrences, 2 position-differences occurred twice, in each case, in two different cell-lines. 2 of these differences consisted of T→C at positions 453 and 455, respectively, but were either sequencing ambiguities or they were within a region of high ambiguity. The other 2 position-differences were a missing T at position 214 and a T→G at the identical position, within a string of 10 Ts (converted sequence). Thus, these differences were attributed to *Taq* polymerase errors also.

It is important to mention that 100% of the subcloned PCR inserts (approximately 112 total) were inserted in the *same* orientation. Contamination was ruled out because all 112 sequence clones consisted of different inserts. Firstly, four different oligo sets were used to PCR amplify bisulphite-treated DNA. Secondly, approximately 20 different methylation patterns were observed. Thirdly, even similarly methylated sequence clones differed regarding efficiency of conversion and/or position-differences. Thus, it was concluded that something about the structure of high thymine content inserts influenced the orientation of insertion. Clark has reported a similar finding, in that she found that the use of pBluescript vectors, such as pBS KS (the vector used in this project) provided a selective advantage for certain molecular species from the pool of PCR-amplified bisulphite-modified DNA. More precisely, her group noted that partially converted PCR products were preferentially ligated to pBluescript vectors, over fully converted products (Clark *et al.*, 1994). The *MRED* technique, developed in this project, circumvents this problem. Moreover, a single reaction volume of bisulphite-treated DNA, as per the *MRED-Improved Bisulphite Genomic Sequencing Protocol*, can be used for the methylation analyses of many genes, even difficult genes such as RARβ.

Finally, though DNA methylation generally involves CpG sites in mammalian DNA (Gruenbaum *et al.*, 1981), non-CpG-cytosines have been reported to be methylated also (Clark *et al.*, 1995; Toth *et al.*, 1990). Thus, among the sequence clones with 98.6%

conversion efficiency, the *non-CpG-cytosine* which was *not* converted may be considered to be methylated.

Six different sequence clones obtained from five different cell lines presented an unconverted non-CpG-cytosine per sequence.

Pos.	Site	Cell line	S.C.	Meth.
200	CpC	COLO-205	099	H
208	CpT	CACO-2	087	h
210	CpC	SW1222	102	h
211	CpC	C-19	088	h
293	CpT	HCT-15	101	H
456	CpA	COLO-205	098	H

Table XXII: Putative positions of observed non-CpG-cytosine methylation.
S.C. = sequence clone; **Meth.**= methylation status; **H** = hyper-methylated; **h** = hypo-methylated.

The locations of five of these putative non-CpG methylcytosines do not coincide with any known *cis*-acting elements, nor do they show any association with hyper- or hypomethylation. However, one coincides with the transcription start [CpA at position 456, (see Table XXII)]. Furthermore, a similarity which may be noted is that four of the non-CpG methylcytosines occurred within a 12 bp sequence [CCCGGCTCCTCC, positions 200 - 211, (Figure 23, p.82)]. Finally, one of them occurred at a CpNpG site (position 293). The CpNpG site has been previously reported to be methylated in mammalian cells (Clark *et al.*, 1995). Thus, it is possible that non-CpG methylation may play a role at specific sequences.

V. REFERENCES

Aaronson, S.A. *Growth factors and cancer.* Science 1991; **254,** 1146-1153.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. *Molecular Biology of the Cell*, 3rd ed., New York, 1994.

Allam, M., Bertrand, R., Zhang-Sun, G., Pappas, J. and Viallet, J. *Cholera toxin triggers apoptosis in human lung cancer cell lines*. Cancer Res. 1997; **57**, 2615-2618.

Antequera, F. and Bird, A. *Number of CpG islands in human and mouse*. Proc. Natl. Acad. Sci. USA. 1993; **90**, 11995-11999.

Antequera, F., Boyes, J. and Bird, A. *High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines*. Cell.1990; **62**, 503-514.

Attadia, V. *Effects of 5-aza-2'-deoxycytidine on differentiation and oncogene expression in the human monoblastic leukemia cell line U-937.* Leukemia. 1993; **7**(Supp.1), 9-16.

Begg, C.B., Zang, Z.F., Sun, M. Herr, H.W. and Schantz, S.P. *Methodology for evaluating incidence of second primary cancers with application to smoking-related from S.E.E.R.* Am. J. Epidemiol. 1995; **142**, 653-665.

Bérard, J., Laboune, F., Mukuna, M., Massé, S., Kothary, R. and Bradley, W.E.C. *Lung tumors in mice expressing an antisense RARβ2 transgene*. FASEB J. 1996; **10**, 1091-1097.

Bestor, T.H. *DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes.* Phil. Trans. Res. Soc. (London). 1990; **B326**, 179-187.

Bhattacharyya, N., Dey, A., Minucci, S., Zimmer, A., John, S., Hager, G. and Ozato, K. *Retinoid-induced chromatin structure alterations in the retinoic acid receptor β2 promoter*. Molec. Cell. Biol. 1997; **17**, 6481-6490.

Bird, A. *The essentials of DNA methylation*. Cell. 1992; **70**, 5-8.

