

*Occasional article*

# A mutational theory of leukaemogenesis

J A MORRIS *Department of Pathology, Lancaster Moor Hospital, Lancaster*

Recent advances in molecular biology have clarified the understanding of the process of carcinogenesis.<sup>1,2</sup> It is now thought that quantitative or qualitative anomalies of expression of proto-oncogenes may lead to uncontrolled cell proliferation. This could arise from one or more mutations in the oncogenes or in the genes controlling their expression.<sup>3</sup> These mutations might be caused by chemical carcinogens, viruses, or radiation. Even in the absence of known mutagens, however, there is a spontaneous mutation rate of the order of  $1 \times 10^{-6}$  per gene for each cell generation which could, theoretically, lead to neoplasia.<sup>4,5</sup>

Greaves and Chan recently suggested that spontaneous mutation might be the cause of most cases of childhood acute lymphoblastic leukaemia (ALL).<sup>6</sup> They argued that measured spontaneous mutation rates are sufficient to account for the observed incidence of ALL because there is a very high rate of lymphocyte division in early childhood. The pronounced variation in the recorded incidence of ALL among different countries,<sup>7</sup> different socioeconomic groups,<sup>8</sup> and different time periods<sup>9</sup> seems to contradict this idea. But Greaves and Chan suggest that much of this variation is a consequence of selective underreporting and premature death in places where socioeconomic conditions are poor.<sup>6,10</sup> Furthermore, Greaves has recently pointed out that the lymphocyte proliferation rate in early childhood will be influenced by the pattern of exposure to micro-organisms which will vary with social, environmental, and geographical factors.<sup>11</sup>

The idea that spontaneous mutations might be important in cancer is not new. Indeed, Den Otter *et al* attempted to calculate the incidence of so-called endogenous malignancy and the number of mutations required for carcinogenesis using measured mutant frequencies.<sup>4,5</sup> Unfortunately, their mathematical

treatment contains errors, which the authors appreciate, as they regard the formula used as a first approximation. An improved mathematical formula is presented for application specifically to the incidence of lymphoid malignancy.

### The model

Successive divisions of a cell and its progeny through  $x$  generations, without cell loss, will produce a total of  $2(2^x)$  cells after  $2^x$  cell divisions with  $2^x$  cells in the final generation. The probability that any one of the cells in the final generation has avoided a specific mutation is  $(1-m)^x$  where  $m$  is the mutation rate per gene for each cell generation. Conversely, the probability that a specific mutation has occurred is  $1-(1-m)^x$  which simplifies to  $mx$  when  $m$  is far less than 1. The value  $mx$  specifies the mutant frequency after  $x$  generations. Thus if  $m = 1 \times 10^{-6}$ , the mutant frequency after 20 generations is  $2 \times 10^{-5}$ , and after 40 generations it is  $4 \times 10^{-5}$ . The probability that a cell has undergone  $n$  specific mutations after  $x$  generations is  $(mx)^n$ , assuming that the genes mutate independently. Thus if  $n$  specific mutations are required for carcinogenesis the number of neoplasms arising at the  $x$ th generation is:

$$[(mx)^n - [m(x-1)]^n]2^x$$

and the total number of neoplasms ( $N$ ) is obtained by summing over  $x$  generations:

$$N = \sum_{x=1}^{x=x} [(mx)^n - [m(x-1)]^n]2^x$$

In practice the total number of neoplasms is a little less than double the number produced in the final generation.

If there is extensive cell loss during the process the number of cells in the final generation is  $2^{(a-b)}$ , or

$$N \approx \sum_{x=b}^{x=x} [(mx)^n - [m(x-1)]^n]2^{(x-b)}$$

## Lymphocyte kinetics

Several cell generations obviously occur before the lymphocyte stem cell differentiates from other embryonic cells. The stem cell then undergoes successive divisions to produce a large pool of precursor lymphocytes. During this process germ line genes are randomly rearranged to specify a wide range of lymphocyte receptors.<sup>12</sup> It is thought that many of the genetic rearrangements will code for receptors which cannot be assembled and that these lymphocytes will be lost. Lymphocytes with receptors for self antigens might also be destroyed or suppressed. After birth lymphocytes with functional receptors will divide in response to antigen so that the repertoire of mature lymphocytes is appropriate to the antigenic experience of the host. Exposure to common antigens obviously occurs early in life and the period of rapid cell division is concentrated in the first few years of life.

