Chapter 1

Introduction

1.1 The Cell Theory

The question of how cells grow and divide has fascinated biologists for over a century. In 1665, Robert Hooke published *Micrographia*, the first important work devoted to microscopial observations, in which he described that he had been looking at something similar to *cellula*. What he had actually been looking at were non-living cells of cork and was possibly the first use of the word 'cell' in a historical context. He was describing the cellulose walls that no longer contained living components and was therefore unwittingly describing programmed cell death. The first man to observe a live cell under the microscope was Anton Van Leeuwenhoek in 1674, who described the algae *Spirogyra*. Under the microscope he saw particles which were motile, and after assuming that motility was equivalent to life, he concluded in 1676, that they were indeed living organisms, and went on to describe specific forms, including microorganisms (Mazzarello, 1999). From this, in 1682, Nehemiah Grew described 'cells or bladders' of root parenchyma and building on these descriptions, several advances, including the discovery of the nucleus, before the cell theory was proposed in 1838 by Schielden and Schwann (Scott & Logan, 2004). The two main tenets of the cell theory are as follows:

- All living organisms are composed of one or more cells.
- New cells can arise only by the division of pre-existing cells.

Viruses are the only apparent exception to this theory, but since they can only replicate inside cells their continued existence depends on the continued reproduction of cells (Murray & Hunt, 1993).

In order to understand the complexity of the cell cycle, basic properties of cells must first be understood. Cells are complicated objects of which the two major types are eukaryotic and prokaryotic. The main distinction between eukaryotic and prokaryotic cells is that eukaryotic cells have discrete organelles, including a nucleus, mitochondria, lysosomes and their DNA is enclosed within the nucleus. Prokaryotic cells lack any distinction between the nucleus and cytoplasm (Murray & Hunt, 1993).

A cell is a discrete entity with the ability to both grow and divide and so must have a number of constituents in place, including a set of instructions and production facilities to essentially make the components that will form the new cell. These instructions are encoded within the DNA sequence of the chromosomes. The cell cycle is the universal process by which cells reproduce and essentially underlies the growth and development of all living organisms (Nurse *et al,* 1998). During the cell cycle, chromosomes have two main functions since they carry the genetic information which is duplicated into the two daughter cells by replication enzymes. They must also express their genes for the production of two protein classes which act as cell cycle regulators and form the next generation of cells. Other components required for cell growth and division are a mechanism in place for dividing one cell into two, and a defined boundary between itself and its environment across which nutrients and wastes are imported and exported, respectively (Murray & Hunt, 1993).

1.2 The Cell Cycle

Following on from the work of Schielden and Schwann, mitosis was first described in 1875, but following this, cell cycle research essentially 'slept' for almost 100 years, and very little progress was made until 1951 when the concept of the cell cycle was proposed by Howard and Pelc. They succinctly described the four phases of the cell cycle that are known today, i.e. G_1 phase, S phase, G_2 phase and M phase.

The second breakthrough came with the introduction of high-resolution autoradiography with ^{3}H-Tdr (tritiated thymidine), which has two advantages:

tritium is a weak emitter and its label is localised precisely within the cells and in cell compartments. Tdr (thymidine) is also a specific precursor of DNA. Hughes *et al* in 1958, using epithelial cells discovered that newer cells replaced older cells, allowing other scientists to discover that the cell cycle exists in a variety of cells (Baserga, 1999).

The third breakthrough came in 1962 when Stanley Cohen, using mice salivary glands, purified a polypeptide causing the early eruption of the incisor teeth and the precocious opening of the eyelids in newborn mice. It later appeared he had managed to purify the first growth factor (GF), which was termed the 'epidermal growth factor' (EGF). By 1962, it was known that cells cycle by moving from one phase to another or even out of the cell cycle and is stimulated by GFs (Baserga, 1999).

1.2.1 Basic Properties of the Cell Cycle

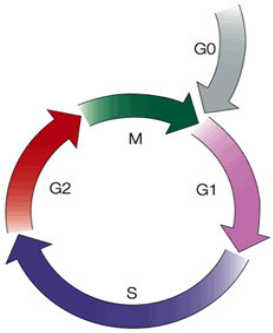
The cell cycle is a 'conscious' series of events whose primary functions are to first replicate and then segregate the genome. Criteria such as cell size and nutrient status must be first fulfilled before the cells 'decide' to enter the cell cycle and divide. At any one time, most cells are in a state known as G_0, which is the resting state. Once a cell is committed to the cell cycle, the cells pass through the first gap phase (known as G_1), followed by DNA synthesis (known as S phase), a second gap phase (known as G_2) before finally dividing by mitosis (known as the M phase) (Hagan, 2005).

1.2.2 Phases of the Cell Cycle

The cell cycle is a ubiquitous and complex process involved in the growth and proliferation of cells, organismal development, regulation of DNA repair and tissue hyperplasia as a result of tissue injury and diseases such as cancer. As stated previously, the cell cycle can be morphologically divided into phases known as interphase and M phase. Interphase encompasses the G_1, S and G_2 phases and M phase encompasses prophase, metaphase, anaphase and telophase (Schafer, 1998). During interphase, the cells simply grow in size. DNA replication occurs

during the S phase and is preceded by the first gap phase G_1 during which the cell prepares for DNA synthesis. This stage is then followed by a second gap phase known as G_2, during which the cell prepares for mitosis. Figure 1.1 shows the traditional subdivisions of the cell cycle which are G_1 » S » G_2 » M. Cells present in G_1 are able (before committing to DNA replication) to enter a resting state called G_0, which accounts for the major part of non-growing, non-proliferating cells in the human body (Vermeulen *et al*, 2003). The cell cycle lasts approximately 24 hours with interphase occupying the majority of the cell cycle. During interphase, the DNA is diffusely distributed within the nucleus and individual chromosomes cannot be distinguished and little activity is detected (Murray & Hunt, 1993).

The two most important events of the eukaryotic cell cycle are both S phase and M phase, where chromosomes are replicated and segregated into dividing cells. S phase and M phase form an ordered developmental sequence that ensures each newly divided cell receives a full complement of the hereditary material and so are discussed in more detail in sections (1.2.3) – (1.2.4) (Nurse, 1994).



Nature Reviews | Molecular Cell Biology

Figure 1.1 The eukaryotic cell cycle. The cell cycle is the ordered set of events through which a cell goes from one cell division to the next. The cell cycle is divided into four distinct phases; the first gap phase known as the G_1 phase, DNA synthesis phase known as S phase, the second gap phase known as the G_2 phase and mitosis known as M phase. Cells in the G_0 phase are said to be 'quiescent' and are no longer part of the cell cycle (Taken and adapted from Coleman *et al*, 2004).

1.2.3 DNA synthesis (S Phase)

As stated previously, the two most important phases of the cell cycle are considered to be the S phase and M phase and here I will go into more detail about S phase to emphasize its importance. S phase, short for synthesis phase is the phase of the cell cycle during which chromosomal DNA replication occurs. The length of S phase varies significantly between cell types and several absolute controls exist to dictate when and how many times during a cell cycle DNA replication can occur. For example, controls are in place to ensure that chromosomes are duplicated just once during a cell cycle therefore once this takes place, further duplication within the same cycle is not permitted and in only a few cases can this be breached, for example, polyploidy in mouse embryonic trophectoderm arises from multiple rounds of DNA replication without an intervening mitosis. Research has shown that under normal circumstances, a cell undergoes mitosis before it can re-replicate its chromosomal DNA and this is controlled via several mechanisms (Dalton, 1998).

The discovery of the S phase identified two key questions:

(1) How does the machinery of DNA replication work?

(2) What determines the onset of the S phase during the cell cycle?

The first problem has been worked on for over 40 years and began with the discovery of DNA polymerase and has resulted in the unravelling of the molecular mechanisms and enzymology of DNA replication (Murray & Hunt, 1993). In order for DNA replication to occur, two tasks must first be performed; activation of the replication origins (specialised regions of the chromosomes where DNA replication starts) and activation of the enzymatic machinery that replicates the DNA. Upon initiation of replication at an origin, the two strands of DNA unwind to form a replication fork whereby each acts as a template for the synthesis of a new strand. From this, two double-stranded daughter DNA molecules are formed and each is identical to the parent molecule. The DNA helix is a very stable structure and under laboratory conditions requires temperatures of around 90 °C before separation of the two strands. In a cell, it is the combined actions of several proteins that help the two strands to separate and much of this knowledge comes

from studies involving *Escherichia coli*. The synthesis of the new DNA strand is catalysed by DNA polymerase whose substrates are the four-deoxyribonucleoside triphosphates dATP, dCTP, dGTP and dTTP. DNA polymerase catalyses the formation of a phosphodiester link between the 3'-hydroxyl group of one sugar residue and the 5'-phosphate group of the second sugar residue and the base sequence of the new DNA strand dictated by the base sequence of the parental strand (figure 1.2). DNA polymerase (an enzyme that catalyses DNA replication) adds the nucleotide to a free 3'-hydroxyl group and so DNA is synthesised in the 5' to 3' direction (Bolsover, 2004). In addition to DNA polymerase, RNA primers are also required, which is an enzyme that catalyses the formation of a short stretch of RNA complimentary in sequence to the DNA template strand. RNA primers are a small piece of RNA that attaches to the single stranded (ss) DNA and acts as the starting point for DNA replication since DNA polymerase can only add new nucleotides to an existing strand of DNA. The RNA primer, using DNA polymerase adds nucleotides matched to the template strand (Kornberg, 1988). Due to the antiparallel structure of DNA, one new DNA strand is synthesized continuously in the direction of fork movement and is called the leading strand. The other, known as the lagging strand, grows in the direction away from fork movement. On this strand, short discontinuous segments of DNA, called Okazaki fragments, are synthesized from RNA primers. During replication, the RNA primers are removed, and each fragment is joined to complete lagging strand synthesis (Bambara *et al*, 1997).

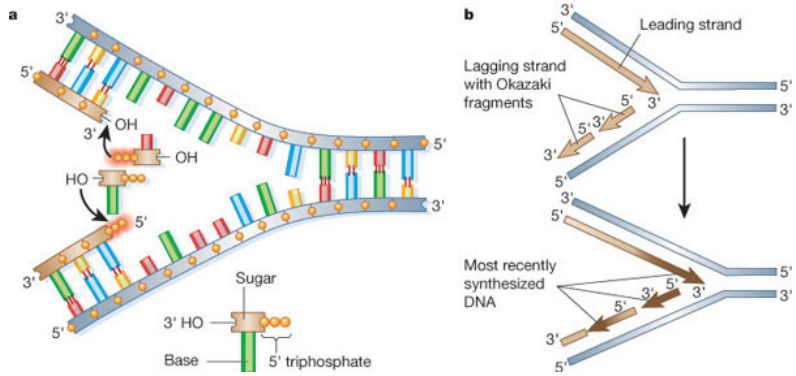


Figure 1.2 The replication fork. (a) shows how nucleoside tripohsphates, such as dATP, act as the substrates for DNA polymerase. Each of the nucleoside triphosphates is made up of 3 phosphates (yellow circles),a deoxyribose sugar (beige triangle) and one of the four bases. (b) shows how DNA polymerases catalyse the growth in the 5' to 3' direction only and the new daughter strands grow at the fork. The leading strand grows continuously, whilst the 'lagging' strand is synthesized by DNA polymerase through the backstitching mechanism illustrated (Taken from Alberts, 2003).

1.2.4 Mitosis (M phase)

It is during the mitotic phase of the cell cycle (often referred to as the M phase) that cell division occurs (Huberman, 2003). As cells enter mitosis, they undergo a sequence of structural changes in which the chromosomes condense, the nuclear envelope breaks down, the membranes of the Golgi apparatus and the endoplasmic reticulum fragment and the cytoskeleton undergoes remodelling to form a completely new cellular structure known as the mitotic spindle. Like every stage of the cell cycle, the M phase is tightly controlled (Bolsover, 2004). The purpose of M phase is to produce two progeny cells that each contains an identical set of the progenitor cell's chromosomes. To allow this, the chromosomes execute a precisely choreographed sequence of movements first described over a century ago. Figure 1.3 shows how mitosis is divided into five stages, each of which is characterised by changes in the appearance of the chromosomes and their organisation with respect to the mitotic spindle, which is responsible for their segregation.

