ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 03, 2021

A STUDY ON PARTIAL CALCULUS WITH DIFFERENTIAL AND INTEGRAL VIEWS OF GENETICS IN COMPUTATIONAL SYSTEMS BIOLOGY

Sarwar Sultana
Assistant Professor in Mathematics , Government women's first grade college, jewargi colony kalaburagi.

Abstract:

This article uses an integrative systems biological view of the relationship between genotypes and phenotypes to clarify some conceptual problems in biological debates about causality. The differential (gene-centric) view is incomplete in a sense analogous to using differentiation without integration in mathematics. Differences in genotype are frequently not reflected in significant differences in phenotype as they are buffered by networks of molecular interactions capable of substituting an alternative pathway to achieve a given phenotype characteristic when one pathway is removed. Those networks integrate the influences of many genes on each phenotype so that the effect of a modification in DNA depends on the context in which it occurs. Mathematical modelling of these interactions can help to understand the mechanisms of buffering and the contextual-dependence of phenotypic outcome, and so to represent correctly and quantitatively the relations between genomes and phenotypes. By incorporating all the causal factors in generating a phenotype, this approach also highlights the role of non-DNA forms of inheritance, and of the interactions at multiple levels.

Keywords: genotype; phenotype; computational systems biology.

INTRODUCTION

Organisms encoded as molecular descriptions in their genes? By analysing the genome, could we solve the forward problem of computing the behaviour of the system from this information, as was implied by the original idea of the ‘genetic programme’ [1] and the more modern representation of the genome as the ‘book of life’? In this article, I will argue that this is both impossible and incorrect. We therefore need to replace the gene-centric ‘differential’ view of the relation between genotype and phenotype with an integrative view.

IMPOSSIBILITY

Current estimates of the number of genes in the human genome range up to 25 000, though the number would be even larger if we included regions of the genome forming templates for non-protein coding RNAs and as yet unknown numbers of microRNAs [2]. With no further information to restrict them, the number of conceivable interactions between 25 000 components is approximately 1070000 [3]. Many more proteins are formed than the number of genes, depending on the number of splice variants and post-transcriptional modifications. Proteins are the real workhorses of the organism so the calculation should really be based on

this number, which may be in excess of 100 000, and further increased by a wide variety of post-translational modifications that influence their function. Of course, such calculations are not realistic. In practice, the great majority of the conceivable interactions cannot occur. Compartmentalization ensures that some components never interact directly with each other, and proteins certainly do not interact with everything they encounter. Nevertheless, we cannot rely on specificity of interactions to reduce the number by as much as was once thought. Most proteins are not very specific [4,5]. Each has many interactions (with central hubs having dozens) with other elements in the organism [6], and many (around 30%) are unstructured in the sense that they lack a unique three-dimensional structure and so can change to react in variable ways in protein and metabolic networks [7]. In figure 1, I show the calculations for a more reasonable range of possible interactions by calculating the results for between 0 and 100 gene products for each biological function (phenotype characteristic) for genomes up to 30 000 in size. At 100 gene products per function, we calculate around 10300 possible interactions. Even when we reduce the number of genes involved in each function to 25 we still calculate a figure, 1080 , which is as large as the estimated number of elementary particles in the universe. These are therefore literally ‘astronomic’ numbers. We do not yet have any way of exploring interaction spaces of this degree of multi-dimensionality without insight into how the interactions are restricted. Computational biology has serious difficulties with the problem of combinatorial explosion even when we deal with just 100 elements, let alone tens of thousands.

Given these estimates of the scale of the forward problem, no-one should contemplate calculating the interactions in this massively ‘blind’ bottom-up fashion. That is the reason why the middle-out approach has been proposed [8]. This was originally a suggestion made by Brenner et al. [9]. The quotations from that Novartis Foundation discussion are interesting in the present context. Brenner wrote ‘I know one approach that will fail, which is to start with genes, make proteins from them and to try to build things bottom-up’ ([9], p. 51) and, then later, ‘Middle-out. The bottom-up approach has very grave difficulties to go all the way’ ([9], p. 154). My interpretation of the ‘middle-out’ approach is that you start calculating at the level at which you have the relevant data. In my work, this is at the level of cells, where we calculate the interactions between the protein and other components that generate cardiac rhythm, then we reach ‘out’ to go down towards the level of genes [10] and upwards towards the level of the whole organ [11,12].1 By starting, in our case, at the level of the cell, we focus on the data relevant to that level and to a particular function at that level in order to reduce the number of components we must take into account. Other computational biologists choose other levels as their middle In practice, therefore, even a dedicated bottom-up computational biologist would look for ways in which nature itself has restricted the interactions that are theoretically possible. Organisms evolve step by step, with each step changing the options subsequently possible. I will argue that much of this restriction is embodied in the structural detail of the cells, tissues and organs of the body, as well as in its DNA. To take this route is therefore already to abandon the idea that the reconstruction can be based on DNA sequences alone.