Unfortunately, although the general features of the above scheme are widely accepted, there is a paucity of precise kinetic data in man and only broad estimates can be used for the model. There are roughly  $7 \times 10^{15}$  cell divisions in a lifetime,<sup>13</sup> and in this paper it is assumed that between 1% and 10% of these are lymphocytic. This is a reasonable approximation because it does not seem likely that lymphocytes comprise more than 10% of all cells, and yet Greaves and Chan have estimated that  $1.8 \times 10^{13}$  lymphocytes are produced in the first year of life.<sup>6</sup> It takes 46 generations without cell loss to produce  $7 \times 10^{13}$  cells from a single stem cell and approximately 49 generations to produce  $7 \times 10^{14}$  cells. The number of cell generations from the zygote to the lymphocyte stem cell will be about seven if lymphocytes comprise 1% of all cells and three or four if they comprise 10% of all cells. Thus the minimum number of cell generations to produce a lymphocyte complement of 1% to 10% is 53. It is thought that the number of cell generations in a lifetime is between 55 and 60,<sup>14</sup> and this would fit with the above analysis because there is extensive cell loss during lymphocyte development. Finally it is assumed that half of the lymphocytes are produced in the first few years of childhood and the rest throughout the remaining years of life.

### The mutation rate

The mutation rate per locus per human generation has been calculated for a large number of inherited diseases and lies between  $1 \times 10^{-4}$  and  $1 \times 10^{-6}$  with an average of  $2 \times 10^{-5}$ .<sup>4,5</sup> As there are 50 to 60 cell generations a generation this gives a mutation rate per gene for each cell generation of the order of  $4 \times 10^{-7}$ . Obviously this is only an approximation as mutations can occur at times other than cell division, certain

mutations can confer a selective disadvantage on the cell, and factors such as the size of the gene will be important. The estimate is of the correct order because measured mutation rates of cells in culture give similar values, usually around  $1 \times 10^{-6}$ .<sup>4,5</sup> Most mutations lead to loss of gene function rather than enhanced or new function. Thus if this rate is relevant to carcinogenesis it presumably acts by deleting regulatory genes and allowing suppressed growth control genes to function. The best example of this is retinoblastoma in which loss of a pair of anti-oncogenes leads to neoplasia.<sup>15</sup>

Cancer could also arise by mutations which directly activate growth control genes. In this case, however, only a subset of mutations would be relevant and the mutation rate would be less than  $4 \times 10^{-7}$ .

### Predictions of the model

#### CUMULATIVE INCIDENCE OF ALL

In the table the predicted cumulative incidence of childhood ALL is shown for varying values of  $m$ ,  $n$ , and  $x$ . The incidence is calculated by assuming that at the  $x$ th generation  $3.5 \times 10^{13}$  or  $3.5 \times 10^{14}$  lymphocytes are produced. The model is relatively robust in that varying the value of  $x$  from 53 to 58 makes only a small difference to the predicted incidence. Equally the rate is linearly related to the number of cells produced but is more affected by variation in  $m$  or  $n$ . For instance, if  $m$  is doubled when  $n = 4$  the predicted incidence rises sixteen-fold. The actual cumulative incidence of ALL in developed countries is 30 to 45/100 000 by age 15 years and the closest fit is with  $n = 4$  and  $m$  between  $9 \times 10^{-7}$  and  $1.7 \times 10^{-6}$ , or with  $n = 2$  and  $m$  between  $6 \times 10^{-11}$  and  $2.4 \times 10^{-10}$ .

The condition  $n = 4$  can be regarded as the loss of four regulatory genes controlling two pairs of growth control genes. The predicted mutation rate between  $9 \times 10^{-7}$  and  $1.7 \times 10^{-6}$  is reasonably close to the

Table Cumulative incidence of ALL using the formula

$$N \approx \sum_{x=b}^{x=x} [(mx)^n - [m(x-1)]^n] 2^{(x-b)}$$

| $m$                   | $n$ | $x$ | Cumulative incidence of ALL   |                               |
|-----------------------|-----|-----|-------------------------------|-------------------------------|
|                       |     |     | $2^{14} = 3.5 \times 10^{13}$ | $2^{14} = 3.5 \times 10^{14}$ |
| $1 \times 10^{-6}$    | 4   | 58  | $5.05 \times 10^{-5}$         | $5.05 \times 10^{-4}$         |
| $1 \times 10^{-6}$    | 4   | 53  | $3.83 \times 10^{-5}$         | $3.83 \times 10^{-4}$         |
| $1 \times 10^{-6}$    | 5   | 58  | $1.1 \times 10^{-10}$         | $1.1 \times 10^{-9}$          |
| $1 \times 10^{-6}$    | 3   | 58  | $8.34 \times 10^{-2}$         | $8.34 \times 10^1$            |
| $9 \times 10^{-7}$    | 4   | 58  |                               | $*3.3 \times 10^{-4}$         |
| $1.7 \times 10^{-6}$  | 4   | 58  | $*4.2 \times 10^{-4}$         |                               |
| $6 \times 10^{-11}$   | 2   | 58  |                               | $*2.9 \times 10^{-4}$         |
| $2.4 \times 10^{-10}$ | 2   | 58  | $*4.5 \times 10^{-4}$         |                               |

\*Closest fit to the observed cumulative incidence of  $3 \times 10^{-4}$  to  $4.5 \times 10^{-4}$  by age 15 years.