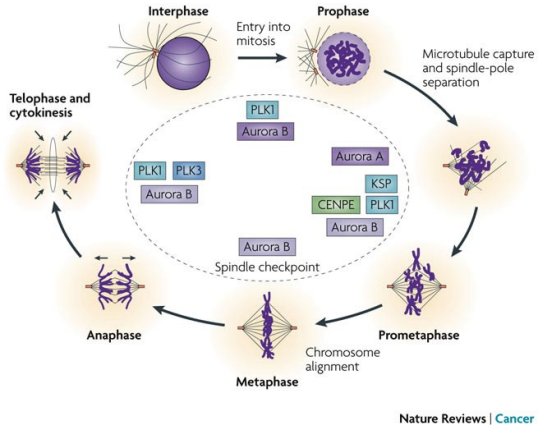


Figure 1.3 Phases of mitosis. Mitosis is sub-divided into 5 main stages, prophase, prometaphase, metaphase, anaphase and telophase. Prior to mitosis, cells undergo interphase. Also shown in this figure are specific targets for each stage of mitsosis and represents a potential therapeutic target (Taken from Jackson *et al*, 2007).

1.2.5 Cell Cycle Checkpoints

Both the S phase and M phase must occur in the correct order AND only once per cycle. To ensure that this occurs, the two phases do not immediately follow each other but are separated by 'gaps' which allow checks to take place ensuring everything is correct before proceeding onto the next stage of the cell cycle. For dividing cells, two points of the cell cycle are critical. The first is the G_1/S boundary during which the cell may be temporarily stopped and enter G_0 phase or the cell is committed to DNA replication and enter S phase after passing the retinoblastoma (Rb) gene in late G_1 (Foster, 2008). The second is the G_2/M transition at which point the cell is committed to mitosis. It is at these positions where the major control points of the cell cycle are carried out, and before 'crossing' them, a cell must be sure that conditions are such that both S and M phase are executed successfully. For example, at G_1/S, the cell must essentially 'decide' whether it is big enough and if nutritional conditions are appropriate to begin the crucial process of replicating its genome. It is essential that G_1 occurs quickly as DNA is very sensitive to damage since the bases are exposed and the DNA strands are separated. During G_2 phase the main concern is that the DNA is in perfect condition before entering mitosis. These mechanisms are sensitive and are explained below and in general cells do not commit themselves to mitosis until any defects have been 'repaired' (Foster, 2008; Bolsover, 2004).

1.2.6 Conventional Cell Cycle Regulation and Control

Cell cycle transition from one phase to another occurs in an orderly fashion, and can progress in one direction only, that is G_1 » S » G_2 » M and for this reason, the cell cycle is highly regulated and controlled. Cell cycle control mechanisms either allow or restrain its progression. When DNA damage occurs, it is essential that the cell cycle is stopped in its progression for DNA repair. The main families of regulatory proteins are the cyclin-dependent kinases (CDKs), which are a family of serine/threonine protein kinases and are activated at specific points of the cell cycle only and play roles in controlling cell-cycle progression. Other key regulators are the cyclins, the CDK substrate proteins, the CDK inhibitors (CKI) and the tumour suppressor gene products p53 and Rb. These families comprise the basic regulatory machinery responsible for catalysing cell cycle transition and checkpoint traversal (Golias *et al*, 2004)**.** Several types of cyclins have been identified by immunohistochemistry namely cyclin A, B, D_1, D_3 and E and each has a specific role and each activates specific CDKs. Cyclin A is particularly interesting since it can activate two different CDK functions to regulate both S phase and M phase. Other cyclins can regulate the entry of cells into S phase and the disruption of these regulatory proteins can lead to disease state, for example, the over expression of cyclin D_1 in G_1 frequently occurs in human oesophageal cancer. The activity of different CDKs and their regulatory subunits delicately control progression through each phase of cell cycle. CDKs 4, 6 and 2 are active during G_1 to S phase transition whereas complexes such as CDK2 and cyclin E are important for initiation of the S phase. It has also been identified that CDK1 is active during G_2 and M phase (Vermeulen *et al*, 2003; Murray, 2004).

The progression of the cell cycle is dependent on the assembly and consequential activation of these CDKs. Active CDKs are comprised of a catalytic subunit, typically around 34 kDa, and a regulatory subunit, known as a cyclin. The activation of a CDK requires the cyclin to associate with the catalytic subunit that is both phosphorylated on a threonine residue and dephosphorylated on a tyrosine residue (figure 1.4). Once activated, the CDKs induce downstream processes via phosphorylation of selected proteins. The levels of CDK proteins remain

relatively stable throughout the cell cycle, however, the levels of cyclins rise and fall throughout the cell cycle and so periodically activate the CDKs. There are different cyclins, each being required at different stages of the cell cycle, for example, there are 3 types of cyclin D (cyclin D1, D2 and D3), which bind to CDK4 and 6, and are essential for entry into G_1. Another cyclin required during G_1 is cyclin E, which associates with CDK2 to allow progression into S phase. Once activated, the CDKs can be inactivated in a number of ways:

- Cyclin levels are reduced by turning down their transcription or by degradation of the protein itself.
- The catalytic subunit can be dephosphorylated at the activating threonine residue and can be inhibited by phosphorylation of the tyrosine residue.
- CKI's bind to the CDK-cyclin complex and inhibit its activity (Peter & Herskowitz, 1994).

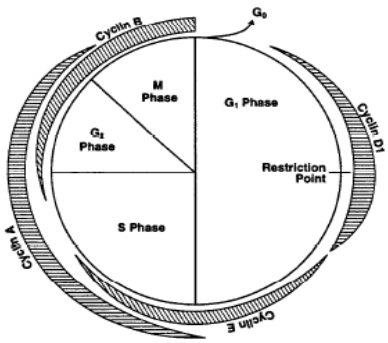


Figure 1.4 Regulation of the cell cycle. The cell cycle is highly regulated as DNA synthesis and replication must occur once, and once only during the cell cycle. The events of the cell cycle must also occur in one direction only and to regulate this are a series of cyclins and cyclin dependent kinase which act together to ensure the occurance of the correct events in the cell cycle (Taken from Schafer, 1998).

When these checkpoints are eliminated, be it by mutation or some other means, the end result can be cell death, infidelity of chromosome transmission or increased susceptibility to environmental perturbations, such as DNA damaging agents (Hartwell & Weinert, 1989).

The cyclins are structurally identified by conserved 'cyclin box' regions. These are 56 kDa proteins first discovered in rapidly dividing cells and are implicated in the mitosis of all eukaryotic cells. Cyclins are regulatory subunits of holoenzyme CDK complexes controlling progression through cell cycle checkpoints by phosphorylating and inactivating target substrates. The cyclins are expressed and degraded in a concerted fashion to drive the individual stages of the cell cycle to ensure the phases of the cell cycle occur in the correct order and just once per cycle. The cyclins combine with CDKs to form activated kinases which then phosphorylate specific targets and regulate the cell cycle (Foster, 2008).

1.2.7 DNA Damage and Repair

Aside from the normal cell cycle regulation as described in section 1.2.6, there are other regulatory processes which exist. DNA is constantly bombarded with both exogenous and endogenous agents and so it is necessary for a complex, highly preserved and regulated system to be present that routinely checks for and repairs any damage. For example, there are restriction points and blocks which result from DNA damage, to ensure DNA damage is not passed to the daughter cell as the result would be cancer (reviewed in section 1.3). Normal animal cells *in-vivo* have alternative modes of existence in that they can either be proliferating or quiescent and can survive in either state. Deoxynucleotide depletion or even a single double-strand break in DNA can create a major checkpoint which is necessary to halt the cell cycle and prevent mutations (Ford & Pardee, 1999). These checkpoints allow for DNA repair to occur before entry into either S phase or M phase, again highlighting the importance of these two phases in the cell cycle. A breakdown in regulation of the cell cycle leads to uncontrolled growth of the cell and defects in many of the molecules that regulate the cell cycle have been implicated in disease states, such as cancer. Key amongst the regulatory proteins are p53, CKI (p15, p16, p18, p19, p21, p27), and Rb, all of which act to keep the cell cycle from progressing until all DNA repair has been completed. There are 3 types of DNA repair depending upon the damage sustained:

1. Simple lesions can be quickly repaired with no effect on the cell cycle.

2. Complex lesions undergo a local response which entails the recruitment of additional factors and provides the first line of defence.
3. Failure of these leads to activation of the DNA damage response, known as the 'global response', the specific details of which have not yet been fully elucidated (Wu *et al*, 2005).

1.2.8 The Cell Cycle and DNA Damage

Cellular responses to DNA damage constitute one of the most important fields in cancer biology and damage to cellular DNA is the cause of cancer. Cancer is a collection of diseases with the common feature of uncontrolled proliferation and the inhibition of repair processes. There are several cellular changes required to generate a cancer, and these are discussed in more detail in section 1.3.2. All cell types are susceptible to changes, however, it is epithelial cells which are the most prone (King, 2000). Many already existing cancer chemotherapies act by interfering with DNA replication and formation of the mitotic spindle and as already stated, these are processes central to the cell cycle. Other strategies which target cell cycle components and specifically deregulate them in tumour cells are also being explored. DNA is under constant attack by agents with the ability to either directly damage one of its three billion bases or break the phosphodiester backbone upon which the bases reside. A good example of this is oxygen free radicals, which cause base damage and DNA breakage. It is the cellular response to this DNA damage, which constitutes one of the most important fields of cancer biology (Wiseman & Halliwell, 1996). According to epidemiological studies, animal models and other sources such as the observation that many human-cancer-susceptibility syndromes arise from mutations in genes involved in DNA-damage responses, it is DNA damage that causes cancer and DNA damage that cures cancer (Kastan & Bartek, 2004).

Given the critical significance of error-free cell cycle progression and DNA replication for the maintenance of genomic integrity and preventing cancer, the most vulnerable stages of the cell cycle are protected by a wider spectrum of checkpoint effector mechanisms (Kastan & Bartek, 2004). For example, during

normal cellular proliferation, cells withdraw from the cell cycle after GF deprivation or growth inhibitory signals in early-to-mid G_1 phase. This is before the cells pass through the Rb/E2F (transcription factor)-controlled restriction point. After this checkpoint, they are committed to a round of DNA replication and cell division (Nevins, 2001; Massague, 2004). However, the ATM/ATR−CHK2/CHK1-controlled checkpoint network response to genotoxic stress can transiently delay the progression of the cell-cycle in G_1, S or G_2 phases, or even impose prolonged cell-cycle arrests in either G_1 or G_2, prior to entry into the S phase or M phase, respectively. It is well known that ATR is activated following DNA damage induced by ultra-violet (UV) and ATM is activated in response to double strand breaks (Helt *et al*, 2005).

To summarise sections (1.2.6) – (1.2.8), it is evident that there are environmental changes which induce DNA damage and as a result it is essential that the cell cycle is stopped in its progression to ensure DNA repair, and we have these repair processes in place in the form of regulatory proteins and cell cycle checkpoints. It is also evident that cellular DNA damage and inhibition of these repair processes is the cause of cancer, and it is DNA damage that is used to cure this disease state.

1.2.9 Senescence

Senescence is a complex phenomena and is another important process involved with the cell cycle and refers to the loss of the ability to divide and occurs after approximately fifty cell divisions *in vitro*, though can occur before as a result of DNA damage or toxins. Cellular senescene was first recognized over forty years ago as a process that prevented normal human fibroblasts from growing indefinitely in culture (Campisi, 2001). Senescence is a process activated by normal cells in response to various types of stress, including oncogene activity, oxidative stress and others. Once a cell has entered senescence, they lose the ability to divide, and undergo a series of dramatic morphologic and metabolic changes whereby the cell stops responding to mitogenic stimuli and undergo changes in their chromatin structure and gene expression in addition to becoming enlarged and flattened. These are evidence to suggest that cellular senescence is a

cellular program that acts to prevent further cell proliferation in situations of physiologic stress and acts parallel to apoptosis (Ben-Porath & Weinberg, 2004; Ben-Porath & Weinberg, 2005). In addition to a flattened morphology, there is also an increase in the size of their nucleus and nucleoli, an increase in the number of lysosmed and the appearance of vacuoles in the cytoplasm in senescent cells (Cristofalo *et al*, 2004). Senescent cells, even though they lack the ability to replicate their DNA, still have their replicative machinery in place. Importantly, senescence has been proposed as a mode of tumour suppression by limiting the growth potential of cells (Sager, 1991).

