Reprint - Classics in Biology

Reprint from: *Speculations in Science and Technology*, Vol. 6. No. 5 (1983) p.471- 478 (249L) 471
Vol. 4, No. 1 (2020)
ISSN: 2532-5876

DOI: 10.13133/2532-5876/16964

A Probabilistic Theory for Cell Differentiation, Embryonic Mortality and DNA C-value Paradox

Jean-Jacques Kupiec[a]*

[a] *Systems Biology Group Lab, University Sapienza, Rome, Italy*

*Corresponding author: Jean-Jacques Kupiec, jj.kupiec@sbglab.org

Abstract
A probabilistic theory for cell differentiation is proposed in which it is postulated that differential gene expression is provoked by random events. An analysis of determinist theories is made, and two predictions based on the probabilistic theory are compared to experimental fact. A probabilistic model of gene regulation is also given. This theory can account for several phenomena: differential gene expression, embryonic mortality, DNA C-value paradox; and it does not need t refer to a wide diversity of specific regulators.

Reprint from: *Speculations in Science and Technology*, vol. 6. no. 5 (1983) p. 471-478 (249L) 471

Citation: Kupiec, JJ, 2020, "A Probabilistic Theory for Cell Differentiation, Embryonic Mortality and DNA C-value Paradox", *Organisms: Journal of Biological Sciences*, vol. 4, no. 1, pp. 80-85. DOI: 10.13133/2532-5876/16964

Introduction

In this article a theory for cell differentiation is presented in which differential gene expression in the different cell lines constituting an organism is the result of random events which occur within the cells at the level of interactions between regulatory molecules and genes. In the frame of this theory, interactions between the different cell types do not play any inductive role in differential gene expression, but intervene secondarily to control and co-ordinate the development of the different tissues.

Three elements have led me to elaborate this theory: (i) the idea that cybernetic concepts cannot describe biological reality because it is radically different from the machine world. (ii) The analysis of determinist models of cell differentiation. (iii) The work of Geissler et al. (Geissler et al., 1977), which shows that the survival probability of an embryo is independent of that of the other embryos carried by the same female. This will be analyzed in the third part of this article.

Since the experiments of Spemann and Mangold in the early twenties, which gave rise to the concept of induction, most of the models that have been proposed to explain cell differentiation have a common basis: it is considered that a cell is determined to differentiate because it has received specific information. This information is generally thought to come from another cell in the form of a regulatory molecule (Figure1).

Differences between models depend on: (i) the chemical nature of the regulatory molecules; (ii) the type of control of the regulatory molecules (activation or repression) (iii) the direct action of these regulatory molecules at the chromatin level, or by the intermediary of membrane signals.

These models contain a contradiction since the cell, which transmits the information (Cell 1 in Figure 1), is already different from the one that receives the infor-

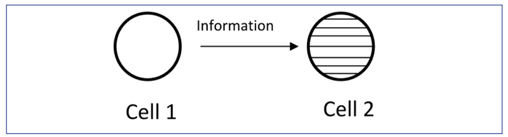

Figure 1

(cell 2 of Figure 1) at the beginning or the process because it synthesizes one or several informative molecules not synthesized by the other cell. In order to resolve this contradiction, the morphogenetic gradient theory is usually used (Davidson and Britten, 1971).

The implicit negative hypothesis of these models is to consider that embryonic cells, when left to themselves, are in a stable state, that they replicate identically and that they cannot express other genes without the action of an inductor.

This hypothesis has never been (and maybe cannot be) experimentally demonstrated. The fact that in certain cases cell differentiation or a specific gene expression can be induced chemically (by hormones, for example) proves that cells are inducible for certain genes, but it does not prove that this is true for all genes and that, in general, cells cannot express different genes without an induction.

On the contrary, the probabilistic theory's initial hypothesis assumes that, because of internal and random events which cause certain genes to be activated or repressed, eukaryotic cells can differentiate without the intervention of external signals.

The theoretical framework, which is thus defined, allows us to conceive that two cells become different from each other even though they are identical at the beginning of the process: in Figure 2 below, cells 1 and 2 are identical. According to whether the random event a or b occurs in either of them, they are transformed into a type A or B cell respectively.