Birrer, M.J., Alani, R., Cuttitta, F., Preis, L.H., Sabich, A.L., Sanders, D.A., Siegfried, J.M., Szabo, E., Brown, P.H. *Early events in the neoplastic transformation of respiratory epithelium*. J. Natl. Cancer Inst. 1992; **13**, 31- 37.

Bos, J.L. *Ras oncogenes in human cancer: A review*. Cancer Res. 1989; **49**, 4682-4689.

Boyes, J. and Bird, A. *DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein*. Cell. 1991; **64**, 1123-1134.

Boyes, J. and Bird, A. *Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein.* EMBO J. 1992; **11**, 327-333.

Campoy, F.J., Meehan, R.R., McKay, S., Nixon, J. and Bird, A. *Binding of histone H1 to DNA is indifferent to methylation at CpG sequences.* J. Biol. Chem. 1995; **270**, 26473-26481.

Canadian Statistics on Cancer 1997. Natl. Cancer Inst. Toronto, Canada. 1997.

Canadian Statistics on Cancer 1991. Natl. Cancer Inst. Toronto, Canada. 1991.

Carney, D.N. *Lung Cancer Biology.* Eur. J. Cancer.1991; **27**, 366-369.

Carney, D.N. and Deleij, L. *Lung cancer biology.* Semin. Oncol. 1988; **15**, 199-214.

Carr, B.I., Reilly, J.G., Smith, S.S., Winberg, C. and Riggs, A. *The tumorigenicity of 5-azacytidine in the male Fischer rat.* Carcinogenesis. 1984; **5**, 1583-1590é

Castonguay, A., Foiles, P.G., Trushin, N. and Hecht, S.S. *Study of DNA methylation by tobacco-specific N-nitrosamines.* Environ. Health Perspect. 1985; **62**, 197-202.

Chambon, P. *The retinoid signaling pathway: molecular and genetic analyses.* Cell Biol. 1994; **5**, 115-125.

Chambon, P. *A decade of molecular biology of retinoic acid receptors.* FASEB J. 1996; **10**, 940-954.

Chopra, D.P. and Wilkoff, L.J. *Reversal by vitamin A analogues (retinoids) of hyperplasia induced by N'-methyl-N'-nitro-N-nitroso-guanidine in mouse prostate organ cultures.* J. Natl. Cancer Inst.1977; **58**, 923-930.

Christman, J.K., Mendelsohn, N., Herzog, D. and Schneiderman, N. *Effect of 5-azacytidine on differentiation and DNA methylation in human promyelocytic leukemia cells (HL-60).* Cancer Res. 1983; **43**, 763-769.

Chung, C.T., and Miller, R.H. *A rapid and convenient method for the preparation and storage of competent bacterial cells.* Nucl. Acids Res. 1988; **16**, 3580.

Chung, G.T.Y., Sundaresan, V., Hasleton, P., Rudd, R., Taylor, R., Rabbitts, P.H. *Sequential molecular genetic changes in lung cancer development.* Oncogene. 1995; **11**, 2591-2598.

Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. *High sensitivity mapping of methylated cytosines.* Nucl. Acids Res. 1994; **22**, 2990-2997.

Clark, S.J., Harrison, J.and Frommer, M. *CpNpG methylation in mammalian cells*. Nat. Genet. 1995; **10**, 20-27.

Comstock, G.W., Alberg, A.J., Huang, H.Y., Wu, K., Burke, A.E., Hoffman, S.C., Norkus, E.P., Gross, M., Cutler, R.G., Morris, J.S., Spate, V.L. and Helzlsouer, K.J. *The risk of developing lung cancer associated with antioxidants in the blood: ascorbic acid, carotenoids, alpha-tocopherol, selenium, and total peroxyl radical absorbing capacity*. Cancer Epidemiol. Biomarkers. 1997; **6**, 907-916.

Cooper, D.N. and Krawczuk, M. Cytosine *methylation and the fate of CpG dinucleotides in vertebrate genomes*. Hum. Genet. 1989; **83**, 181-188.

Côté, S. and Momparler, R. *Activation of the retinoic acid receptor β gene by 5-aza-2'-deoxycytidine in human DLD-1 colon carcinoma cells*. Anti-cancer Drugs. 1997; **8**, 56-61.

Counts, J.L. and Goodman, J.I. *Alterations in DNA methylation may play a variety of roles in carcinogenesis*. Cell. 1995; **83**, 13-15.

Cross, S.H., Meehan, R.R., Nan, X. and Bird, A. *A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins*. Nature Genetics. 1997; **16**, 256-259.

deBustros, A., Nelkin, B.D., Silverman, A. Ehrlich, G., Poiesz, B. and Baylin, S.B. *The short arm of chromosome 17 is a "hot-spot" for hypermethylation in human neoplasia*. Proc. Natl. Acad. Sci. USA. 1988; **85**, 5693-5697.

DeLaat, C.A. and Lampkin, B.C. *Long term survivors of childhood cancer: evaluation and identification of sequelae of treatment*. CA Cancer J. Clin.1992; **42**, 263.

DeLuca, L.M. *Retinoids and their receptors in differentiation, embryogenesis and neoplasia*. FASEB J. 1991; **5**, 2924-2933.