COMPARING THE DIFFERENT FORMS OF INHERITANCE

How does non-DNA inheritance compare with that through DNA? The eukaryotic cell is an unbelievably complex structure. It is not simply a bag formed by a cell membrane enclosing a protein soup. Even prokaryotes, formerly thought to fit that description, are structured [13] and some are also compartmentalized [14]. But the eukaryotic cell is divided up into many more compartments formed by the membranous organelles and other structures. The nucleus is also highly structured. It is not simply a container for naked DNA, which is why nuclear transfer experiments are not strict tests for excluding non-DNA inheritance. If we wished to represent these structures as digital information to enable computation, we would need to convert the three-dimensional images of the cell at a level of resolution that would capture the way in which these structures restrict the molecular interactions. This would require a resolution of around 10 nm to give at least 10 image points across an organelle of around 100 nm diameter. To represent the three-dimensional structure of a cell around 100 mm across would require a grid of 10 000 image points across. Each gridpoint (or group of points forming a compartment) would need data on the proteins and other molecules that could be present and at what level. Assuming the cell has a similar size in all directions (i.e. is approximately a cube), we would require 1012 gridpoints, i.e. 1000 billion points. Even a cell as small as 10 mm across would require a billion grid points. Recall that the genome is about three billion base pairs. It is therefore easy to represent the threedimensional image structure of a cell as containing as much information as the genome, or even more since there are only four possible nucleotides at each position in the genome sequence, whereas each grid point of the cellular structure representation is associated with digital or analogue information on a large number of features that are present or absent locally. There are many qualifications to be put on these calculations and comparisons. Many of the cell structures are repetitive. This is what enables cell modellers to lump together compartments like mitochondria, endoplasmic reticulum, ribosomes, filaments, and other organelles and structures, though we are also beginning to understand that, sometimes, this is an oversimplification. A good example is the calcium signalling system in muscles, where the tiny spaces in which calcium signalling occurs, that couples excitation to contraction have to be represented at ever finer detail to capture what the experimental information tells us. Current estimates of the number of calcium ions in a single dyad (the space across which calcium signalling occurs) is only between 10 and 100 [15], too small for the laws of mass action to be valid.

Nevertheless, there is extensive repetition. One mitochondrion is basically similar to another, as are ribosomes and all the other organelles. But then, extensive repetition is also characteristic of the genome. A large fraction of the three billion base pairs forms repetitive sequences. Protein template regions of the human genome are estimated to be less than 1.5 per cent. Even if 99 per cent of the structural information from a cell image were to be redundant because of repetition, we would still arrive at figures comparable to the effective information content of the genome. And, for the arguments in this paper to be valid, it does not really matter whether the information is strictly comparable, nor whether one is greater than the other. Significance of information matters as much as its quantity. All I need to establish at this point is that, in a bottom-up reconstruction or indeed in any other kind of reconstruction it would be courting failure to ignore the structural detail. That is precisely what restricts the combinations of interactions (a protein in one compartment cannot interact

directly with one in another, and proteins floating in lipid bilayer membranes have their parts exposed to different sets of molecules) and may therefore make the computations possible. Successful systems biology has to combine reduction and integration [16]. There is no alternative. Electrophysiological cell modellers are familiar with this necessity since the electrochemical potential gradients across membranes are central to function. The influence of these gradients on the gating of ion channel proteins is a fundamental feature of models of the Hodgkin– Huxley type. Only by integrating the equations for the kinetics of these channels with the electrochemical properties of the whole cell can the analysis be successful. As such models have been extended from nerve to cardiac and other kinds of muscle the incorporation of ever finer detail of cell structure has become increasingly important.