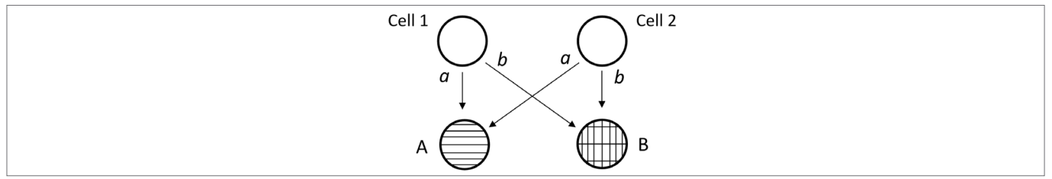

Figure 2

The implicit negative hypothesis of these models is to consider that embryonic cells, when left to themselves, are in a stable state, that they replicate identically and that they cannot express other genes without the action of an inductor.

This hypothesis has never been (and maybe cannot be) experimentally demonstrated. The fact that in certain cases cell differentiation or a specific gene expression can be induced chemically (by hormones, for example) proves that cells are inducible for certain genes, but it does not prove that this is true for all genes and that, in general, cells cannot express different genes without an induction.

On the contrary, the probabilistic theory's initial hypothesis assumes that, because of internal and random events which cause certain genes to be activated or repressed, eukaryotic cells can differentiate without the intervention of external signals.

The theoretical framework, which is thus defined, allows us to conceive that two cells become different from each other even though they are identical at the beginning of the process: in Figure 2 below, cells 1 and 2 are identical. According to whether the random event a or b occurs in either of them, they are transformed into a type A or B cell respectively.

Before raising the question or the nature of the random events who provoke cell differentiation, two predictions based on this hypothesis can be made and compared to experimental facts.

First Prediction: If the appearance of the different cell types that constitute an organism depends on the occurrence of random events, this phenomenon must only have a certain probability to succeed each time that it occurs. Therefore, it must also have a certain failure probability.

Second Prediction: In order to succeed, each of the random events leading to a cell type must occur at least once within one of the cells that constitute the embryo. If this does not happen, one or several cell types will be missing and it will be a failure.

The failure probability should be a function of, on the one hand, the probabilities of the random events in each cell leading to the different cell ty pes; and, on the other hand, the number of cells, which constitute the embryo the greater this number, the smaller the failure probability, will be.

These predictions are compatible with some experimental facts. There is always a certain failure rate in embryogenesis. This phenomenon has mainly been studied by agro-biologists with the purpose of increasing the profitability of stock. According to Vandeplassche (Vandeplassche,1968) “All species have +/- 25 % embryonic mortality which occurs before, during and shortly after implantation, so that one is inclined to believe that embryonic mortality is, at least to a certain extent, a normal phenomenon.”

Several hypotheses (either genetic or physiological) have been made to explain the cause or embryonic mortality (Bishop,1964). Various factors may influence it: breed, age, genotype, temperature (for reviews on this subject (Edey, 1969; Ayalon, 1978; Fechheimer, 1979; Gustavsson, 1979). But no definitive explanation bas been given for this phenomenon. The probabilist theory provides a simple and coherent explanation since it predicts a failure probability for cell differentiation. Moreover it can give an explanation for certain experimental results that classical theories can hardly account for. Allison (Allison, 1975) has shown that in sheep there is a relationship between the survival probability of an embryo and the number of embryos carried simultaneously by the gestating female. When the number of embryos carried simultaneously by the female increases, the survival probability of each embryo decreases. Geissler et al. (Geissler et al., 1976) have done a mathematical analysis of these results. They have shown that the survival of a fertilised ovum depends only on the number of ova carried with it and is independent of the survival or death of those carried with it.

Now, if the survival probability of one embryo is independent of those of the other embryos carried by the same female, this indicates that the determining cause of this probability is internal and not external to the embryo.