1.2.10 Methods of Cell Cycle Analysis

In studies of the mammalian cell cycle, one of the prime objectives is to identify cells in the process of replication and as a result, the incorporation of DNA precursors, such as ^{3}H-Tdr has been one of the main tools for such analysis. Scintillation spectrophotometry provides the quantitative information on the amount of precursor incorporated into the DNA of replicating cells, and both electron microscopy and light microscopy allows the localisation of the incorporation and identification of the replicating population (Krishnan, 1975). Cell cycle assays have been redefined over the past several years and is largely due to new technologies such as flow cytometry. Flow cytometry is a technique used for the rapid measurement of particles or cells as they flow in a fluid stream one by one through a sensing point. The basic principles of flow cytometry arise from very old ideas generated early in the 20th Century and basically follow the principles of laminar flow defined by Reynolds in the late 19th Century. Fifty years later, Maldavan designed a machine with the ability to identify single cells using a microscope and photodetector. Flow cytometry allows a very large number of cells to be evaluated in a very short time with some systems having the ability to run particles at rates of up to 100,000 particles per second whilst also collecting 10 to 20 parameters from each cell (Robinson, 2004). An important feature of flow cytometric analysis is that measurements are made separately on each particle within the suspension in turn, thus what may appear to be a

homogenous population using conventional bulk assays may be demonstrated to be multiple populations by flow cytometry (Carter, 1994; Robinson *et al*, 1997). Flow cytometry is the major direct way of analysing stages of the cell cycle, however, there are also indirect methods, such as radiolabelled antibodies which label a protein of interest, such as CDKs. With the knowledge of the different proteins in abundance at certain stages of the cell cycle, the point in the cycle of a population of cells can be determined. However, flow cytometry remains the most accurate approach.

A number of different protocols have been proposed that employ propidium iodide (PI) as a fluorescent stain for DNA content in flow cytometry. The earlier attempts required fixation of the sample in ethanol whilst later methods required a single step procedure utilising a hypotonic fluorochrome solution to lyse the unfixed cells and allow PI to penetrate the cells. PI intercalates with both double-stranded (ds) RNA and DNA and therefore there is the need to denaturate the RNA using a hypotonic citrate buffer (Deitch *et al*, 1982). Incubation of the cells in a hypotonic solution containing PI allows the structural ability of the plasma membrane for dye exclusion to be lost and within a few minutes of staining, a fluorescent stain of the DNA complex is observed (Krishnan, 1975). Using PI to determine cell cycle distributions is based on a mathematical model that assumes the population of cells is separated into compartments each consisting of cells with approximately the same amount of DNA content. Cells in G_1 phase are in one compartment, those in G_2 and M are grouped in another and S phase cells are in several compartments according to their position in S. The fluorescence intensity of cells in each compartment is distributed normally with the mean intensity of those in G_2/M equal to twice that of those in G_1 (Fried *et al*, 1976).

Cell cycle analysis techniques also exist which allow specific phases of the cell cycle to be essentially 'highlighted' and therefore accurately distinguished. Two of these techniques are described below.

Mitosis is a valuable indicator of active tissue proliferation yet other than morphological characteristics, previous to 1999 there were no markers available to detect only M phase cells. Hirata *et al* in 2004 discovered a monoclonal antibody

(MAb) known as HTA28 which recognizes histone H3 harbouring phosphoserine 28. Histones are the major protein constituents of chromatin in eukaryotic cells and are composed of an octamer of core histones, two of each of H2A, H2B, H3 and H4, around which there are two superhelical turns of DNA are wrapped, and are approximately 80 bp in length (Goto *et al*, 1999; Hans & Dimitrov, 2001). Histone H3 is specifically phosphorylated during both mitosis and meiosis in patterns that are coordinated with both space and time. Chromosome condensation, as described in section 1.2.4, is accompanied by the phosphorylation of histone H3 and the hyperphosphorylation of its linker histone H1. Histone H3 phosphorylation at serine 10 in conjunction with phosphorylation at serine 28 is also required for the proper segregation and condensation during mitosis (Bradbury, 1992; Hans & Dimitrov, 2001). The use of the antibody HTA28 that can recognise H3 phosphorylated at serine 28 but not 10 and is useful in distinguishing M phase cells within a population of cells.

The Ki-67 antibody recognises a nuclear antigen present in proliferating cells, but is absent in G_0 phase cells. In 1983, a study by Gerdes *et al* showed that the Ki-67 antibody did not react with cells known to be in the resting stage such as lymphocytes, monocytes and parietal cells. It was also observed that expression of the antigen recognised by Ki-67 could be induced in peripheral blood lymphocytes after stimulation with phytohaemagglutinin (PHA) whereas it disappeared from HL-60 cells stimulated with phorbol esters to differentiate into mature macrophages in a resting stage. From this, they concluded that Ki-67 is directed against a nuclear antigen associated with cellular proliferation (Gerdes *et al*, 1983). A further study carried out in 1984 by Gerdes *et al* consistently showed that the Ki-67 nuclear antigen is present in S, G_2 and M phase but is absent in G_0 (Gerdes *et al*, 1983).

The key marker of senescence is the senescence-associated β-galactosidase (SA-β-gal) stain and involves the staining of cells with the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at pH 6.0. At pH 6.0, there is sufficient suppression of lysosomal β-galactosidase activity to ensure that non-senescent cells remain unstained (Gary & Kindell, 2005).

1.3 Cancer

The earliest recorded references to what later became known as cancer was found in the *Ebers Papyrus* written in Egypt in about 1500 B.C. which is a summary of medical knowledge known at that time. By the fourth century, approximately 1000 years later, Greek physicians, under the leadership of Hippocrates, had become familiar with cancers of the breast, stomach and uterus. It was at this time that Hippocrates assigned the name 'cancer'. It is not known exactly why the term cancer was applied, however, it is thought to be due to cancers of the breast resembling a crab with the claw-like processes buried or that the term was used to describe pain similar to that experienced from the bite or a crab. Another 500 years passed before the last physician of antiquity, Galen, carried the problem of cancer to the next logical step to develop a remarkable classification, in which he called tumours according to their nature, those being malignant and benign (Braun, 1977).

Cancer is often referred to as the 'failure of medicine' and is understandable when it was estimated that in 2002 there were 10.9 million cases of cancer worldwide with 24.6 million people living with cancer (within 5 years of diagnosis). In the same year, there were 6.7 million deaths from cancer (Parkin *et al*, 2005). Cancer was responsible for 26% of all deaths in the UK in 2005 with the three most common being lung, colorectal and breast which accounted for 21.8%, 10.5%, and 8.1% of deaths from cancer, respectively (URL accessed 10/11/2008: http://publications.cancerresearchuk.org/WebRoot/crukstoredb/CRUK_PDFs/CSMORT06.pdf.)

1.3.1 Cancer and its Characteristics

Cancer is a complex family of diseases, and carcinogenesis (the process by which a normal cell is transformed into a cancer cell) is a complex, multi-step process resulting from the accumulation of errors in vital regulatory pathways. Cancer is initiated in a single cell, which then multiplies and acquires additional changes that give it a survival advantage over its neighbours. These altered cells then amplify and generate billions of cells eventually establishing a cancer. It takes time for the

billions of cells to accumulate and for the errors to generate and so as the life expectancy of the population increases, it is likely that the incidence of cancer will increase. The main characteristics of a cancer cell include uncontrolled proliferation, metastasis, insensitivity to growth-inhibitory signals, evasion of apoptosis and self-sufficiency. Cancer is a large group of approximately 100 or more diseases, varying in age of onset, rate of growth, state of cellular differentiation, invasiveness, metastatic potential, responsiveness to treatment and prognosis. In terms of molecular and cell biology cancer is caused by molecular defects in both cell function and alterations of a cells genes and so from this we can see that cancer is a disease of abnormal gene expression (Ruddon, 1995; Ullah & Aatif, 2008).

1.3.2 Hallmarks of Cancer

The discovery of mutations which produce oncogenes (dominant gain of function) and deactivate tumour suppressor genes (recessive loss of function) have revealed cancer to be a disease of the genome. Both classes of cancer genes have been identified through their alteration in human and animal cancer cells and by their elicitation of cancer phenotypes in experimental models. It is suggested that the search for the origin and treatment of cancer may continue over the next 25 years in the same fashion that it has for the last 25 years and will simply add further layers of complexity to an already complex scientific literature (Hanahan & Weinberg, 2000). However, it is anticipated that the science of cancer will merge into a logical science whereby the complexities of the disease will become more understandable in terms of a small number of underlying principles, and in fact this has already begun. Research over the last decade has revealed that a small number of molecular, biochemical and cellular traits are shared by most cancer types and are required for the transformation of a normal cell into a cancer cell. Carcinogenesis is a multistep process, and cancer cell genotypes are a manifestation of seven essential alterations in cell physiology that collectively dictate malignant growth (Renan, 1993). The acquired characteristics of cancer, according to Hanahan & Weinberg are shown in figure 1.5. They are self-

sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis, and genomic instability. Each of these physiologic changes represents the successful breaching of anti-cancer defence mechanisms present within cells. The multiplicity of defence mechanisms may explain why cancer is relatively rare during the average human lifetime (Hanahan & Weinberg, 2000).

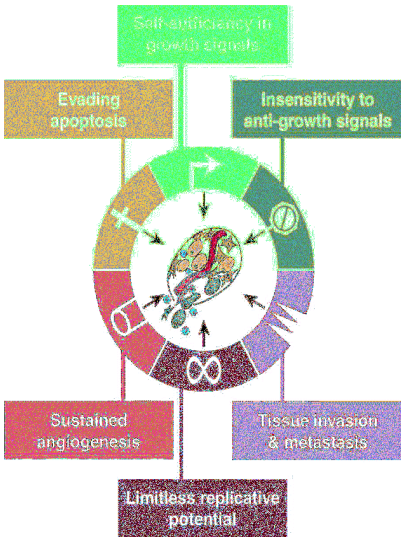


Figure 1.5 The acquired capabilities of cancer. According to Hanahan and Weinberg, all cancers have acquired capabilities shown above with each representing the successful breaching of anti-cancer defence mechanisms present within cells (Taken from Hanahan & Weinberg, 2000).

1.3.2.1 Self-Sufficiency

Under normal circumstances, cells require mitogenic growth signals (GS) before they moving from a quiescent state into an active proliferative state i.e. the cell cycle. Normal cells cannot proliferate in the absence of these stimulatory signals and many oncogenes are able to mimic the normal GS in one way or another (Hanahan & Weinberg, 2000).

Acquired GS autonomy was the first of the seven hallmarks of cancer to be identified and is mainly due to the prevalence of dominant oncogenes that have

been found. Three strategies that oncogenes use to achieve this GS autonomy are known and are as follows:

1. Alteration of extracellular growth signals
2. Alteration of transcellular transducers of these signals
3. Alteration of the intracellular circuits that translate these signals into action.

The manufacture of a GF by cancer cells obviates dependence of GFs from other cells within the tissue, for example, the production of platelet-derived growth factor (PDGF) and tumour growth factor α (TGFα) by glioblastomas and sarcomas, respectively (Fedi *et al*, 1997). Cancer cells also possess the ability to switch the types of extracellular matrix (ECM) they express and favour those that transmit progrowth signals (Lukashev & Werb, 1998).

Cell surface receptors which are responsible for the transduction of growth-stimulatory signals into the cell interior are often the targets of deregulation in tumour pathogenesis and GF receptors carrying tyrosine kinase activities in their cytoplasmic domains are overexpressed in many cancers (Hanahan & Weinberg, 2000).

The most complex mechanisms of acquired GS autonomy are derived from alterations in the components of the downstream cytoplasmic circuitry that receive and process the signals emitted by ligand-activated GF receptors and integrins. A central role is played here by the SOS-Ras-Raf-MAPK cascade and in approximately 25% of human tumours; Ras proteins are present in structurally altered forms that allow them to release a flux of mitogenic signals into cells in the absence of their normal upstream regulators (Medema & Bos, 1993).

1.3.2.2 Evasion of Apoptosis

The ability of cancer cells to expand in number is not only determined by the rate of cell proliferation, but also by the rate of cell attrition. Apoptosis, or programmed cell death, represents a major source of this attrition and acquired resistance to apoptosis is a hallmark of most types of cancer (Hanahan & Weinberg, 2000).