THE DIFFERENTIAL VIEW OF GENETICS

These points are so obvious, and have been so ever since electron microscopes first revealed the fine details of those intricate sub-cellular structures around 50 years ago, that one has to ask how mainstream genetics came to ignore the problem. The answer lies in what I will call the differential view of genetics. At this point, a little history of genetics is relevant. The original concept of a gene was whatever is the inheritable cause of a particular characteristic in the phenotype, such as eye colour, number of limbs/ digits, and so on. For each identifiable phenotype characteristic, there would be a gene (actually an allele—a particular variant of a gene) responsible for that characteristic. A gene could be defined therefore as something whose presence or absence makes a difference to the phenotype. When genetics was combined with natural selection to produce the modern synthesis [26], which is usually called neo-Darwinism, the idea took hold that only those differences were relevant to evolutionary success and all that mattered in relating genetics to phenotypes was to identify the genetic causes of those differences. Since each phenotype must have such a cause (on this view at least) then selection of phenotypes amounts, in effect, to selection of individual genes. It does not really matter which way one looks at it. They are effectively equivalent [17]. The gene'seye view then relegates the organism itself to the role of disposable carrier of its genes [18]. To this view we can add the idea that, in any case, only differences of genetic make-up can be observed. The procedure is simply to alter the genes, by mutation, deletion, addition and observe the effect on the phenotype. I will call this gene-centric approach the 'differential view' of genetics to distinguish it from the 'integral view' I will propose later. To the differential view, we must add an implicit assumption. Since, on this view, no differences in the phenotype that are not caused by a genetic difference can be inherited, the fertilized egg cell (or just the cell itself in the case of unicellular organisms) does not evolve other than by mutations and other forms of evolution of its genes. The inherited information in the rest of the egg cell is ignored because (i) it is thought to be equivalent in different species (the prediction being that a cross-species clone will always show the phenotype of whichever species provides the genes), and (ii) it does not evolve or, if it does through the acquisition of new characteristics, these differences are not passed on to subsequent generations, which amounts to the same thing. Evolution requires inheritance. A temporary change does not matter. At this stage in the argument, I will divide the holders of the differential view into two categories. The 'strong' version is that, while it is correct to say that the intricate

structure of the egg cell is inherited as well as the genes, in principle that structure can be deduced from the genome information. On this view, a complete bottom-up reconstruction might still be possible even without the non-genetic information. This is a version of an old idea, that the complete organism is somehow represented in the genetic information. It just needs to be unfolded during development, like a building emerging from its blueprint. The 'weak' version is one that does not make this assumption but still supposes that the genetic information carries all the differences that make one species different from another. The weak version is easier to deal with, so I will start with that. In fact, it is remarkably easy to deal with. Only by restricting ourselves to the differential view of genetics it is possible to ignore the non-genetic structural information. But Nature does not play just with differences when it develops an organism. The organism develops only because the non-genetic structural information is also inherited and is used to develop the organism. When we try to solve the forward problem, we will be compelled to take that structural information into account even if it were to be identical in different species. To use a computer analogy, we need not only the 'programme' of life, we also need the 'computer' of life, the interpreter of the genome, i.e. the highly complex egg cell. In other words, we have to take the context of the cell into account, not only its genome. There is a question remaining, which is whether the weak version is correct in assuming the identity of egg cell information between species. I will deal with that question later. The important point at this stage is that, even with that assumption, the forward problem cannot be solved on the basis of genetic information alone. Recall that genes need to be activated to do anything at all. Proponents of the strong version would probably also take this route in solving the forward problem, but only as a temporary measure. They would argue that, when we have gained sufficient experience in solving this problem, we will come to see how the structural information is somehow also encoded in the genetic information.

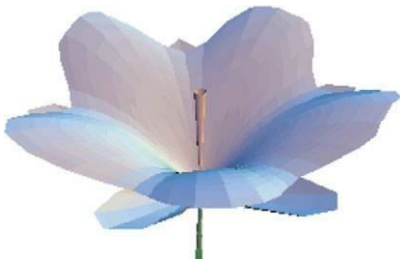

Figure1.Solutions of a generalized Schro¨dinger equation for diffusive spheric growth from a centre.

Neo-Darwinists dismissed Waddington's work largely because it did not involve the environment actually changing individual DNA gene sequences. But this is to restrict acquisition of evolutionarily significant change to individual DNA sequences (the gene's-eye view). On an integrative view, a new combination of alleles is just as significant from an evolutionary point of view. Speciation (defined, e.g., as failure of interbreeding) could occur just as readily from this process—and, as we now know, many other processes, such as gene transfer, genome duplication, symbiogenesis as it might through the accumulation of mutations. What is the difference, from the organism's point of view, between a mutation in a particular DNA sequence that enables a particular phenotype to be displayed and a new

combination of alleles that achieves the same result? There is an inherited change at the global genome level, even if no mutations in individual genes were involved. Sequences change, even if they do not occur within what we characterize as genes. Taking the integrative view naturally leads to a more inclusive view of the mechanisms of evolutionary change. Focusing on individual genes obscuresthis view. In this article, I have been strongly critical of the gene-centred differential view. Let me end on a more positive note. The integral view does not exclude the differential view any more than integration excludes differentiation in mathematics. They complement each other. Genome sequencing, epigenomics, metabolomics, proteomics, transcriptomics are all contributing basic information that is of great value. We have only to think of how much genome sequencing of different species has contributed to evolutionary theory to recognize that the huge investment involved was well worth the effort. As integrative computational biology advances, it will be using this massive data collection, and it will be doing so in a meaningful way. The 'meaning' of a biological function lies at the level at which it is integrated, often enough at the level of a whole cell (a point frequently emphasized by Sydney Brenner), but in principle, the integration can be at any level in the organism. It is through identifying that level and the meaning to the whole organism of the function concerned that we acquire the spectacles required to interpret the data at other levels.