Dobrovic, A., Houle, B., Belouchi, A. and Bradley, W.E.C. *ErbA-related sequence coding for DNA-binding hormone receptor localized to chromosome 3p21-3p25 and deleted in small cell lung carcinoma*. Cancer Res. 1988; **48**, 682-685.

Dwyer, T., Blizzard, L., Shugg, D., Hill, D. and Ansari, M.Z. *Higher lung cancer rates in young women than young men: Tasmania, 1983-1992*. Cancer Causes Control. 1994; **5**, 351-358.

Eguchi, K., Kanai, Y., Kobayashi, K. and Hirohashi, S. *DNA hypermethylation at the D17S5 locus in non-small cell lung cancers: Its association with smoking history*. Cancer Res. 1997; **57**, 4913-4915.

Eliason, J.F., Kaufmann, F., Tanaka, T. and Tsugaguchi, T. *Anti-proliferative effects of arotinoid Ro 40-8757 on human cancer cell lines in vitro*. Br. J. Cancer. 1993; **67**, 1293-1298.

Fazely, F., Ledinko, N. and Smith, D.J. *Inhibition by retinoids of in vitro invasive ability of human lung carcinoma cells*. Anticancer Res. 1988; **8**, 1387-1391.

Feil, R., Charlton, J., Bird, A.P., Walter, J. and Reik, W. *Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing*. Nucl. Acids Res. 1994; **22**, 695-696.

Feinberg, A.P., Gehrke, C.W., Kuo, K.C. and Ehrlich, M. *Reduced genomic 5-methylcytosine content in human colonic neoplasia*. Cancer. 1988; **48**, 1159-1161.

Frank, A.L. *Epidemiology of lung cancer*. Cancer of the lung. Thoracic Oncology. (Roth, J.A., Ruckdeschel, J.C. and Weisenburger, T.H. eds.) Harcourt Brace Jovanovitch Inc. Philadelphia. 1989; 6-15.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, D.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. *A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands*. Proc. Natl. Acad. Sci. USA. 1992; **89**, 1827-1831.

Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gerke, C.W. and Ehrlich, M. *The 5-methylcytosine content of DNA from human tumours*. Nucl. Acids Res. 1983; **11**, 6883-6894.

Gambari, R., del Senno, L., Barbieri, R., Viola, L., Tripodi, M., Raschella, G. and Fantoni, A. *Human leukemia K-562 cells: induction of erythroid differentiation by 5-azacytidine*. Cell Diff. 1984; **14**, 87-97.

Gerbert, J.F., Moghal, N., Frangioni, J.V., Sugarbaker, D.J. and Neel, B.G. *High frequency of retinoic acid receptor* β *in human epidermoid lung cancer*. Oncogene. 1991; **6**, 1859-1968.

Giguère, V., Ong, E.S., Segui, P. and Evans, R.M. *Identification of a receptor for the morphogen retinoic acid*. Nature (London) 1987; **330**, 624-629.

Ginsberg, R.J., Vokes, E.E. and Raben, A. *Non-small cell lung cancer*. Cancer of the lung. Cancer principles & Practice of Oncology, 5th ed., 1997; 858-867.

Gonzalgo, M.L., Liang, G., Spruck, C.H., 3rd, Zingg, J.M., Rideout, W.M., 3rd and Jones, P.A. *Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily-primed PCR*. Cancer Res. 1997; **57**, 594-599.

Gould, V.E. and Warren, W.H. *Epithelial neoplasms of the lung*. Thoracic Oncology. (Roth, J.A., Ruckdeschel, J.C. and Weisenburger, T.H. eds.) Harcourt Brace Jovanovitch Inc. Philadelphia. 1989; 77-93.

Graham, S. *Epidemiology of retinoids and cancer*. J. Natl. Cancer Inst. 1984; **73**, 1423-1428.

Grigg, G. and Clark, S. *Sequencing 5-methylcytosine residues in genomic DNA*. BioEssays. 1994; **16**, 431-436.

Grigg, G. *Sequencing 5-methylcytosines by the bisulphite method*. J. Sequ. Mapping. 1996; **6**, 189-198.

Gruenbaum, Y., Stein, R., Cedar, H. and Razin, A. *Methylation of CpG sequences in eukaryotic DNA*. FEBS Lett. **124**, 67-71.

Hajj, C., Akoum, R., Bradley, W.E.C., Paquin, F. and Ayoub, J. *DNA alterations at proto-oncogene loci and their clinical significance in operable non-small cell lung cancer*. Cancer. 1990; 733-739.

Handelman, G.J., Packer, L. and Cross, C.E. *Destruction of tocopherols, carotenoids, and retinol in human plasma by cigarette smoke*. Am. J. Clin. Nutr. 1996; **63**, 559-565.

Harris, C.C., Silverman, T., Smith, J.M., Jackson, F. and Boren, H.G. *Proliferation of tracheal epithelial cells in normal and vitamin A deficient Syrian golden hamsters*. J. Natl. Cancer Inst.1973; **51**, 1059-1062.

Hayatsu, H., Wataya, Y., Kai, K. and Iida, S. *Reaction of sodium bisulphite with uracil, cytosine and their derivatives*. Biochemistry. 1970; **9**, 2858-2866.

Henderson, B.E., Ross, R.K. Pike, M.C. *Toward the primary prevention of cancer*. Science 1991; **254**, 1131-1137.

Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., Linehan, W.M. and Baylin, S.B. *Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma*. Proc. Natl. Acad. Sci. USA. 1994; **91**, 9700-9704.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D. and Baylin, S.J. *Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands*. Proc. Natl. Acad. Sci. USA. 1996; **93**, 9821-9826.

Hoffmann, D., Melikian, A., Adams, J.D. New aspects of tobacco carcinogenesis. Mass, M.J., Kaufman, D.G., Siegfried, J.M. (eds). *Carcinogenesis*. Compr. Surv. 1985; **8**, 239-256.

Hong, W.K. and Hri, L.M. *Retinoids and human cancer*. The Retinoids: Biology, Chemistry and Medicine. 2nd ed. Raven Press, New York. 1994; 597-658.

Houle, B., Leduc, F. and Bradley, W.E.C. *Implication of RAR*β *in epidermoid (squamous) lung cancer*. Genes Chrom. 1991; **3**, 358-366.

Houle, B. *Étude des gènes THR*β *et RAR*β *dans les cellules de cancer pulmonaire. Ph.D. thesis*. Université de Montréal, Montréal, Canada. 1992.

Houle, B., Rochette-Egly, C. and Bradley, W.E.C. *Tumor-suppressive effect of the retinoic acid receptor* β *in human epidermoid lung cancer cells*. Proc. Natl. Acad. Sci. USA.1993; **90**, 985-989.

Houle, B., Pelletier, M., Wu, J., Goodyer, C. and Bradley, W.E.C. *Fetal isoform of human retinoic acid receptor* β *expressed in small cell lung cancer lines*. Cancer Res. 1994; **54**, 365-369.

Hsieh, C.-L. *Dependence of transcriptional repression on CpG methylation density*. Molec. Cell. Biol. 1994; **14**, 5487-5494.

Hu, L., Crowe, D.L., Rheinwald, J.G., Chambon, P. and Gudas, L.J. *Abnormal expression of retinoic acid receptors and keratin 19 by human oral and epidermal squamous cell carcinoma cell lines*. Cancer Res. 1991; **51**, 3972-3981.

Jackson-Grusby, L., Laird, P.W., Magge, S.N., Moeller, B.J. and Jaenish, R. *Mutagenicity of 5-aza-2'deoxycytidine is mediated by the mammalian DNA methyltransferase*. Proc. Natl. Acad. Sci. USA. 1997; **94**, 4681-4685.

Jehn, U. *New drugs in the treatment of acute and chronic leukaemia: current role of mAMSA*. Bone Marrow Transplant. 1989; **4**(Supp.3), 53-58.

Jones, P.A. *Altering gene expression with 5-azacytidine*. Cell. 1985; **40**, 485-486.

Kastner, P., Mark, M. and Chambon, P. *Nonsteroid nuclear receptors: what are genetic studies telling us about their role in real life?* Cell. 1995; **83**, 859-869.

Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A., Pembrey, M.E., Holland, J., Oostra, B.A., Bobrow, M. and Davies, K.E. *Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation*. Cell. 1993; **74**, 127-134.

Kornberg, A. and Baker, T. *DNA Replication*. W.H. Freeman & Company. (2nd ed.) 1992; 783, 831-833.

Laird, P.W., Jacksongrusby, L., Fazeli, A., Dickinson, S.L., Jung, W.E., Li, E., Wainberg, R.A. and Jaebisch, R. *Suppression of intestinal neoplasia by DNA hypomethylation*. Cell 1995; **81**, 197-205.

Lehmann, J.M., Zhang, X.-K. and Pfahl, M. *RARγ2 expression is regulated through a retinoic acid response element embedded in Sp-1 sites*. Molec. Cell. Biol. 1992; **12**, 2976-2985.

Leroy, P., Krust, A., Zelent, A., Mendelsohn, C., Garnier, J.M., Kastner, P., Dierich, A. and Chambon, P. *Multiple isoforms of the mouse retinoic acid receptor α are generated by alternative splicing and differential induction by retinoic acid*. EMBO J. 1991; **10**, 59-69.

Leroy, P., Nakshatri, H. and Chambon, P. *Mouse retinoic acid receptor α2 isoform is transcribed from a promoter that contains a retinoic acid response element*. Proc. Natl. Acad. Sci. USA. 1991; **88**, 19137-10142.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. *Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA*. Cell. 1992; **69**, 905-914.

Li, E., Bestor, T. and Jaenisch, R. *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality*. Cell. 1992; **69**, 915-926.

Li, X.-S., Shao, Z.-M., Scheick, M.S., Eiseman, J.L., Sentz, D., Jetten, A.M., Chen, J.C., Dawson, M.I., Aisner, S., Rishi, A.K., Gutierrez, P., Schnapper, L. and Fontana, J.A. *RARβ inhibits breast cancer anchorage independent growth*. J. Cell Physiol.1995; **165**, 449-458.

Lipscomb, M.F. *The respiratory system*. Basic Pathology, 5th ed. (Kumar, V., Cotran, R.S. and Robbins, S.L. eds.) 1992; 385-436.