The apoptotic machinery of normal cells is divided into two classes: sensors and effectors. Sensors are responsible for monitoring the extracellular and intracellular environment for conditions of normality or abnormality that influence whether a cell should live or die. Effectors of apoptosis include an array of intracellular proteases termed 'caspases'. There are two 'gatekeeper' caspases – these are caspases 8 and 9, that are activated by death receptors such as FAS or by the cytochrome C released from mitochondria, respectively (Thornberry & Lazebnik, 1998). Resistance to apoptosis is acquired by cancer cells through a variety of strategies with the most commonly occurring loss of proapoptotic regulation being mutation of the *p53* tumour suppressor gene. The resulting functional inactivation is observed in more than 50% of human cancers and the result is the removal of a key component of the DNA damage sensor that induces the apoptotic effector cascade (Hanahan & Weinberg, 2000).

1.3.2.3 Insensitivity to Anti-Growth Signals

In normal cells, homeostasis and cellular quiescence are maintained through multiple anti-proliferative signals, which include both soluble growth inhibitors and immobilized inhibitors. Anti-growth signals are able to block proliferation in normal cells via two very different mechanisms:

1. Cells can be forced out of the active proliferative cycle into a quiescent state from which they re-emerge in the future when extracellular signals permit.
2. Cell can be induced to permanently relinquish their proliferative potential by being induced to enter into postmitotic states.

Imperfectly formed cancer cells must evade these anti-proliferative signals if they are to thrive (Hanahan & Weinberg, 2000).

1.3.2.4 Tissue Invasion and Metastasis

Cancer cells sooner or later spawn cells which move away from the primary tumour masses and invade adjacent tissues and from here 'travel' to distant sites and essentially form new colonies. These are known as metastases and are

responsible for 90% of deaths from cancer. Metastasis is a very complex process and represents the final frontier in exploratory cancer research (Hanahan & Weinberg, 2000).

1.3.2.5 Sustained Angiogenesis

Cell function and survival depends on oxygen and nutrients, supplied by the vasculature, which obliges all cells within a tissue to be within 100 μM of a capillary blood vessel. The formation of new blood vessels is via the process of angiogenesis and is a carefully regulated and transitory process. The ability to induce and sustain angiogenesis is acquired in a discrete step during tumour development and is via an 'angiogenic switch' from vascular quiescence. Without angiogenesis, tumours cannot survive and therefore it offers a uniquely attractive therapeutic target (Hanahan & Folkman, 1996; Hanahan & Weinberg, 2000).

1.3.2.6 Limitless Replicative Potential

The three acquired capabilities of growth signal autonomy, insensitivity to anti-growth signals and resistance to apoptosis lead a cell to be uncoupled from signals within its environment (Hanahan & Weinberg, 2000).

Hayflick demonstrated that cells in culture have a finite replicative potential and in certain cell populations, once they have progressed through a certain number of doublings, they stop growing via a process called senescence (Hayflick, 1997). The 'counting' device present within cells has been discovered at the ends of chromosomes- telomeres. Telomeres are composed of several of thousands of repeats of a short 6bp tandem repeat, -TTAGGG. The generations of replication are counted by the loss of 50-100 bp of the telomeric DNA during each cell cycle. The shortening is attributed to the inability of DNA polymerases to completely replicate the 3' ends of chromosomal DNA during each S phase and is termed the 'end-replication problem' (Counter *et al*, 1992). When telomeres are short, cells stop dividing and undergo growth arrest, or senescence. The 'counting' device and ability of the cells to become senescent is a powerful tumour suppressor mechanism as senescence occurs before the telomere length becomes short enough

to compromise chromosomal integrity (Campisi, 2001; Reddel, 2003). Cellular senescence is the ultimate irreversible loss of replicative capacity occurring in somatic cell culture (von Zglinicki *et al*, 2005). Almost all cancer cells are immortal as they can overcome senescence by either reactivation or upregulation of telomerase. Telomerase is a reverse transcriptase enzyme involved in the maintenance of telomeres (Shay *et al*, 2001). Telomere maintenance is evident in nearly all malignant cells, hence the ability to replicate uncontrollably as shown in figure 1.6.

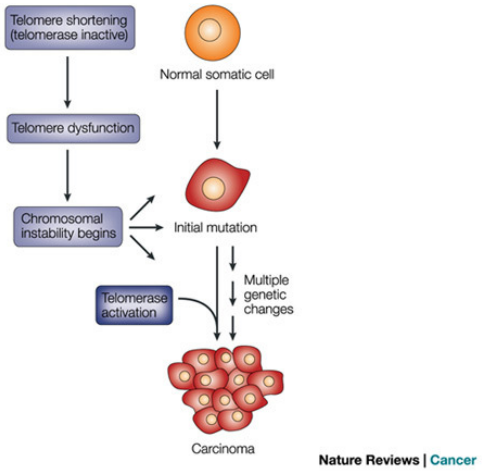


Figure 1.6 Limitless replicative potential. Upon division, normal somatic cells lose telomeric repeats in the absence of telomerase and eventually become dysfunctional leading to chromosomal instability. This instability leads to the formation of a carcinoma upon which telomerase activation occurs late in tumorigenesis (Taken from Feldser *et al*, 2003).

1.3.2.7 Genomic Instability

The final part in the production of a cancer cell is genomic instability, which is the progressive destabilisation of the cancer cell's genome. The increasing frequency of mutational events accompanying tumour progression provides a driving force for several of the hallmarks of malignancy as well as the ability of cancers to become resistant to therapy (Sledge & Miller, 2003).

Whilst the hallmarks of cancer, reviewed in sections 1.3.2.1 – 1.3.2.7, are considered important phenotypic markers, other important concepts of the role of pluripotent stem cells and gap junctional intercellular communication (GJIC) should also be considered when analysing carcinogenesis. The hypothesis which integrates these hallmarks shows that stem cells and their early progenitor cell are the target cells for the initiation event. These cells are immortal and only become mortal when they are induced to terminally differentiate thereby losing their telomerase activity. These two cell types are suppressed by GJIC and when permanent down-regulation of GJIC occurs, the cell enters the progression phase thereby contradicting the current paradigm that the first phase of carcinogenesis is the immortalisation of a normal cell followed by its neoplastic transformation and are considered the 'ignored' hallmarks of carcinogenesis (Trosko *et al*, 2004).

1.3.3 Cancer Epidemiology

Cancer is the leading cause of death in the Western world and in the US alone, over 1 million new cases of cancer occur annually. Cancer is an ancient condition yet despite this ancient lineage, the modern components of longevity and lifestyle affect both the number of and type of cancers encountered. Changes in lifestyle have impacted on the prevalence of certain cancers with breast and prostate cancer being the most prevalent cancers in Western countries and stomach cancer being the most prevalent cancer in China. It is possible, however, that these geographical differences are due to more than genetic differences between the races (King, 2000). Cancer can occur at any time, though the average age of onset is 67 years. Although this is true, cancer, particularly leukaemia, lymphoma and Wilms' tumour, are a leading cause of death in children aged up to 14 years. In the 1940's, 25% of people diagnosed with cancer lived at least 5 years following treatment, increasing to 40% by the 1990's meaning that almost 85 000 people are now surviving who would have possibly died within 5 years in the 1940's and so illustrating that the diagnosis, treatment, and knowledge of cancer is continuously improving (Ruddon, 1995).

1.3.4 Cancer Genomics

Cancer is a disease of the genome and in multiple sites of the cancer cell; it is the genome which is altered. The goal of cancer research is to define these molecular defects and turn any discoveries into effective treatment and prevention regimes. Until relatively recently these efforts relied on laborious and limited one-gene-at-a-time approaches (Weber, 2002).

In cancer, different genes are altered and in general, cancer arises from the accumulation of mutations in a series of genes over time. Despite this being widely known, not one human cancer has yet been completely described with a clear picture of all the genes altered within. To further add to the problem, even a complete profile of the mutations in any given cancer would represent only a single point in time (Weber, 2002).

1.4 Systems Biology

It is suggested that systems biology is nothing new or that it is simply irrelevant and is therefore not often appreciated. The central paradigm of systems biology is that much of biological function arises from the interactions of macromolecules. This area of science attempts to analyse, with new and improved reincarnations of existing and new technologies, scientific issues that already existing scientific disciplines are able to describe but not resolve (Westerhoff & Alberghina, 2005). So what is systems biology? There is new excitement about systems biology, and with that is an understandable interest in its definition. Scientific fields, like species, arise by descent with modification, so in their earlier forms even the founders of great dynasties are only marginally different than their sister fields and species and it is in retrospect that we can recognize the significant founding events (Kirschner, 2005).

Systems level approaches to biology are not new, however, the foundations are being achieved only now at the beginning of the 21st Century and this renewed interest in a systems level approach is linked to the progress in collecting experimental data and to the limits of the ‘reductionist’ approach and also the

requirement of system level understanding of native biological and pathological systems in providing potential therapeutic targets (Friboulet & Thomas, 2005).
The emergence of a systems biology approach has 3 possible impacts:

1. A systems level understanding of native biological systems (animals, plants and microorganisms) with not only system structures but also system dynamics, including metabolic analysis and sensitivity analysis.

2. A system level understanding of pathology and malfunction in order to control the state from cell to the whole body and provide therapeutic targets for treatment of diseases.

3. The development of a system level approach in biotechnology to design biological systems having the desired properties which don't exist in nature (Friboulet & Thomas, 2005).

1.4.1 The Meaning of Systems Biology

It has been said that there are as many definitions of systems biology as there are research institutes in the world (Beltrao *et al*, 2007). The definition of systems biology is dependent on the person you ask, but in general there are two groups of people doing, and talking about systems biology. The first group are the panomicists who are researchers that combine different high-throughput data, such as transcriptosomes, to formulate and test hypotheses about the components and connectivity of biological networks. The second group are the dynamicists, who apply principles of systems theory to the modelling and testing of biological system dynamics. In addition to these two groups, there is a great deal of interest in developing tools to design and control biological systems (Levesque & Benfey, 2004). The definition of systems biology varies considerably according to the expertise of the person in question and the lack of a clear definition of systems biology is due to a number of factors, including:

1. Language- the language of systems biology needs to cross several scientific disciplines and is therefore, at the moment, not very well developed.

2. Multi-disciplinary- the practitioners of systems biology come from a multitude of disciplines in the scientific field and each has their own scientific

perspective. Typical practitioners of systems biology include analytical chemists, mathematicians, statisticians, biologists, technologists, computational biologists and informaticians.

3. Approach- systems biology can be approached in two different ways: the 'top-down' or 'bottom-up' approach. The 'top-down' approach aims to target analysis concerned with mechanism and is the more focused of the two approaches. The 'bottom-up' approach is concerned with a global attempt to identify and understand all the components within the system and how they interact (illustrated in figure 1.7). In reality, systems biology approaches appear to be more of a hybrid of the two approaches and are often dependent on the biological process in question (illustrated in figure 1.8) (Naylor & Cavanagh, 2004; Naylor *et al*, 2008). The 'top-down' approach to systems biology identifies molecular interaction networks on the basis of correlated molecular behaviour observed in genome wide 'omics' studies, whereas the 'bottom-up' approach examines the mechanisms through which functional properties arise in the interactions of known components (Bruggeman & Westerhoff, 2007).

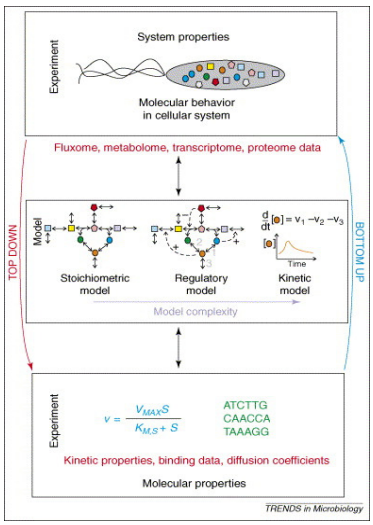


Figure 1.7 The 'top-down' and 'bottom-up' approach to systems biology. Approaching systems biology in a 'bottom-up' manner is to start with the molecular properties from which models are constructed and used to predict systemic properties. The models are validated by experiments and are then refined. A 'top-down' approach to systems biology tends to be data driven and so starts with the experimental data to be used to discover new models or refine existing models to describe the experimental data successfully. Both approaches are depicted here (Taken from Bruggeman & Westerhoff, 2007).