### *A mutational theory of leukaemogenesis*

assumed spontaneous mutation rate of  $4 \times 10^{-7}$ , particularly as it has been argued that the mutation rate in precursor lymphocytes might be higher than in other cells because of breaks in DNA induced by enzymes occurring naturally during germ line gene rearrangement.<sup>6</sup>

The condition  $n = 2$  can be regarded as the direct activation of two growth control genes. In this case the spontaneous mutation rate per gene is between  $3 \times 10^{-11}$  and  $1.2 \times 10^{-10}$  because each gene will be represented twice (maternal and paternal copies) unless it is sex linked. Unfortunately, there is no independent, experimental, or observational support for mutation rates of such rarity.

#### INCIDENCE OF LYMPHOID MALIGNANCY

If lymphoid malignancy in adults was also a consequence of spontaneous mutation occurring at mitosis then according to this model the total incidence would equal that of ALL in childhood. This is because it is assumed that half of all lymphocytes are produced in childhood. In fact, the incidence of non-Hodgkin's lymphoma alone is 9/100 000/year<sup>16</sup> (the incidence of ALL is 2-3/100 000/year) and therefore there must be additional environmental agents increasing the chance of genetic damage.

#### INTRAUTERINE AND EXTRAUTERINE MUTATIONS

Consider a simplified scheme in which the lymphocyte stem cell differentiates after seven generations and then divides through 45 generations to produce in the last generation the childhood lymphocyte complement of  $3.5 \times 10^{13}$  cells. It is possible, using the model, to determine the effect of changing the mutation rate at different periods. If the mutation rate is doubled throughout the 52 generations then, with  $n = 4$ , the incidence of ALL would rise sixteen-fold. If the mutation rate is doubled for the first 49 generations and then returned to normal for the final three generations the incidence of ALL would rise ten-fold. If the mutation rate is constant for 49 generations and then doubled for the final three generations the incidence of ALL would double. As 87.5% of lymphocyte divisions occur in the final three generations it follows that the number of generations over which the increased mutation rate acts is more important than the number of cell divisions affected by the increased rate. This is obviously relevant to consideration of the differential effect of carcinogens or mutagens during intrauterine and extrauterine life. It is also of particular interest in view of the recent evidence that the increased risk of ALL in Seascale is confined to children who are born there.<sup>17,18</sup> This implies exposure in the uterus or early in extrauterine life. If the spontaneous mutation rate is doubled throughout the period of lymphocyte development the incidence of

ALL would rise sixteen-fold. This is the order of increase seen in Seascale.

It must be emphasised that the above analysis is a simplification. It does not allow for cell loss, and the situation is complicated by the likelihood that cell generations are not necessarily in synchrony. Thus some late cell generations will occur in the uterus and some earlier cell generations after birth.

This simplified scheme equating 49 generations with intrauterine life, three generations with early childhood, and one final generation with adult life, is, however, of some value. It predicts that about 10% of ALL clones will arise in the uterus. Thus the concordance rate of ALL in identical twins who share an intrauterine circulation should be about 20% which is the observed rate.<sup>19</sup>

#### THE AGE INCIDENCE OF ALL

The two hit ( $n = 2$ ) and four hit ( $n = 4$ ) models both predict that the incidence of clonal origin of ALL will be roughly proportional to the number of lymphocyte divisions a year. This is because most lymphocytes are produced as a result of the final few generations, during which the probability of neoplasia only changes slightly. The a priori expectation is that lymphocyte division will be maximal in the first year of life, when most common antigens are met, and decrease thereafter. The actual incidence of ALL, however, rises sharply from birth to a peak in the fourth year of life after which there is a rapid fall.<sup>7</sup> Thus there is possibly a three year latent interval between clonal origin and disease presentation. The median age of onset in identical twins who are concordant for ALL is less than 12 months,<sup>20</sup> however, indicating that the median latent interval from clonal origin to presentation is less than 21 months and is probably around 12 months.

The model could be adapted by assuming that either the spontaneous mutation rate rises and then falls in childhood or that the number of lymphocyte divisions rises and then falls, but it is difficult to substantiate either assumption. Delayed exposure to micro-organisms in certain socioeconomic groups will affect lymphocyte proliferation rates in childhood as implied by Greaves.<sup>11</sup> The age incidence of first contact with common antigens, however, must be a falling curve, and the effect of delayed exposure will be to decrease the rate of fall not to reverse it. It is unlikely, therefore, that this will translate into an increasing rate of cell division.