Lohnes, D., Kastner, P., Dierich, A., Mark, M. and Chambon P. *Function of retinoic receptor γ (RARγ) in the mouse*. Cell. 1993; **73**, 643-658.

Lohnes, D., Mendelson, M.M., Dollé, C., Dierich, A., Gorry, P., Gansmuller, A. and Chambon, P. *Function of the retinoic acid receptors (RARs) during development.* Cranofacial and skeletal abnormalities in RAR double mutants. Development. 1994; **120**, 2723-2748.

Lotan, R. *Effects of vitamin A and its analogs (retinoids) on normal and neoplastic cells.* Biochim. Biophys. ACTA. 1980; **605**, 33-91.

Lotan, R., Xu, X.-C., Lippman, S.M., Ro, J.Y., Lee, J.S., Lee, J.J. and Hong, W.K. *Suppression of retinoic acid receptor-β in premalignant oral lesions and its upregulation by isotretinoin.* New Engl. J. Med. 1995; **332**, 1405-1410.

Lu, X.P., Fanjul, A., Picard, N., Pfahl, M., Rungta, D., Nared-Hood, K., Carter, B., Peidrafita, J., Tang, S., Fabbrizio, E. and Pfahl, M. *Novel retinoid-related molecules as apoptosis inducers and effective inhibitors of human lung cancer cells in vivo.* Nat. Med. 1997; **3**, 686-690.

Luo, J., Pasceri, P., Conlon, R.A., Rossant, J. and Giguère, V. *Mice lacking all isoforms of retinoic acid receptor β develop normally and are susceptible to the teratogenic effects of retinoic acid.* Mech. Dev. 1995; **53**, 61-71.

Mabry, M., Nelkin, B.D., Falco, J.P., Barr, L.F., Baylin, S.J. *Transitions between lung cancer phenotypes - implications for tumour progression.* Cancer Cells 1991; **3**, 53-58.

Makos, M., Nelkin, B.D., Lerman, M.I., Latif, F., Zbar, B. and Baylin, S.J. *Distinct hypermethylation patterns at altered chromosome loci in human lung and colon cancer.* Proc. Natl. Acad. Sci. USA. 1992; **89**, 1929-1933.

Mangelsdorf, D.J., and Evans, R.M. *The RXR heterodimers and orphan receptors.* Cell. 1995; **83**, 841-850.

Mangelsdorf, D.J., Kliewer, S.A., Kakizuka, A., Umesono, K. and Evans, R.M. *Retinoid receptors.* Recent Progress Horm. Res. 1993; **48**, 99-121.

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. and Evans, R.M. *The nuclear receptor superfamily: the second decade.* Cell. 1995; **83**, 835-839.

Marwick, C. *Trials reveal no benefit, possible harm of beta carotene and vitamin A for lung cancer prevention.* JAMA. 1996; **275**, 422-423.

Maxwell, S.A. and Mukhopadhyay, T. Transient stabilization of p53 in non-small cell lung carcinoma cultures arrested for growth by retinoic acid. Exp. Cell Res. 1994; **214**, 67-74.

Mayne, S.T., Graham, S. and Zheng, T. *Dietary retinol: prevention or promotion of carcinogenesis in humans?* Cancer Causes Control. 1991; **2**, 443-450.

Merlo, A., Herman, J.G., Mao, Li, Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B. and Sidransky, D. *5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers.* Nature Med. 1995; **1**, 686-692.

Migeon, B.R. *X-chromosome inactivation: molecular mechanisms and genetic consequences*. Trends Genet. 1994; **10**, 230-235.

Minna, J.D. *The molecular biology of lung cancer pathogenesis*. Chest Supp. 1993; **103**, 449S-456S.

Minna, J.D., Sekido, Y., Fong, K.M., Gazdar, A.F. *Cancer of the lung*. Cancer principles & Practice of Oncology, 5th ed., 1997; 849-857.

Modrich, P. *DNA mismatch correction*. Annu. Rev. Biochem. 1987; **56**, 435-466.

Moghal, N. and Neel, B.G. *Evidence for impaired retinoic acid receptor-thyroid hormone receptor AF-2 cofactor activity in human lung cancer*. Molec. Cell Biol. 1995; **15**, 3945-3959.

Momparler, R.L., Momparler, L.F. and Samson, J. *Comparison of the antileukemic activity of 5-aza-2'-deoxycytidine, 1-beta-D-arabinofuranosylcytosine and 5-azacytidine against L-1210 leukemia*. Leuk. Res. 1984; **8**, 1043-1049.

Mooi, W.J. *Common lung cancers*. Spencer's Pathology of the Lung. (Hasleton, P.S. ed.) 5th ed. McGraw Hill. New York. 1996; 1009-1064.

Moon, R.C., Rao, K.V. N., Detrisac, C.J., Kelloff, G.J. *Animal models for chemoprevention of respiratory cancer*. J. Natl. Cancer Inst. 1992; **13**, 45-49.

Moon, R.C., Rao, K.V.N., Detrisac, C.J., Kelloff, G.J. *Hamster lung cancer model of carcinogenesis and chemoprevention*. Adv. Exp. Med. Biol. 1994; **66**, 649-653.

Morriss-Kay, G.M. and Sokolova, N. *Embryonic development and pattern formation.* FASEB J. 1996; **10**, 961-968.