Systems biology is also described as 'integrative biology' or 'predictive biology'. Operationally, systems biology requires an ability to digitalize biological output so that it can be computed and then the computational power manipulated to analyze highly comprehensive and huge datasets. This data then needs to be integrated into a user-friendly format therefore allowing systems biology to be described as 'integrative biology'. The ultimate aim of 'integrative biology' is the ability to predict new biological outcomes given the list of components involved within a system. The description of a network as a precise functional model mapped by a control diagram and therefore advancing the prospects of forward engineering of a complex biological process and it is in this respect, systems biology can be described as 'predictive biology' (Liu, 2005).

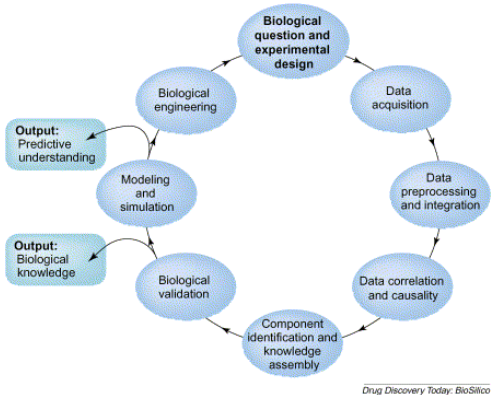


Figure 1.8 The process of systems biology. According to Naylor and Kavanagh, systems biology is a process consisting of numerous interchangeable modules. The order of the process is determined by the biological question that is being investigated (Taken from Naylor & Cavanagh, 2004)

1.4.2 Origin of Systems Biology

Systems biology is a well-established area of human activity that can be traced back at least to the writings of Aristotle around 350 B.C.; however, as stated previously the foundations are only being achieved now at the beginning of the 21st Century. When referring to systems biology, the term 'system' refers to an entity that represents wholeness, and can be superficially divided into components whose pivotal properties cannot be fully explained solely from knowledge of these

parts. The paradigms of life sciences were changing with time, but the core fundamental concepts of systems biology were apparently resistant to these changes and remain the same today as they were in the time of Aristotle (Konopka, 2007).

Over the past 150 years, there have been four major groups of theories which have contributed to the deciphering of biological information for all living forms;

1. 'The cell theory' led to the establishment of the cell as a basic unit of biological information for all living forms.
2. 'Theories of evolution' led to information regarding how biological information has changed over time.
3. 'Theories of hereditary' told us about how biological information is passed from one generation to the next.
4. 'Theories of biochemistry' told us how biological information is organized and structured in macromolecules, such as nucleic acids and proteins (Auffray *et al*, 2003).

The past 20 years has seen the emergence of the Human Genome Project and with it the knowledge of how biological information is organized on a global scale. The next step in the Human Genome Project is how to make sense of all this biological information, and it is proposed that the most sensible approach would be a systems biology approach. Systems biology will allow the study of all the elements of a system in response to both internal and external signals and allow the understanding of emergent properties (Auffray *et al*, 2003).

The cell cycle is one of the most frequently studied biological processes in systems biology and involves the knowledge of a large number of genes and protein interactions (Alfieri *et al*, 2007). With the discovery of how functions arise in dynamic interactions, systems biology aims to address the missing links between molecules and physiology (Bruggeman & Westerhoff, 2007).

1.4.3 Modern Day Systems Biology

Although systems biology has been well-established since 350 B.C., it is still considered a new field in the molecular life sciences and is as new now as

molecular biology was in the fifties and cell biology in the seventies. According to Westerhoff, systems biology is a science that aims to elucidate the general principles that govern the emergence of biological function from the interactions of living systems (Westerhoff, 2005).

Three emergences have driven systems biology:

1. The internet has made it possible to capture, store and transmit massive data sets and communicate on a global scale.
2. A cross-disciplinary approach has been essential for the emergence of systems biology and it is therefore necessary to integrate technology, biology and computation.
3. The Human Genome Project has had a major impact on biology and has transformed the way in which we view and practice this area of science (Hood, 2002).

Typically, how we view and practice biology has changed massively since the Human Genome Project and we have observed the emergence of biology as an informational science through the application of high-throughput genomic and proteomic platforms for global analyses leading to the term systems biology (Ideker *et al*, 2001).

1.4.4 The Challenges of Systems Biology

Systems biology approaches have long been taken in particular areas such as immunology, physiology, development and neurobiology and technological advances in the 1980s have led to the analysis of years of reductionist inquiries in the context of the entire system. For example, we now have microarray analysis which permits global transcriptional profiling and mass spectrometry advances allow large-scale proteomic and metabolomic analysis. Mountains of data is generated by these high-throughput methods leading to the rapid growth of the bioinformatics and computational biology areas of science. It is thought there are three concepts which are crucial to the understanding of systems biology, those being emergence, robustness and modularity (Aderem, 2005).

Technical challenges faced by systems biology are outlined below:

1. Network Biology
    - Network biology is an area of science still in its infancy.
    - The future needs of systems biology includes the development of new theoretical methods to characterize network topology to insights of biological function.
2. Data Quality
    - Systems biology approaches rely heavily on information available in public databases, which are often incomplete or not standardized and/or annotated properly.
    - The quality of the data is often not known with confidence.
2. Equipment
    - There is a need for sensitive tools to identify and quantify the concentrations, fluxes and interactions of various types of molecules at high resolutions in terms of both space and time.
    - The measurements made must also be in a context which is appropriate to specific networks, cells or organisms.
    - There is a need for equipment that is both miniaturized and automated to allow parallel multiparameter analysis which integrates operations such as cell sorting. It may also be necessary to quantify the forces and kinetics associated with protein/protein or protein/DNA interactions which would require nanomechanical or nanoelectric devices.
3. Imaging
    - Imaging will need to be extended to dynamic, spatial and multiparameter measurements within a single cell.
    - Hypotheses will not only need to be tested at the cellular level, but also at the whole animal level. This kind of testing would require there to be advances in molecular imaging ranging from bioluminescence and fluorescence to positron emission tomography (PET) and magnetic resonance imaging (MRI) (Aderem, 2005).

To conclude, systems biology is the analysis of a system by looking at the different levels of that particular system, for example, looking at the molecular level to the cellular level and so on. Systems biology depends on the simultaneous understanding of hundreds of processes within a system, and the human mind is only able to deal with a few processes and so it is understandable that help is needed via computation and modelling (Westerhoff, 2005).

1.4.5 The Future of Systems Biology

The future of systems biology relies on the further development of theoretical frameworks and tools for integrating the various levels of biological data, displaying them graphically and mathematical modelling and simulations of this data. There is also a requirement for the education of cross-disciplinary scientists whom should have a deep understanding of biology. The solution would be to teach biology as an informational science. The other major challenge facing both academia and industry and the potential future of systems biology is that of being able to integrate technology, biology and computation (Ideker *et al*, 2001).

Systems biology reflects a true paradigm shift of thinking how biology is investigated and reflects the arrival of 'big biology'. For systems biology to be sustainable and 'here to stay' there is the need for a new generation of specially trained scientists in the systems biology field. We also need to improve data sharing and encourage cooperation and whilst there is this new arrival of 'big-biology', the old-fashioned, hypothesis-driven biology is still alive and well and the data that already exists from this approach can be integrated into new data sets derived from systems biology experiments. Systems biology is able to provide us with new and dynamic ways of approaching science, and simply now needs to be exploited (Naylor & Cavanagh, 2004).

1.4.6 Cancer as a Systems Biology Disease

Cancer research has focused on the identification of molecular differences between cancerous and healthy cells and the emerging picture is overwhelmingly complex. During the evolutionary process of malignant transformation, cancer cells

accumulate multiple genetic alterations providing them with several capabilities, such as the ability to escape normal growth control, evading the suicidal apoptotic program, the induction of sustained angiogenesis, the ability to metastasize and invade healthy tissues (Cahill *et al*, 1999). A central aim of cancer research has been the identification of the mutated genes that are implicated in cancer, the so-called 'cancer genes'. A 'census' of these cancer genes indicated that mutations in more than 1% of genes contributed to human cancer (Futreal *et al*, 2004). The vast majority of genes that have been identified either function in signal transduction processes within or between cells or govern cell cycle progression, apoptosis and infiltration (Vogelstein & Kinzler, 2004). Cancer is a disease involving several levels and if this is true, how can the effect of a mutation in a particular gene on the functionality of the entire system be predicted, and as a consequence, how can the identification of drug targets progress? It has been argued by Hornberg *et al* that it is crucial to recognize cancer as a systems biology disease which would lead to more cancer biologists becoming systems biologists and provide answers to some of the systemic questions regarding cancer. On the other side, applying systems biology to drug discovery could lead to rationalized therapies (Hornberg *et al*, 2006). To practice systems biology, global sets of biological data must be captured and integrated from as many hierarchical levels of information as possible. Types of biological data include DNA sequences, protein measurements and interactions. The data is then analysed and annotated, however, there is a need for the formulation of detailed graphical mathematical models. These models are formulated and then refined on the basis of hypothesis-driven, iterative systems perturbations and data integration (Aderem, 2005). It is clear therefore that systems biology leads to mathematical modelling and in the next section, I will go into more detail about this area

1.5 Mathematical Modelling

A mathematical model is an abstract model using mathematical language to describe the behaviour of a system. Eykhoff in 1974 described mathematical models as a 'representation of the essential aspects of an existing system (or

systems to be constructed) which presents knowledge of that system in usable form.' Mathematical models of biochemical control systems attempt to derive observed physiological properties of cells from the underlying molecular regulatory networks. For example, it is possible to model cell growth and division, chemotaxis, secretion and circadian rhythms. The aim of modelling such systems is that by creating them there will be a greater understanding of the biological processes involved (Allen *et al*, 2003).

In addition to mathematical modelling there is the area of computational modelling. A major objective of computational modelling is to model the kinetics of entire living cells at the molecular mechanistic level. If developed further, such models would represent the most compact, unambiguous and unified form of biological hypotheses, and as such could then be used to quantitatively explore inter-relationships at both the molecular and cellular levels (Morgan *et al*, 2004). They also provide useful insights into the mechanisms of healthy and diseased cells and give a better understanding of how system-level or whole-cell properties emerge from intracellular interactions of molecular components (Lindon *et al*, 2004).

1.5.1 Modelling the Cell Cycle

The cell cycle is a fundamental process of cell biology and a detailed understanding of its function, regulation and other underlying mechanisms is critical to many applications in both biotechnology and medicine (Fuß *et al*, 2005). As stated before, mathematical models use abstract language to describe a system. Models which describe the events of cell growth and division are useful in understanding the functional control mechanisms. Models are also useful for the theoretical support of the automated analysis of flow cytometric data and of cell volume distributions (Alberghina *et al*, 1986). The cell cycle is a fundamental phenomenon of life of which irregularities are the one common factor in the manifestation of human cancer and investigations at the molecular level show that the cycle is strictly linked with oscillations in the concentrations of a limited number of substances such as cyclins, maturation-promoting factor (MPF) and

Cdc2. Mathematical modelling of the cell cycle in terms of nonlinear models of the interaction between these substances started in the early nineties (d'Onofrio, 2006).

1.5.1.1 Mathematical Modelling to Describe the Cell Cycle

Mathematical modelling of the cell cycle has generally followed the current, existing knowledge of the cell cycle as opposed to being involved in driving it (Clyde *et al*, 2006). Before the 1980's, there was very little known with respect to the biochemistry or the physical processes of the proteins involved in the cell cycle. A view commonly taken by embryologists and physiologists and with the knowledge of the cell cycle being a sequential process of growth and division has resulted in the idea that the cell cycle should be modelled as an oscillating system which is driven by size and mass. This view of the cell cycle is often referred to as the 'clock theory'. Another view is that held by geneticists whom consider the cell cycle as a series of independent reactions whereby the completion of each reaction is dependent upon the previous one and is often referred to the 'domino theory'. In 1989, Murray & Kirschner proposed a unified view of both the 'clock theory' and 'domino theory' and was based on the idea that the key components of the cell cycle regardless of being embryonic or somatic, each appeared to be the products of homologous genes (Murray & Kirschner, 1989).

One of the early models of the cell cycle was proposed by Tyson in 1991 and depicted cellular physiology in terms of molecular concentrations. This model was a series of differential equations to describe the concentrations of the proteins, cdc2 and cyclin, which form a heterodimer that control the major events of the cell cycle. The differential equations were used to describe the unification of cdc2 and cyclin to form complexes and modify their functions through both phosphorylation and dephosphorylation. The model shows that it can operate in three modes; a steady state, a spontaneous oscillator or as an excitable switch (Tyson, 1991). Further models developed in the 1990's by Thron provided a series of models of the mitotic clock, in addition to the dynamics of the cyclin B-MPF and the

activation of a cell cycle kinase which is involved in the down-regulation of its own inhibitor (Clyde *et al*, 2006).