The discussion so far, however, is based on the assumption that malignant clones arise as a result of multiple hits in a single cell and this determines the age incidence. But it is possible that there are mechanisms for eliminating malignant clones, although if so, that value of  $n$  would be reduced in relation to the efficiency

of those mechanisms. It has been argued elsewhere that the efficiency of biological detection systems will deteriorate with time.<sup>21</sup> Thus the age distribution of ALL will be a function of the rate of lymphocyte division, the latent period from clonal origin to disease presentation, and the efficiency of clonal elimination. In this way the age incidence of ALL can be explained.<sup>21</sup>

### Conclusion

Epidemiological and kinetic data are consistent with the concept that childhood ALL is a consequence of spontaneous mutation. This could occur by the activation of two growth control genes either directly ( $n = 2$ ) or by the loss of two pairs of regulatory genes ( $n = 4$ ). Measured mutation rates fit best with the latter possibility. The mathematical model shows that if an increased mutation rate affects a subset of cell divisions then the number of cell generations affected is more important than the absolute number of cell divisions affected. The model, however, fails to explain the age distribution of ALL without the additional assumption of a fallible mechanism which normally eliminates malignant clones—in which case, the activation of one growth control gene by the loss of two regulatory genes would be sufficient for neoplasia.

### References

- 1 Paul J. Oncogenes. *J Pathol* 1984;**143**:1–10.
- 2 Weiss RA, Marshall CJ. Oncogenes. *Lancet* 1984;ii:1138–42.
- 3 Green AR, Wyke JA. Anti-oncogenes: a subset of regulatory genes involved in carcinogenesis? *Lancet* 1985;ii:475–7.
- 4 Den Otter W, Kolen JW, Derkinderen DJ. A new model of oncogenesis. *Anticancer Res* 1986;**6**:509–14.
- 5 Den Otter W, Kolen JW, Derkinderen DJ. Carcinogenesis revisited. *Cancer Invest* 1987;**5**:69–74.
- 6 Greaves MF, Chan LC. Is spontaneous mutation the major cause of childhood acute lymphoblastic leukaemia? *Br J Haematol* 1986;**64**:1–13.
- 7 Greaves MF, Pegram SM, Chan LC. Collaborative group study of the epidemiology of acute lymphoblastic leukaemia subtypes: background and first report. *Leuk Res* 1985;**9**:715–33.
- 8 Ramot B, Magrath I. Hypothesis: the environment is a major determinant of the immunologic subtype of lymphoma and acute lymphoblastic leukaemia in children. *Br J Haematol* 1982;**52**:183–9.
- 9 Court-Brown WM, Doll R. Leukaemia in childhood and young adult life. Trends in mortality in relation to aetiology. *Br Med J* 1961;**26**:981–8.
- 10 Stewart A, Kneale GW. Role of local infections in the recognition of haemopoietic neoplasms. *Nature* 1969;**233**:741–2.
- 11 Greaves MF. Speculation on the cause of childhood acute lymphoblastic leukaemia. *Leukaemia* 1988;**2**:120–5.
- 12 Korsmeyer SJ, Arnold A, Bakshi A, et al. Immunoglobulin gene rearrangement and cell surface antigen expression in acute lymphocytic leukaemia of T cell and B cell precursor origin. *J Clin Invest* 1983;**71**:301–13.
- 13 Falkman J. The vascularization of tumours. *Sci Am* 1976;**234**:59–73.
- 14 Hayflick L. Biological aspects of human ageing. *Br J Hosp Med* 1982;**12**:366–72.
- 15 Goudie RB. What are antioncogenes? *J Pathol* 1988;**154**:297–8.
- 16 Leukaemia Research Fund 1984 Data Collection Group. Distribution of leukaemia, lymphoma and allied diseases in parts of Great Britain: analysis by administrative districts and simulations of adjacencies. *Leukaemia* 1987;**1**:78–81.
- 17 Gardner MJ, Hall AJ, Downes S, Terrell JD. Follow up study of children born elsewhere but attending schools in Seascale, West Cumbria (schools cohort). *Br Med J* 1987;**295**:819–22.
- 18 Gardner MJ, Hall AJ, Downes S, Terrell JD. Follow up study of children born to mothers resident in Seascale, West Cumbria (birth cohort). *Br Med J* 1987;**295**:822–7.
- 19 Clarkson BD, Boyce EA. Possible explanation of high concordance for acute leukaemia in monozygous twins. *Lancet* 1971;i:699–701.
- 20 Zuelzer WF, Cox DE. Genetic aspects of leukaemia. *Semin Haematol* 1969;**6**:228–49.
- 21 Morris JA. Autoimmunity: decision theory model. *J Clin Pathol* 1987;**40**:210–5.

Requests for reprints to: Dr J A Morris, Consultant Pathologist, Department of Pathology, Lancaster Moor Hospital, Lancaster LA1 3JR, England.