Motoji, T., Hoang, T., Tritchler, D. and McCullough, E.A. *The effect of 5-azacytidine and its analogues on blast cell renewal in acute myeloblastic leukemia*. Blood. 1985; **65**, 894-901.

Nagpal, S., Saunders, M., Kastner, P., Durand, B., Nakshatri, H. and Chambon, P. *Promoter context - and response element-dependent specificity of the transcriptional activation and modulating functions of retinoic acid receptors*. Cell. 1992; **70**, 1007-1019.

Nan, X., Campoy, J. and Bird, A. *MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin*. Cell. 1997; **88**, 471-481.

Nan, X., Meehan, R.J. and Bird, A. *Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2*. Nucl. Acids Res. 1993; **21**, 4886-4892.

National Research Council. *Implications for reducing cancer risk*. Washington, D.C. National Academy Press. 1989.

Nelkin, B.D., Przepiorka, D., Burke, P.J., Thomas, E.D. and Baylin, S.B. *Abnormal methylation of the calcitonon gene marks progression of chronic myelogenous leukemia*. Blood. 1991; **77**, 2431-2434.

Nervi, C, Vollberg, T.M., George, M.D., Zelent, A., Chambon, P. and Jetten, Am. Exp. Cell Res. 1991; **195**, 163-170.

Ngo, V., Gourdji, D. and Laverrière, J.-N. *Site-specific methylation of the rat prolactin and growth hormone promoters correlates with gene expression*. Molec. Cell. Biol. 1996; **16**, 3245-3254.

Ohtani-Fujita, N., Fujita, T., Aoike, A., Osifchin, N.E., Robbins, P.D. and Sakai, T. *CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene*. Oncogene. 1993; **8**, 1063-1067.

Osann, K.E. *Lung cancer in women: the importance of smoking, family history of cancer, and medical history of respiratory disease*. Cancer Res. 1991; **51**, 4893-4897.

Osann, K.E., Anton-Culver, H., Kurosaki, T. and Taylor, T. *Sex differences in lung cancer risk associated with cigarette-smoking*. Int. J. Cancer. 1993; **54**, 44-48.

Park, J.-G. and Chapman, V.M. *CpG island promoter region methylation patterns of the inactive X-chromosome hypoxanthine phosphoribosyltransferase (hprt) gene*. Molec. Cell. Biol. 1994; **14**, 7975-7983.

Paroush, Z. Keshet, I., Yisraeli, J. and Cedar, H. *Dynamics of demethylation and activation of the α-actin gene in myoblasts*. Cell. 1990; **63**, 1229-1237.

Petkovitch, M., Brand, N.J., Krust, A. and Chambon, P. *A human retinoic acid receptor which belongs to the family of nuclear receptors*. Nature (London) 1987; **330**, 444-450.

Pfeiffer, G.P., Steigerwald, S.D., Hansen, R.S., Gartler, S.M. and Riggs, A.D. *Polymerase chain reaction-aided sequencing of an X chromosome-linked CpG island; methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability*. Proc. Natl. Acad. Sci. USA. 1990a; **87**, 8252-8256.

Pfeiffer, G.P., Tanguay, R.L., Steigerwald, S.D. and Riggs, A.D. *In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1*. Genes Dev. 1990b; **4**, 1277-1287.

Raizis, A.M., Schmitt, F. and Jost, J.P. *A bisulphite method of 5-methyl-cytosine mapping that minimises template degradation*. Anal. Biochem. 1994; **226**, 161-166.

Ramchandani, S., MacLeod, A.R., Pinard, M., von Hofe, E. and Szyf, M. *Inhibition of tumorigenesis by a cytosine-DNA, methyltransferase, antisense oligodeoxynucleotide*. Proc. Nat'l. Acad. Sci. USA. 1997; **94**, 684-689.

Razin, A. and Riggs, A.D. *DNA methylation and gene function*. Science. 1980; **210**, 604-610.

Reddel, R.R. and Harris, C.C. *Carcinogenesis*. Cancer of the lung. Thoracic Oncology. (Roth, J.A., Ruckdeschel, J.C. and Weisenburger, T.H. eds.) Harcourt Brace Jovanovitch Inc. Philadelphia. 1989; 16-37.

Rhodes, K., Rippe, R.A., Umezawa, A., Nehls, M., Brenner, D.A. and Breindl, M. *DNA methylation represses the murine α1(I) collagen promoter by an indirect mechanism*. Molec. Cell. Biol. 1994; **14**, 5950-5960.

Richardson, G.E., Johnson, B.E. *The biology of lung cancer*. Seminars in Onc. 1993; **20**, 105-127.

Risch, H.A., Howe, G.R., Jain, M. Burch, D.J., Holowaty, E.J. and Miller, A.B. *Are female smokers at higher risk for lung cancer than male smokers? A case control analysis by histologic type*. Am. J. Epidemiol. 1993; 138, 281-293.

Roberts, A.B. and Sporn, M.B *The Retinoids*, Vol 2. (Sporn, M.B., Roberts, A.B., Goodman D.S. eds.) New York Academic Press, Inc., 1984.