It is important to ensure that the mathematical models that are developed are representative of the cell cycle in the most accurate way possible. Current knowledge of the cell cycle clearly shows that the cell cycle is a complex process involving the interplay of a number of regulatory proteins that drive cells through a specific sequence of events culminating in mitosis and the production of two daughter cells (Schafer, 1998). As a result of this, a number of models have been developed which describe the transitions of the cell cycle, in particular the G_1/S and the G_2/M transition points. The G_1/S transition point has been modelled by Hatzimanikatis *et al* (1999) and Thron (1999). Hatzimanikatis *et al* described the G_1/S transition point as a limit cycle, whereas the model developed by Thron simulates the reactions between the heterodimeric kinase cyclin E: Cdk2 and the inhibitor p27 (Hatzimanikatis *et al*, 1999; Thron, 1999).

The G_2/M transition has also been extensively modelled by Thron and in one example, the activation of the MPF, which is involved in starting mitosis, is modelled as a biochemical switch or trigger. The model proposed by Thron (1996) does not allow for cyclin biosynthesis or degradation and there is only one state variable, that being M, and there no possible oscillations.

Progress on modelling the spatial considerations of the cells has already started as depicted by a model proposed by Meinhardt & de Boer (2001). The Min proteins in *Escherichia coli* undergo a highly dynamic localization cycle, during which they oscillate between the membrane of both cell halves. The model proposed here shows the Min protein dynamics using pattern formation, resulting from the activator/depleted substrate mechanisms in *Escherichia coli.* Yang *et al* (2006) have also developed a mathematical model that not only describes the regulatory proteins of the cell cycle on the basis of time, but also considers protein translocations between the cytoplasm and the nucleus.

Many of the cell cycle models that have been produced to date are often simplistic and are so to avoid issues such as considerations of space and the changing shape and volume of the cells. In addition to this, there is a neglect of the high level of

complexity involved in signal-transduction pathways within the cells and so there is the need for these to be incorporate in future mathematical models of the cell cycle.

1.5.1.2 What has Mathematical Modelling Delivered?

Mathematical models have not yet reached a sufficient complexity that is required to be effectively involved with clinical research of diseases relevant to the cell cycle. Mathematical modelling of the cell cycle has been presented by a number of authors and some have been more optimistic about the progress of mathematical modelling than others. Ingolia & Murray (2004) offer a pessimistic view of mathematical modelling since they conclude that whilst there have been successes, generally, mathematical modelling has added little understanding to the biological processes of the cell cycle. Ingolia & Murray conclude that mathematical modelling is required to stimulate experiments, rather than simulate them. Additionally, since mathematical models at present are derived with the intention of being interpreted qualitatively, they cannot be proven to be right or wrong in quantitative terms (Clyde *et al*, 2006). The basis of a successful model is the source and quality of the data, in addition to the format of the data (Davidoff *et al*, 2003).

Mathematical modelling has the potential to assist in the research of potential drug treatments for certain types of cancers. The networks involved in cancer are highly complex and nonlinear and so models of the protein-level dynamics will highly assist in the investigation of the effects of potential drugs or combinations of drugs. Models could also be used in the determination of the best drug regimes in the treatment of specific cancers. For accuracy, the models used for this would have to be not only highly comprehensive but also quantitavely accurate and at present this has not been achieved (Clyde *et al*, 2006).

The derivation and organization of data remains a substantial task, but is considered to be smaller than that of repeating the large number of experiments to derive numerical data. It is considered that gaps will remain and future modelling must accommodate the uncertainty and so future models should be optimized

against the available data to infer the properties and associated parameters in the regions of the network where data is missing (Clyde *et al*, 2006). Both mathematics and computation have played critical roles in the understanding of complex systems, particularly in physics and chemistry. This approach is now entering the field of biology and it is suggested that biology has not yet met its Einstein, however, some important beginnings have been made, for example, mathematical frameworks for the characterization of metabolic systems. There are also certain mathematical models of certain molecular subsystems involved in DNA replication and cell cycle control and regulation. The mathematical models that have already been developed are useful in playing a central role in the testing of hypotheses, however, at present are inferior to other models that discuss how individual components work together to enable the system to work (Bailey, 1999). Davidoff *et al*, 2003 suggest that advances in mathematical modelling tools and computer simulations have already positioned *in silico* biology to capitalize on technological breakthroughs across all sectors of biomedical research, from target selection to the field of personalized medicine.

It is obvious that mathematical models represent an accurate tool for the research of potential drug treatments, however, what is lacking is a unified approach to the problem and an organizational overlay which would ensure that present and future research is directed efficiently into a structured format that is suitable for modelling (Clyde *et al*, 2006). The future of mathematical modelling in biology is dependent on the greater co-operation between cell biologists and theoreticians and although this has been realised in the past, there has been very little change and may only change when cell biology begins delivering in a manner that other sciences already do. The first step to this approach requires the knowledge, in quantitative terms, the level of proteins involved in the cell cycle and how they are altered by varying anti-cancer drug treatment regimes. Kholodenko (2003) indicates that this is no small task and requires considerable resources, particularly in the form of teams of theoreticians and biologists. Any mathematical model describing the cell cycle must be founded on the underlying biology (Clyde *et al*,

2006). To highlight mathematical models of the cell cycle and where these are at present, in the next section I will present a few examples.

1.5.1.3 Examples of Mathematical Models

In 2004, Basse *et* al developed a model of a population of cells and the effects of cancer therapy on this population of cells. The model began with a theoretical one compartment size structures cell population model and was used to investigate its asymptotic steady size distributions (SSDs) on plankton. This model lacked accuracy, however, due to the differences in size distribution between plankton and DNA size distributions observed via flow cytometry of human tumour cell lines. The models analysed the density of cells in each phase of the cell cycle over time, which was differentiated by their DNA content and based on distinct phases of the cell cycle and the transition of cells between these phases. From this first model, a second model was devised for a cell line perturbed by paclitaxel, with this second obtained from the first with the cell division parameter set to zero and the inclusion of a possible transition of the cells from the M- phase to a defined A-phase in which cellular DNA decreases with time (Basse *et al*, 2004).

Carrying on from this model, Basse *et al* in 2005 proposed three alternative formulations of a mathematical model for human tumour cell lines unperturbed by cancer therapy. The model for a cell line unperturbed by cancer treatment is based on the distinct phases of the cell cycle as described in section 1.2.2 and the transitions of cells between those phases as shown in figure 1.9. The model is a system of ordinary differential equations (ODEs) and the two alternative formulations give a temporal ODE and partial differential equation (PDE) system that allow the number of cells in each phase over time to be analysed without any knowledge of DNA distribution, which is required in the first formulation. The three formulations derived provide a link between many of the examples of cell population models. The basis of each formulation is a simple mathematical model of the cell cycle differentiated by the phases G_1, S, G_2 and M with rates of transitions between the phases. The models allowed Basse *et al* to gain estimates of the DNA distribution over time in which they found the distribution eventually

settled into steady DNA distribution (SDD) which can then be compared with experimental DNA distributions thereby providing the starting point for cancer therapy experiments and could be considered a 'four-compartment model' (Basse *et al*, 2005).

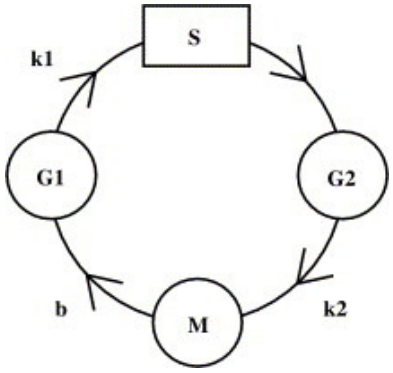


Figure 1.9 The phase transition model proposed by Basse *et al* (2005). The model proposed by Basse *et al* decomposes the cell cycle into its four distinct phases, G_1, S (synthesis), G_2 and M (mitosis). It was proposed that k_1 is the hourly rate of transition of cells between G_1 and S whilst k_2 is the hourly rate of transistion of cells between G_2 and M. The rate of cell division is with rate b per hour from M to G_1. Cells are assumed to spend a fixed time in S phase (Taken from Basse *et al*, 2005).

Another model developed by Basse *et al* is based on DNA content. In this model the transitions between the phases were measured and the model is validated using flow cytometric methods. This model represents an approach to the human tumour cell cycle that is considered to extend and compliment previous models and could be considered a 'three-compartment model' (Basse *et al*, 2005).

In addition to modelling cell cycle phase, it is also important to model the controls and checkpoints of the cell. The eukaryotic cell cycle is regulated by three modules that control G_1/S, G_2/M and meta/anaphase transitions. If a cell cycle cannot be completed then checkpoint mechanisms either delay or completely block further cell-cycle progression. By writing a series of ODEs/PDEs for the time rate of change of the components of a proposed network, the network can be transformed into a mathematical model. These equations can be used to find all the possible characteristic states of the controls of that system (Novak & Tyson, 2003). The different approaches to mathematical modelling are illustrated in figure 1.10.

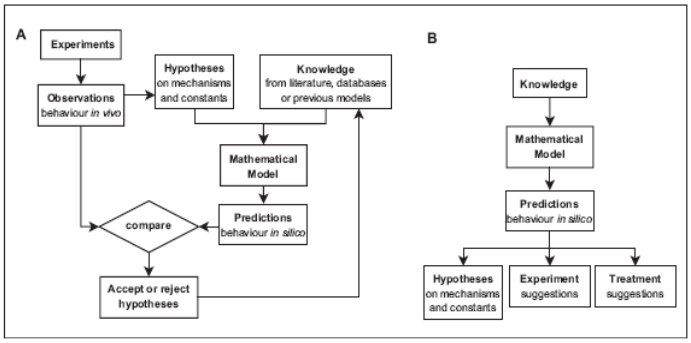


Figure 1.10 Mathematical modelling strategies. With respect to mathematical modelling there are two main strategies, explanatory modelling as shown in A and predictive modelling as shown in B. Explanatory modelling uses experimental observations and the model's predictions (possibly those of several alternative models) are compared in order to confirm or reject the hypotheses of the mathematical model. Using predictive modelling is to conduct artificial experiments, i.e. *in silico* to suggest novel hypotheses or be used to support decisions on therapeutic strategies (Taken from Fuß *et al*, 2005).

The cell theory implies that cells grow in size and divide in order for the formation of more cells with the majority of cells dividing at the same rate at which they double in mass. This therefore preserves the average cell mass that is constant over many cell cycles (Ingolia & Murray, 2004). The regulation of cell growth is crucial in the development of both plants and animals and attempts to mathematically model the cell cycle regulation have been carried out over the past 30 years. Earlier attempts regarded the machinery that controlled the cell cycle as a black box and mathematical models were simply attempts to produce a mathematical description of the rules that were being obeyed (Ingolia & Murray, 2004).

Like many applications, studies of single cells also have their problems. One of the main problems is allowing the cells to show 'balanced growth' which often show distortions in the cell cycle. Cells also need a continuous supply of fresh medium for oxygenation (Mitchison, 2005). Work in physics and engineering suggest that mathematical models are powerful tools and would therefore be useful in understanding the cell cycle. Although cell cycle machinery is oscillatory, its main function is to ensure the orderly progression of the cell cycle and production of a new daughter cell and it differs from other oscillatory systems. Cell growth is continuous and must be properly coupled with cell proliferation and the production

of chromosomes and centrosomes. This means the cell proceeds through a repeating sequence of discrete states and this highly regulated and abrupt system is in contrast to many other oscillatory systems. One of the main goals in mathematically modelling the cell cycle is to verify that known interactions in some systems can produce the observed qualitative behaviour. Mathematical models employ ODEs in which concentrations of different molecules are dynamic variables and must all be taken into account when modelling the cell cycle (Ingolia & Murray, 2004).

An excellent example of a mathematical model which is specifically useful to the cell cycle is presented by Aguda, 2001, in which the pathways involved in the G_1 phase of the cell cycle are described. This model was followed by that of Sear & Nevins, 2002, whom summarize to role of the Ras pathway in E2F and Myc transcription.