The hypotheses that can be made within the frame of classical theories to explain embryonic mortality are hardly able to account for this result. In the case of hypotheses which postulate that in given conditions, there is a limit to the number of embryos that can develop inside the same female (Vandeplassche, 1968) - this limit being determined by factors such as the physical space of the uterus, nutritive conditions, hormonal imbalance, etc. -the survival probabilities of the embryos should be dependent. (For example, from the moment that the maximum number of embryos having a normal development is reached the survival probability of the remaining embryos should become zero.)

In the case of genetical hypotheses: at least an important part of mutations, chromosomal abnormalities and aberrations result from errors that occur before fecundation and there is no reason for the number of these errors to increase when the number of released ova increases. However one might postulate that the number of remaining errors occurring during or shortly after fecundation increases if the number of released ova increases. But such a hypothesis is incompatible with the results of Gates (Gates, 1956) who, by the transplantation of blastocysts from females which had been induced to super- ovulate in to non-treated females, found that these blastocysts were genetically and physiologically normal.

The probabilistic theory explains this result in the following way: when the number of embryos carried by the same female increases, the number of cells constituting each embryo in each phase of its development decreases (This hypothesis is plausible and can be tested). Now, the survival probability is as we have previously seen (second prediction) dependent on the number of cells which constitute the embryo. If this number decreases, the survival probability will also decrease. But in this case, the survival probability of one embryo remains independent of the survival probability of the other embryos since it is only dependent on the number of cells and not on the failure or survival of the other embryos. The fact that this probability decreases has for its ultimate effect that, statistically, the number of surviving embryos decreases if the number of embryos carried by the same gestating female increases.

A relation between the number of cells and the viability of embryos has already been described (Wu, 1976).

Probabilistic theory of gene regulation in eucaryotic cells

Determinist models of gene regulation in eukaryotic cells presuppose the existence of specific regulators which specifically activate or repress different genes during cell differentiation. These models are usually elaborated by a re-thinking of the Monod-Jacob model of the lactose operon of E. Coli. Here again, this is an undemonstrated hypothesis because these models must presuppose a wide diversity of regulators, not yet discovered experimentally in order to explain the diversity of tissues which constitute an organism.

Probabilistic model of gene regulation in eukaryotic cells. This model is based on two basic principles: 1) the molecules which interact with DNA and activate or repress genes are non-specific regulators. Each regulator in the cell’s nucleus is present in a quantity smaller than the N number or DNA sequences with which it can interact.

The choice of the q DNA sequences among the possible N sequences, with which the q regulatory molecules interact in a cell, occurs in a stochastic way.

There is therefore a combination of distributional possibilities of the q molecules, taken the N DNA sequences. Each of these distributions corresponds to the activation or repression of q different genes (or set of genes).

At the time of DNA replication, in each cell the q regulatory molecules are redistributed over the N sequences and a different distribution or the q molecules over the N sequences may result. However, the transition of one distribution to another one is not equiprobable for all possible distributions.

The probability of transition f rom one distribution to another is a function of the relative positions of the different regulatory sequences of DNA (Different parameters might intervene to determine this probability so that it is not necessarily directly proportional to the distance separating the DNA sequences. With a view to simplicity, I have only considered distance along the DNA chain as determining the probability of transition in the following examples. This does not change the logic of the theory).

Let us consider, for example, a very simple case in which q= 1 and N = 4 (Figure 3). When the regulatory molecule is in N1, it has an equal probability of moving into N2 or N5.

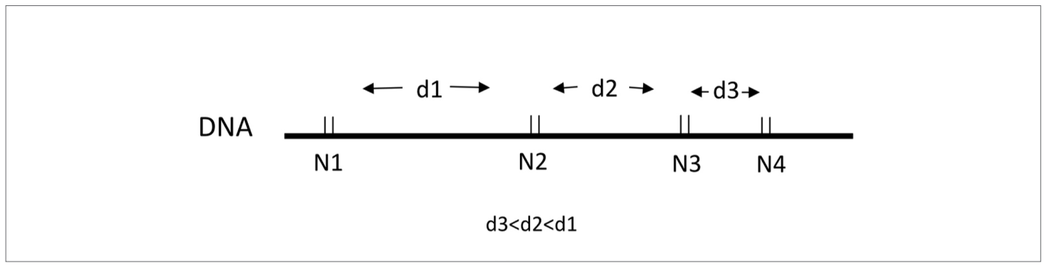

Figure 3

The four DNA sequences Nl, N2, N3 and N4 with which the regulatory molecule can interact have a distance of dl, d2, d3, between them. However these three distances are not equal so that: d3 < d2 < dl. When the regulatory molecule is in Nl, the probability for it to move into N2 at the time of DNA replication is higher than the probability for it to move into N3 which is further away, and even higher than the probability for it to move into N4 which is still further away.