Rother, K.I., Silke, J., Georgiev, O., Schaffner, W. and Matsuo, K. *Influence of DNA sequence and methylation status on bisulfite conversion of cytosine residues*. Analyt. Biochem. 1995; **231**, 263-265.

Royer-Pokora, B. and Schneider, S. *Wilm's tumor-specific methylation pattern in 11p13 detected by PFGE*. Genes Chromosom. 1992; **5**, 132-140.

Ruddon, R.W. *Cancer Biology*, 3rd ed., New York. 1995.

Samet, J.M. *The epidemiology of lung cancer*. Chest Supp.1993; **103**, 20S-29S.

Sankaranarayanan, R. and Mathew, B. *Retinoids as cancer preventive agents. Principles of chemoprevention*. (Stewart, B.W., McGregor, D. and Kleihues, P. eds.) 1996; 47-59.

Sapienza, C., Peterson, A.C., Rossant, J. and Balling, R. *Degree of methylation of transgenes is dependent on gamete of origin*. Nature. 1987; **328**, 251-254.

Schauf, C., Moffett, D. and Moffett, S. *Human Physiology Foundations and Frontiers*. Times Mirror/Mosby College Publishing. 1990; 400-425.

Schroder, E.W. & Black, P.H. *Retinoids: tumor preventors or tumor enhancers*. J. Natl. Cancer Inst. 1980; **65**, 671-674.

Seewaldt, V.L., Johnson, B.S., Parker, M.B., Collins, S.J. and Swisshelm, K. *Expression of RARβ mediates RA-induced growth arrest and apoptosis in breast cancer cells*. Cell Growth Differ. 1995; **6**, 1077-1088.

Segall, L. *Étude de l'adaptation métabolique de la cellule β(INS-1) à des niveaux élevés d'acides gras libres*. M.Sc. Thesis. Université de Montréal. Montréal, Canada, 1997.

Sellers, T.A., Bailey-Wilson, J.E., Elston, R.C., Wilson, A.F., Elston, G.Z., Ooi, W.L., Rothschild, H. *Evidence for Mendelian inheritance in the pathogenesis of lung cancer*. J. Natl. Cancer Inst. 1989; **82**, 1272-1279.

Shapiro, R., Braverman, B., Louis, B. and Servis, R.E. Nucleic acid reactivity and conformation. II. *Reaction of cytosine and uracil with sodium bisulphite*. J. Biolog. Chem. 1974; **248**, 4060-4064.

Shen, S., Kruyt, F.A.E., den Hertog, J., van der Saag, P.T. and Kruijer, W. *Mouse and human retinoic acid receptor β2 promoter: sequence comparison and localization of retinoic acid responsiveness*. J. DNA Sequ. & Mapp. 1991; **2**, 111-119.

Slebos, R.J.C. and Rodenhuis, S. *The molecular genetics of human lung cancer*. Eur. Resp. J. 1989; **2**, 461-469.

Sozzi, G., Veronese, M.L., Negrini, M., Baffa, R., Cotticelli, M.G., Inoue, H., Tornielli, S., Pilotti, S., De Gregorio, L., Pastorino, U., Pierotti, M.A., Ohta, M., Huebner, K. and Croce, C.M. *The FHIT gene at 3p14.2 is abnormal in lung cancer*. Cell. 1996; **85**, 17-26.

Sporn, M.B., Roberts, A.B. and Goodman, eds., D.S. *The Retinoids: Biology, Chemistry and Medicine,* 2nd ed.. Raven Press, New York. 1994.

Stanley, K. and Stjernsward, J. *Lung cancer - a woldwide health problem*. Chest. 1989; 1s-5s.

Stryer, L. *Biochemistry* 3rd ed., W.H. Freeman & Company. New York. 1988.

Subak-Sharpe, H., Burk, R.R., Crawford, L.V., Morrison, J.M., Hay, J. and Keir, H.M. *An approach to evolutionary relationships of mammalian DNA viruses through the analysis of the pattern of nearest neighbor base sequences*. Cold Spring Harb. Symp. 1966; **31**, 737-748.

Sun, S.Y. Yue, P., Shroof, B., Hong, W.K. and Lotan, R. *Induction of apoptosis in human non-small cell lung carcinoma cells by the novel synthetic retinoid CD437*. J. Cell Physiol. 1997; **173**, 279-284.

Sundaresan, V., Rabbitts, P.H. *Genetics of lung tumors*. Spencer's Pathology of the Lung. (Hasleton, P.S. ed.) 5th ed. McGraw Hill. New York. 1996; 987-1009.

Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D. and Warren, S.T. *DNA methylation represses FMR-1 transcription in fragile-X syndrome*. Hum. Mol. Genet. 1992; **1**, 397-400.

Suzuki, H., Ueda, R., Takahashi, T. and Takahashi, T. *Altered imprinting in lung cancer*. Nature Genetics. 1994; **6**, 332-.

Szyf, M. *The DNA methylation machinery as a target for anticancer therapy*. Pharmacol. Ther. 1996; **70**, 1-37.

Tasheva, E.S. and Roufa, D.J. *Densely methylated DNA island in mammalian chromosomal replication origins*. Mol. Cell. Biol. 1994; **14**, 5636-5644.