A paper by Lupi *et al*, 2004 describes their approach to the analysis of the cytotoxic and cytostatic effects of topotecan (TPT) by a novel mathematical simulation approach. TPT, a derivative of camptothecin, is an anti-cancer drug widely used for the treatment of human solid malignancies, including ovarian cancer and small cell lung cancer. The mechanism of action of camptothecin is noncovalently binding to the normally transient DNA-topoisomerase I cleavable complex and in doing so, inhibiting the relegation step of the enzyme resulting in the accumulation of protein-linked ss breaks in the DNA (Lupi *et al*, 2004). The aim of this study was to derive a complete analysis of the effects of TPT on a cell line by the analysis of the experimental data, that being flow cytometric percentages and absolute cell number, by a mathematical model in terms of DNA synthesis, G_1, G_2/M and death or recycling. Each parameter of this mathematical model is viewed as a quantification of the activity of a specific molecular network (Lupi *et al*, 2004).

Another excellent example of a mathematical model is that by Lupi *et al,* 2007 in which the aim is to mathematically interpret the role of p53 in the dynamics of cell cycle response to DNA damage. p53 was the first tumour-suppressor gene to be described and upon DNA damage of aberrant growth signals, there is the

activation of p53 and this can occur in a number of ways, for example, many different kinase families phosphorylate p53, including MAP kinase and the casein kinase family. In addition, there is ATM and ATR, which are the two major upstream regulators of the p53 response to DNA damage (Lupi *et al*, 2007). The response of a cell line to a drug is a mixture of both cytostatic and cytotoxic effects within each phase of the cell cycle and is dependent on time and dose. Studies carried out on the p53 inactivation on the response of cancer cells are controversial with some studies indicating there is enhanced sensitivity and other suggesting there is in fact enhanced resistance in studies using the same compounds. In their study, HCT116 cells are exponentially grown in cisplatin and doxorubicin to obtain monoparametric DNA histograms, biparametric DNA/bromodeoxyuridine (BrdU) dot plots, detection of apoptosis by the TUNEL assay and the biparametric analysis of p53 level in relation to DNA content. The approach used by Lupi *et al* is very similar to the approach used in this study since they model the cell cycle as a series of compartments based on cell cycle data.

The major novelty of this study is the long term nature of the modelling in comparison to other studies of this nature. This is challenging since most studies are carried out up to 72 hours and this is addressed in this study. The data presented by Lupi *et al*, 2007 was carried out up to 72 hours and in 2004 was carried out up to 96 hours. Basse *et al*, 2004 carried out their study up to 50 hours. The present study is carried out up to 21 days.

1.6 Telomerase Inhibitors

The unique biology of telomeres and telomerase plays an important role in many aspects of mammalian cell physiology such as cell ageing and death. Telomeres are nucleoprotein structures found at that the ends of eukaryotic chromosomes, as shown in figure 1.11, and one of their primary functions is to mark the ends of linear chromosomes as distinct from that of a broken DNA end and facilitate in chromosome replication (Blackburn, 2001). Telomeric DNA is mostly composed of 2 to 30 kb of double-stranded $(TTAGGG)_n$ repeats that are necessary for their function (Yang *et* al, 2005).

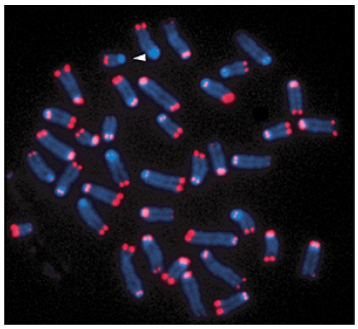


Figure 1.11 The location of telomeres. Telomeres are nucleoprotein structures found at the ends of eukaryotic chromosomes as shown above are stained red. The telomeres are important for maintaining chromosome stability (Taken from Bonetta, 2001).

In general, human cells will senesce or die as the average telomere length approaches 1-4 kb (Li *et al*, 1998). The telomeres are characterized by their 3'-overhang of the G-rich strand. The current model suggests that telomeres form a higher-order structure that hides the 3'-overhang and therefore protecting the chromosome ends from activities in the cell. The higher-order structure is formed from the ability of the 3'-overhang to fold back and essentially invades the double-stranded region of the telomere to form what is known as the 'T-loop' as shown in figure 1.13. The T-loop is held together by seven known proteins: TRF1, TRF2, POT1, TIN1, TIN2, TANK1 and TANK2, collectively referred to as the shelterin complex and are discussed in more detail in section 1.6.1. In forming this T-loop, a displacement loop is formed, known as a 'D-loop' (Blasco, 2003). The telomeres are maintained by a specialized reverse transcriptase, the ribonucleoprotein telomerase, as shown in figure 1.12. Telomerase is composed of a ubiquitously expressed RNA subunit, Telomerase RNA Component (TERC) and a protein catalytic subunit, Telomerase Reverse Transcriptase (TERT), whose expression is highly regulated (Masutomi & Hahn, 2003).

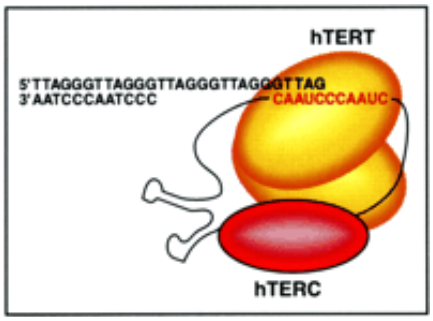


Figure 1.12 Telomerase at the telomere. Telomerase is composed of a RNA subunit, TERC, shown here in red and a protein catalytic subunit, TERT, shown here in yellow. The role of telomerase is to maintain the telomeres (Taken from Hahn, 2003).

It was determined in 1990 that in human somatic cells, telomeres shorten by 5-20 repeats with every cell division (von Zglinicki, 2003). The shortening of the telomeres is due to the 'end-replication' problem, i.e. the inability of DNA polymerase to fully replicate the 3' end of linear molecules during discontinuous lagging strand synthesis (Harley *et al*, 1990).

Without telomeres, chromosomes undergo end-to-end fusions to form dicentric and multicentric chromosomes which would break during mitosis, causing severe damage to the genome and activating DNA damage checkpoints leading to cellular senescence or the initiation of apoptotic cell death pathways (Rezler *et al*, 2003). The type of damage induced is determined by the level of loss of telomeres (Lechel *et al*, 2005). It is important that the ends of telomeres are maintained and in approximately 15% of tumours, telomerase activity is absent and so may allow the cells to avoid senescence and the telomere lengthening is instead obtained by recombination events, such as alternative lengthening of telomeres (ALT)[1]. The lack of telomerase gene expression in ALT cells is associated with chromatin remodelling and the methylation of the TERC and TERT gene promoters (De Cian *et al*, 2008).

[1]Immortal human cells maintain telomere length by expression of telomerase or the ALT mechanism. ALT involves a recombination-like process allowing rapid elongation of the shortened telomeres. It is not known if the ALT pathway affects other sequences in the genome (Jeyapalan *et al*, 2005).

Genetics studies have identified that there are at least three factors essential for telomere protection:

1. A minimum length of TTAGGG repeats to prevent telomere exhaustion and chromosomal instability.
2. Telomere binding proteins as discussed in section 1.6.1.
3. The integrity of the 3'- overhang (Blasco, 2002).

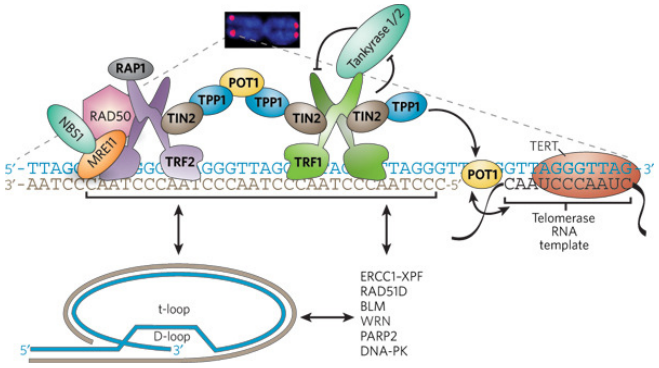


Figure 1.13 The location of telomeres and telomere binding proteins within a chromosome. Telomeres contain TTAGGG repeats with a single-stranded 3' overhang of the G-rich strand. POT1 is 'protection of telomeres 1' protein and is a negative regulator of telomerase activity and is involved in telomere maintenance (Taken from Verdun & Karlseeder, 2007).

1.6.1 Telomere Binding Proteins

One of the central characteristics of cancer is its complexity and the activation of telomerase is strongly correlated with this complex disease implicating telomerase, telomeres and telomere maintenance as a critical step in cancer development (Masutomi & Hahn, 2003). In addition to TERC and TERT, more than 30 proteins have been identified and proposed to be associated with the telomerase complex as shown in table 1.1.

Table 1.1 Telomere binding proteins and their role

Telomere Binding Protein	Role
TRF1/2 (Telomere Binding Factor 1/2)	• Expressed in all human cell types and are associated with telomeric repeats throughout the cell cycle. • Involved in length regulation either directly or by interacting with other telomere binding proteins. • Bind specifically to ds telomeric DNA and the inhibition of TRF2 causes the immediate deprotection of the chromosome ends through activation of p53, the loss of the telomeric 3'-overhang and end-to-end chromosome fusion. • Is also part of the shelterin complex (Shay *et al*, 2001; Cesare & Reddel, 2008; Griffith *et al*, 1999).
POT1 (Protection of Telomeres 1)	• Involved in binding ss DNA and is part of the shelterin complex (Cesare & Reddel, 2008).
TIN2	• Forms a protein complex with POT1 and TRF1. • Modulates the activity of TRF1 by inhibiting tankyrase 1. Also involved in T-loop stabilization and the down-regulation of TIN2 results in telomerase dependent telomere lengthening. • Part of the shelterin complex (Colgin & Reddel, 2004; Smogorzewska & de Lange, 2004).
TPP1 (previously known as PIP1)	• TPP1 and POT1 form a complex with telomeric DNA that increases the activity of the telomerase core enzyme. • It is also possible that TPP1 could recruit telomerase and is involved in the shelterin complex. It is possible that TPP1 is the protein by which the telomere can switch between in the 'open' and 'closed' state. • Enhances the affinity of POT1 for telomeric ss DNA (Xin *et al*, 2007; De Cian *et al*, 2008).
Rap1	• Involved in DNA repair and modulates T-loop formation (Shay *et al*, 2001).
Tankyrase 1/2	• A poly (ADP-ribose) polymerase (PARP) involved in telomere length regulation in co-operation with TRF1. • Its closely related homolog is tankyrase 2 and they may functionally overlap (Hsiao & Smith, 2008).
RAD50/ MRE 11/NBS1	• Are part of a DNA damage response complex and is crucial for DNA end-joining. • Mutations of this complex leads to defects in telomere length maintenance (Granger *et al*, 2002; Lavelle *et al*, 2000).

Cohen *et al* demonstrated that the active telomerase complex is composed only of three components: TERT, TERC and dyskerin, each in two copies. The regulation of these protein levels are poorly understood, however, it is known that the transcriptional control of the TERT gene plays a key role in the regulation of telomerase (Cohen *et al*, 2007).

The shelterin complex has been identified in the protection of telomeres at their extremities and is composed of six proteins. Three of the proteins, TRF1, TRF2 and POT1 bind directly to the telomeric repeats, as shown in figure 1.12, whereas

the other three, TIN2 (binds TRF1 and TRF2), TPP1, known previously as PIP1 (binds TIN2 and POT1) and Rap1 (binds TRF2) (De Cian *et al*, 2008). It has been identified that POT1 acts as a telomerase dependent, positive regulator of telomere length (Colgin *et al*, 2003). It is also reported that the overexpression of TRF2 increases the rate of telomere shortening without increasing senescence and actually reduces the setpoint of senescence defined by telomere length. TRF2 was identified in protecting critically short telomeres and so highlights the importance of these proteins in telomere protection and maintenance (Karlseder *et al*, 2002). The shelterin complex, as shown in figure 1.14, appears to connect the duplex telomeric DNA and the 3'-G-strand-overhang and exists in two states known as a 'closed' state in which the 3'-end cannot be accessed by telomerase but is protected from degradation and the other is an 'open' state in which all the proteins are present and the 3' -end is accessible by telomerase. The third state, as shown in figure 1.14, is as yet unknown and would generate a DNA damage response (De Cian *et al*, 2008).