When the regulator is in N2, the nearest sequence is N3. Therefore, the highest probability is that it will move into N3.

Similarly, when the regulator is in N3, the highest probability is that it will move into N4.

In an embryonic cell which starts dividing and in which the regulatory molecule is in N1, the successive transition into N2, N3 and N4 is most probable and each of these situations will cause different genes to be activated or repressed. So that the relative position of the four sequences Nl, N2, N3 and N4 in relation to one another determines the differentiation program which has the highest probability of being achieved.

With more genes or more regulators one can obtain several cell lines.

For example, let us now consider the case in which one regulatory molecule can interact with seven regulatory sequences (Figure 4). These different sequences are separated from each other by the distances d1, d 2, d3, d4, d5 and d6. N1 is equidistant from N2 and N5: d l = d4. d3 < d 2 < dl, d6 < d5 < d4.

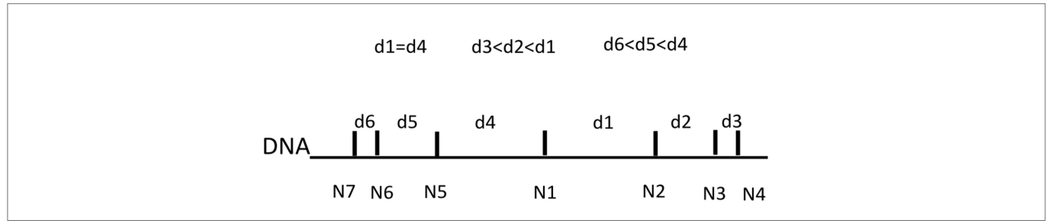

Figure 4

Among the embryonic cells which are in the process of dividing, in certain cases the regulator will move into N2 and in other cases into N5. In the cells where the regulator has moved into N2, the probability to move on will be highest in N3 and N4 successively; in the cells where it has moved into N5, the probability to move on will be highest in N6and N7 successively. In this case, the initial cell where the regulator was in N1 will have given rise to two different cell lines.

The relative position of the different sequences along the DNA chain defines a kind of "supercode" which determines the cell differentiation program. This differentiation program has a certain probability to be achieved, that is, it also has a certain reverse probability not to be achieved.

As we have seen previously, this proposal, which may seem paradoxical, is not incompatible with experimental facts.

Numerous models of varying complexity can be constructed with the same operating mode. These models would differ from each other by the value of q and N, the number of non-specific regulators involved, the type of control and the chemical nature of the regulators.

In reality. it is unlikely that there is a single model, valid for all phases of cell differentiation and for all species.

To my knowledge, there is as yet no direct experimental proof in favor of this type of gene regulation. However certain interesting consequences of these models can be considered.

If the relative position of the different regulatory sequences determines the differentiation program, the portions of DNA between these sequences, including the non-coding portion s, play an essential role because they determine the probability of transition from one sequence to another of the regulatory molecules.

This could partly explain why the eucaryotes have an excess of DNA which has been called the "C value paradox", to which no clearly defined function has been assigned, and which has given rise to the concept of "selfish DNA" (Orgel and Crick, 1980).

The role or certain portions of non-coding DNA might be to keep the genes at a certain distance from each other in order to maintain the relative position of these genes, that is the differentiation program.

From an evolutionist point of view, it may also be advantageous for organisms to have bits of non-coding DNA in reserve which, by changing their position in the DNA molecule, would modify the relative position of the genes, causing the differentiation program to vary.