Tate, P., Skarnes, W. and Bird, A. *The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse*. Nature Genetics. 1996; **12**, 205208.

Tazi, J. and Bird, A. *Alternative chromatin structure at CpG islands*. Cell. 1990; **60**, 909-920.

Tortora, G.J., and Anagnostakos, N.P. *Principles of Anatomy and Physiology*, 5th ed. Harper & Row Publishers, Inc. 1987; 584-621, 690.

Toulouse, A., Morin, J., Pelletier, M., Bradley, W.E.C. *Structure of the human retinoic acid receptor β1 gene*. Biochim. Biophys. ACTA 1996; **1309**, 1-4.

Toulouse, A., Morin, J., Ying, Y., Ayoub, J. and Bradley, W.E.C. *Retinoic acid receptor β: an exploration of its role in lung cancer suppression and its potential in cancer prevention*. Clinical and biological basis of lung cancer prevention. (Martinet, Y., Hirsch, F.R., Martinet, N., Vignailt, J.M. and Mulshine, J.L. eds.) Birkhauser. Switzerland. 1997; 198-209.

Underwood, B.A. *Vitamin A in human nutrition: public health considerations*. The Retinoids: Biology, Chemistry and Medicine, 2nd ed. (Sporn, M.B., Roberts, A.B. and Goodman, D.S., eds.) Raven Press, New York. 1994; 211-227.

Vinocour, M., Minna, J.D. *Cellular and molecular biology of lung cancer*. Thoracic Oncology. (Roth, J.A., Ruckdeschel, J.C. and Weisenburger, T.H. eds.) Harcourt Brace Jovanovitch Inc. Philadelphia. 1989; 38-51.

Vogelstein, B., Kinzler, K.W. *The multistep nature of cancer*. TIG. 1993; **9**, no.4, 138-141.

Wald, G. *Molecular basis of visual excitation*. Science. 1968; **162**, 230-239.

Walker, C. and Nettesheim, P. *In vitro neoplastic progression of rat tracheal epithelial cells transformed by diverse carcinogens*. Cancer Res. 1989; **49**, 4427-4430.

Wang, R.Y.-H., Gehrke, C.W. and Ehrlich, M. *Comparison of bisulphite modification of 5-methyldeoxycytidine and deoxycytidine residues*. Nucleic Acids Res. 1980; **8**, 4777-4790.

Weinberg, R.A. *Tumor suppressor genes*. Science 1991; **254**, 1138-1146.

Whang-Peng, J., Kao-Shan, C.S., Lee, E.C., Bunn, P.A., Carney, D.N., Gazdar, A.F. and Minna, J.D. *Specific chromosome defect associated with human small cell lung cancer: deletion 3p(14-23)*. Science. 1982; **215**, 181-182.

Willey, J.C., Saladino, A.J., Ozanne, C. *Acute effects of 12-O-tetradecanoylphorbol-13-acetate, teleocidin B, or 2,3,7,8,-tetrachlorodibenzo-p-dioxin on cultured normal human bronchial epithelial cells*. Carcinogenesis. 1984; **5**, 209-215.

Wolbach, S.B. and Howe, P.R. *Tissue changes following deprivation of fat soluble A vitamin*. J. Exp. Med. 1925; **42**, 753-777.

World Health Organization Histological Typing of lung tumours, 2nd ed., Geneva, Switzerland. Am. J. Clin. Pathol.1982; **77**, 123-136.

Xiong, Z. and Laird, P.W. *COBRA: a sensitive and quantitative DNA methylation assay*. Nucl. Acids Res. 1997; **25**, 2532-2534.

Yong, L.C., Brown, C.C., Schatkin, A., Dresser, C.M., Slesinski, M.J., Cox, C.S. and Taylor, P.R. *Intake of vitamins E, C, and A and risk of lung cancer. The NHANES I epidemiologic followup study. First national health and nutrition examination survey*. Am J. Epidem. 1997; **146**. 231-243.

Zang, E.A., Wynder, E.L. *Differences in lung cancer risk between men and women: examination of the evidence*. J. Natl. Canc. Inst. 1996; **88**, 183-192.

Zelent, A., Mendelsohn, C., Kastner, P., Krust, A., Garnier, J.M., Ruffenach, Leroy, P. and Chambon, P. *Differentially expressed isoforms of the mouse retinoic acid receptor b are generated by usage of two promoters and alternative splicing*. EMBO J. 1991; **10**, 71-81.

Zhang, X.K., Liu, Y., Lee, M.O. and Pfahl, M. *A specific defect in the retinoic acid response associated with human lung cancer cell lines*. Cancer Res. 1994; **54**, 5663-5669.

Appendix

CpG site number	Sequence Position
1	428
2	459
3	464
4	471
5	478
6	482
7	486
8	492
9	500
10	517
11	520
12	534
13	561
14	577
15	582
16	651
17	654
18	663
19	667
20	784
21	798
22	810

Table XXIII:
DNA sequence position according to CpG site number within the human RARβ P2 promoter region, starting approximately 0.5 Kb upstream from the transcription start site. Please refer to GenBank accession no. X56849. (Based on Shen *et al*, 1991). **RAR** = Retinoic acid receptor; **CpG** = Cytosine-Guanine dinucleotide.