REFERENCES

1 Jacob, F.& Monod, J. 1961 Genetic regulatory mechan-ismsinthesynthesisofproteins.J.Mol.Biol.3,318–356.(doi:10.1016/S0022 836(61)80072-7)

2 Baulcombe,D.2002DNA events.AnRN Amicrocosm. Science297,2002–2003.(doi:10.1126/science.1077906)

3 Feytmans, E., Noble, D. & Peitsch, M. 2005 Genome size and numbers of biological functions. Trans. Comput. Syst. Biol.1,44–49.(doi:10.1007/978-3-540-32126-2_4)

4 Bray,D.2009Wetware.Acomputerineverycell.New Haven, CT: Yale University Press.

5 Kupiec, J. 2009The origin of individuals: a Darwinian approach to developmental biology. London, UK: World Scientific Publishing Company.

6 Bork, P., Jensen, L. J., von Mering, C., Ramani, A. K.,Lee, I.-S.& Marcotte, E. M. 2004 Protein interaction networks from yeast to human.Curr. Opin. Struct.Biol. 14,292–299.(doi:10.1016/j.sbi.2004.05.003)

7 Gsponer, J. & Babu, M. M. 2009 The rules of disorder or whydisorderrules.Progr.Biophys.Mol.Biol.99,94–103. (doi:10.1016/j.pbiomolbio.2009.03.001)

8 Noble, D. 2006The music of life. Oxford, UK: Oxford University Press.

9 Brenner, S., Noble, D., Sejnowski, T., Fields, R. D., Laughlin,S.,Berridge,M.,Segel,L.,Prank,K.& Dolmetsch, R. E. 2001 Understanding complex systems: top-down, bottom-up or middle-out? InNovartis Foun- dation Symposium: Complexity in biological information processing,vol. 239, pp. 150–159.

Chichester, UK: John Wiley.
10 Clancy,C.E.&Rudy,Y.1999Linking ageneticdefectto its cellular phenotype in acardiacarrhythmia.Nature 400, 566–569. (doi:10.1038/23034)
11 Bassingthwaighte, J. B., Hunter, P. J.& Noble, D. 2009 The Cardiac Physiome: perspectives for the future.Exp. Physiol.94,597–605.(doi:10.1113/expphysiol.2008.
044099)

12 Noble,D.2007FromtheHodgkin–Huxleyaxontothevir- tual heart.J. Physiol.580, 15–22. (doi:10.1113/jphysiol.
2006.119370)

13 Gluckman, P.& Hanson, M. 2004The fetal matrix. Evolution, development and disease. Cambridge, UK: Cambridge University Press.
14 Anway,M.D.,Memon,M.A.,Uzumcu,M.&Skinner,
M. K. 2006 Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. J. Androl. 27,868–879.(doi:10.2164/jandrol.106.000349)

15 Barroso, G., Valdespin, C., Vega, E., Kershenovich, R., Avila, R., Avendano, C.& Oehninger, S. 2009 Develop- mental sperm contributions: fertilization andbeyond. Fertil. Steril. 92, 835–848. (doi:10.1016/j.fertnstert.2009.
06.030)

16 Pembrey, M. E., Bygren,L.O.,Kaati,G.,Edvinsson,S.,Northstone,K.,Sjostrom,M.,Golding,J.& ALSPAC study team 2006 Sex-specific, male-line trans- generational responses in humans.Eur.J.Hum. Genet. 14,159–166.(doi:10.1038/sj.ejhg.5201538)
17 Sathananthan, A. H. 2009 Editorial. Human centriole: origin, and how it impacts fertilization,embryogenesis, infertility and cloning. Ind. J. Med. Res. 129, 348–350.
18 Agrawal,A.A.,Laforsch,C.&Tollrian,R.1999Transge- nerational induction of defences in animals and plants. Nature 401, 60–63. (doi:10.1038/43425)
19 Boulinier, T. & Staszewski, V. 2008 Maternal transfer of antibodies: raising immuno-ecology issues.TrendsEcol. Evol.23,282–288.(doi:10.1016/j.tree.2007.12.006)
20 Hasselquist, D. & Nilsson, J. A. 2009 Maternal transfer of antibodies in vertebrates: trans-generational effects on off- spring immunity.Phil. Trans. R. Soc. B364, 51 – 60. (doi:10.1098/rstb.2008.0137)