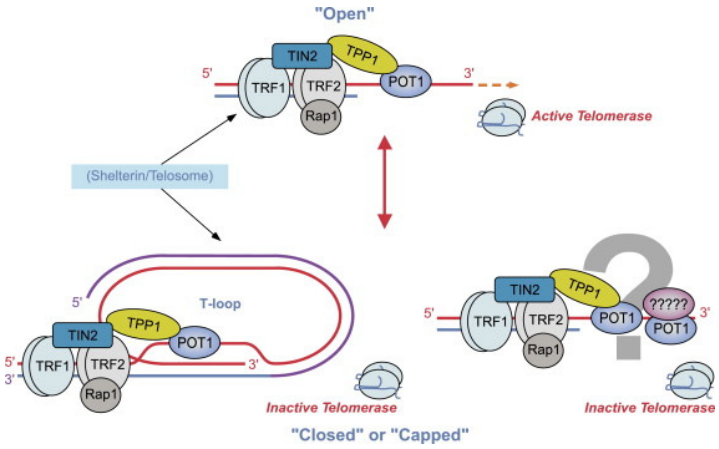


Figure 1.14 The shelterin complex. The 'open' state of the telomere with the associated proteins allows the binding of telomerase and therefore the elongation of the 3' end of the G-strand. The 'closed' state of the telomere and inactivation of telomerase is unknown and two mechanisms have been proposed. The first is the formation of a T-loop and the other is said to involve the capping of the 3' end, possibly by POT1, of the linear telomere (Taken from De Cian *et al*, 2008).

1.6.2 Telomerase Inhibitors as Anti-Tumour Agents

Current cancer research is focused on the identification of therapeutic agents that limit the exponential proliferation of transformed cells and it has recently been identified that targeting telomerase, which is responsible for the addition of the repetitive TTAGGG units to the ends of telomeres, presents a potent factor in the control of invasive cancerous cells. The normal addition of the nucleotide repeats acts to avert a crisis in which the increased shortening and telomeric loss would result in cellular senescence (Odago & Gerson, 2003). The recent reports of telomerase reactivation in a large proportion of human cancers have led to the proposal of telomerase as a promising therapeutic target for anti-cancer drugs and as a marker of cancer. Telomerase is a particularly promising anti-tumour target as it is activated in most tumour cells that concurrently have short telomeres and is not expressed in normal somatic cells. In addition to this, stem cells which also have high telomerase activity and long telomeres would be affected by telomerase inhibitors at a later time point and therefore reducing toxicity and maximizing activity (Raymond *et al*, 1996).

Inhibitory methods towards telomeres and telomerase have focused on three main areas:

1. Antisense molecules, such as RNA molecules and oligonucleotides, directed against the TERC RNA component of telomerase.
2. Small molecule reverse transcriptase inhibitors such as azidothymine (AZT).
3. Small molecules with the ability to interact with and stabilize four-stranded (G-quadruplex) structures formed by telomeres. These agents inhibit telomerase in the sub-micromolar range (Kelland, 2000).

There are several approaches to drug discovery, though traditionally it is driven by chemistry but is increasingly being guided by pharmacology and the clinical sciences. The advent of molecular biology and in particular genomic science is having a deep impact on drug discovery (Drews, 2000). The chemistry driven approach to drug discovery sees that the molecule works first and the mechanism

of action is discussed later. In biology based drug discovery, you start with a target and develop drugs from there. The problem with the chemistry-driven approach is the possibility of compounds that appear to work, however, have complex mechanisms of action. These compounds are known as 'dirty compounds' and with these we have various bits of biological data but no clear target or mechanism of action. It is clear, therefore, that these compounds are ideal for a systems biology approach to decompose their mechanism of action. A prime example is 3, 11- difluoro-6, 8, 13- trimethyl-8*H*- quino [4, 3, 2-kl] acridinium methosulfate (RHPS4) which is discussed below.

1.7 RHPS4

Cookson *et al*, describe the syntheses and biological activities of derivatives of the DNA-affinic *8H-quino* [4, 3, 2,-*kl*] acridine skeleton. This work was inspired by reports of the pharmacological properties of related polycyclic acridines of marine and fungal origin. RHPS4 (see figure 1.15) was of particular interest due to its electron-withdrawing fluoro substituents in the 3- and 11- positions and has a higher affinity for triplex and quadruplex DNA and is only weakly growth-inhibitory overall to tumour cells (Cookson *et al*, 2005).

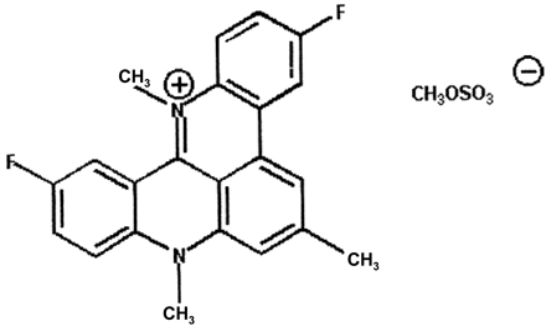


Figure 1.15 Molecular structure of 3,11-difluoro-6,8,13-trimethyl-8*H*-quino[4,3,2-*kl*]acridinium methosulfate (RHPS4). RHPS4 was found to be a potent telomerase inhibitor in the TRAP assay with an IC_{50} of 0.33 μM. It was also found to cause irreversible cessation of growth after long-term culture at non-cytotoxic concentrations in cancer cells with relatively short telomeres (Taken from Leonetti *et al*, 2004).

G-rich DNA sequences, such as those found at the ends of telomeres, are able to adopt an unusual four-stranded DNA structure called G-quadruplex DNA and

represents a potential therapeutic target for anti-tumour agents (Kerwin, 2000). The formation of G-quadruplex DNA has been shown to reduce the activity of telomerase, which is responsible for maintaining telomere length, and is therefore an attractive target for drug discovery.

RHPS4 was synthesized by Robert Heald as part of a range of compounds of related structure with potent telomerase-inhibitory activity. When synthesizing the compounds, three priorities were kept in mind:

1. Potent telomerase-inhibitory activity through selective binding to higher ordered DNA structures such as quadruplex DNA.
2. Possession of robust pharmaceutical properties including chemical and metabolic stability.
3. Accessibility by simple synthetic pathways (Heald *et al*, 2002).

RHPS4 is a water soluble compound and is stable at pH range of 5.0 - 9.0. It is efficiently trafficked into tumour cells and is a difluoro compound with electron-withdrawing fluoro groups and binds more selectively to the higher ordered structures, i.e. triplex and quadruplex DNA, confirmed by DNA melting studies where RHPS4 was shown to increase the melting point (T_m) of salmon testis DNA by 10.8°C. *In-vitro* studies of RHPS4 using the NCI 60 cell line panel, the average GI_{50} was 13.18 μM (Heald *et al*, 2002). Further studies revealed that RHPS4, a novel pentacyclic acridine, was identified as a potent telomerase inhibitor using the telomeric repeat amplification protocol (TRAP) assay (Gowan *et al*, 2001). It was also recently concluded that cancer cells treated with RHPS4 undergo a unique type of telomere damage that is possibly responsible for its anti-tumour effects. RHPS4 has been found, not only to induce telomerase inhibition, but also exert an anti-cancer effect by altering telomeric chromatin as shown in figure 1.16 (Salvati *et al*, 2007).

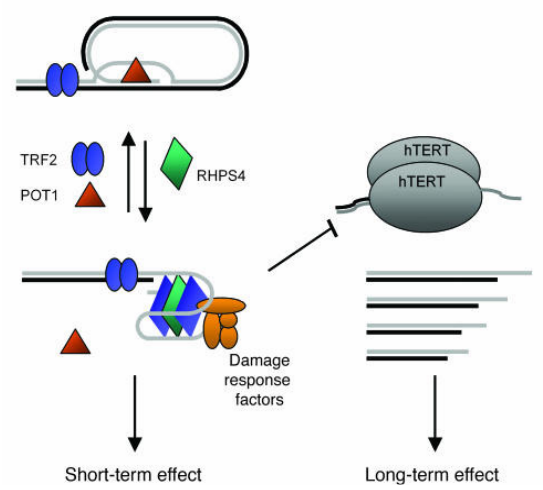


Figure 1.16 The proposed short- and long-term effects of RHPS4. It has recently been identified that cancer cells which have undergone treatment with RHPS4 undergo a unique type of telomere damage which is possibly linked to its anti-tumour effects. RHPS4 has been found, not only to induce telomerase inhibition, but also exert an anti-cancer effect by altering telomeric chromatin (Taken from Salvatti *et al*, 2007).

1.8 Aims and Objectives

Systems biology allows the study of biological systems, such as the cell cycle, by systematically perturbing them either by biological, genetic or chemical means. Upon perturbation, the genes, proteins and pathway responses are monitored and the data is integrated. From the data mathematical models are formulated that describe the system and its response to perturbations (Ideker *et al*, 2001). The aim of this project is to approach the analysis of the complex mechanism of action of RHPS4 in a systems biology manner through the development and application of mathematical models of cell growth and death based on the analysis of cells and their progression through the stages of the cell cycle and alternatively into states of either senescence or death in the absence and presence of anti-tumour agents. Following on from the mathematical modelling is to develop this approach into a robust and easy-to-use tool for other biologists. The approach is similar to that of Basse *et al*, as we use compartments to incorporate as much information regarding the phases of cell cycle as possible, however, we want to derive models that simply describe the flow of cells through the cell cycle in the presence and absence of anti-tumour agents. In the laboratory we are able to derive information with

regards to cell cycle phases using flow cytometry and staining the cells with PI, senescence analysis based on the SA-β-gal stain and number of population doublings using a heamocytometer. This data alone gives information with regards to the effects of RHPS4 and other anti-tumour agents with respect to changes of the proportions of cells in phases of the cell cycle and senescence and also how the number of population doublings changes. What is not obvious from the data is how the kinetics of the cell cycle changes upon the addition of RHPS4 and other anti-tumour agents and the mathematical model aims to resolve this. We have a schematic diagram of the cell cycle as shown in chapter 4 upon which the mathematical models are based and we want to incorporate as much information from this as possible into our compartmental models based on what we can derive in the laboratory and use these to probe the mechanism of action of RHPS4 and allow us to link the data to a possible molecular target.

Mathematical models play an important role in helping to understand complex biochemical processes and the type of insight depends ultimately on the level of understanding already available and the quality of the data. For this reason, the parameterisation of our mathematical model requires robust data from a well-characterised cell line and so the aim of the first part of the project, described in chapter 3, was the analysis of four colorectal cancer cell lines: HCT116, HT-29, KM12 and HCC-2998 to decide which would be suitable to use for further study of the effects of RHPS4 and other anti-tumour agents. Chapter 3 outlines the approach to the selection of a suitable cell line, and the data obtained from these initial studies. The process involved optimisation against a variety of goals, including cell growth characteristics, sensitivity to RHPS4, tractability in a variety of assays, and not least clinical relevance.

Chapter 4 outlines the development of the mathematical models which were used to analyse the data obtained from studies the chosen colorectal cell line. The mathematical models were developed on the basis that in the laboratory we were able to measure the proportion of cells in each phase of the cell cycle, the proportion of cells in senescence and the number of population doublings and so we wanted to incorporate as much information with regards to the cell cycle as

possible. It was also required that the mathematical models were simple enough for the practising biologist, yet complex enough to yield useful information regarding the effects of RHPS4 and other anti-tumour agents that would be difficult to appreciate from the data alone. Following on from the development of the mathematical models was the application of the data obtained in the laboratory to the mathematical models as outlined in chapter 5. The data was applied to all the models that were derived to simply outline that depending upon which mathematical model is used; the conclusions that are derived regarding RHPS4 and other anti-tumour agents can be very different. Data from doxorubicin was also applied to the mathematical models in order to compare the conclusions derived. Doxorubicin has a different mechanism of action than what is proposed for RHPS4. If upon application of the data to the model from both RHPS4 and doxorubicin the same conclusions are generated, the mathematical model fails to decipher between anti-tumour agents and falls apart. In chapter 6, a software application was used to determine whether the results of the mathematical models could be reproduced and essentially 'close the loop'. The data generated regarding the kinetics of the cell cycle from the mathematical model was applied to COPASI to simulate the behaviour of the models and highlight the needs to do this and this is outlined in chapter 6. It is evident that both approaches to systems biology have been taken in this project as we have started with biological data and used this to derive mathematical models and from this we have then used COPASI as a means of validating this approach.