Moreover, if certain of the N genes, which can be regulated by regulatory molecules are repeated several times, they have a higher probability to be activated or repressed. The repetition of a gene that plays an important role during cell life or at a certain stage of development, gives this gene a higher probability to be expressed or repressed than if it only occurs once in the genome.

Role of cell interactions

During development, cells not only differentiate but also give rise to the organized structure, which constitutes an adult organism. This implies a co-ordination in the development of the different tissues.

In the frame of the probabilistic theory this is achieved by means of cell interactions which intervene secondarily to coordinate the development of the different tissues which first emerged in a random manner (It is not a new idea that cell interactions play a role in development, however, in the frame of the probabilistic theory these interactions have no primary inductive role in cell differentiation).

This means that differentiation is achieved in two steps (at each phase). Gene regulation occurs in a stochastic manner (cf. section 4). Different cell types emerge but they might change their determination at each replication. They remain totipotent. No organization can emerge. Different cell types interact and the result of this interaction is that they become monopotent. They cannot change their determination any longer. An organization can appear. In this way, the multiplication of one cell type cannot occur independently of the multiplication of the other cell types: there is a coordination in the development or the different tissues.

Conclusion

It may seem paradoxical to explain a phenomenon such as cell differentiation, which seems to be repeated in exactly the same way each time that it occurs, by a theory based on the occurrence of random events. The probabilistic theory explains this paradox in the following way: when we think of, or look at, cell differentiation, we usually only consider cases that succeed, so that it seems always to be an identical process. But if we consider the cases when embryonic mortality occurs, we must consider that this process is not always the same. Only one (or a very few) out of the different manners in which the differentiation program may occur leads to a viable embryo.

In the future, the probabilistic theory of cell differentiation presented here should be complemented by a theory or cell interactions, in order to account for the whole process of embryonic development.

However, the probabilistic theory already presents different characteristics and advantages: (a) it explains cell differentiation without referring to a preformationist concept such as the morphogenetic gradient; (b) it does not need a wide diversity of specific regulators; (c) it can explain the DNA C-value paradox ; and (d) it can explain embryonic mortality.

Acknowledgments

I would like to thank Professor Jean Tavlitzski for helpful and pleasant discussions.

References

Geissler, P.A., Newton , .J.E. and Mohan , A.E., A mathematical model of fertilization failure and early embryonic mortality in sheep, J. Ag. Sci. Camb., 89, 309- 317 (1977).

Davidson, E.H. and Britten, R.J., Note on the control of gene expression during development, .J. Theor. Biol., 32, 123-130 (1971).

Vandeplassche , M., La mortalité embryonnaire et son diagnostic, Vie. cong. inter. reprod. anim. insem. artif . Paris, Vol. 1 (1968).

Bishop, M.W.H., Paternal contribution to embryonic death, J. Reprod. Fertil., 7, 383-396 (1964).

Edey, T.N., Prenatal mortality in sheep: a review, Animal Breeding Abstracts,37,173-190 (1969).

Ayalon, N., A review of embryonic mortality in cattle, J. Reprod. Fertil ., 54, 484-493 (1978).

Fechheimer, N .S. Cytogenetics i n animal production, J. Dairy Sci., 62, 844-853 (1979).

Gustavsson, I., Distribution and effects of the 1/19 Robertsonian Translocation in cattle, J. Dairy Sci., 62, 825-835 (1979).

Allison, A.J., Effects of nutritionally induced live-weight differences on the ovarian response and fertility in ewes treated with PMS, NZ J.g.Res., 18, 101-107 (1975).

Gates, A.H., Viability and developmental capacity of eggs from immature mice treated with gonadotrophins, Nature, 177, 754-755 (1956).

Wu , J.T., Cell loss of the aged rat blastocysts in relation to their viability, Biol. of the Reprod., 15, 350-355 (1976).

Orgel, L.E. and Crick, F.H.C., Selfish DNA: the ultimate parasite, Nature, 284, 604-607 (1980).

Doolittle, W.F. and Sapienza, C., Selfish genes, the phenotype paradigm and genome evolution, Nature, 284, 601-603 (1980).