The scientific paper

Paul M. Matthews, MD, DPhil, FRCP

Co-Director, WT-GSK INTERDISCIPLINARY TRAINING PROGRAMME FOR CLINICIANS IN TRANSLATIONAL MEDICINE AND THERAPEUTICS

Head of the Division of Brain Sciences and Professor of Clinical Neurosciences, Imperial College

3 Oct 2012

What do I want you to be able to think about

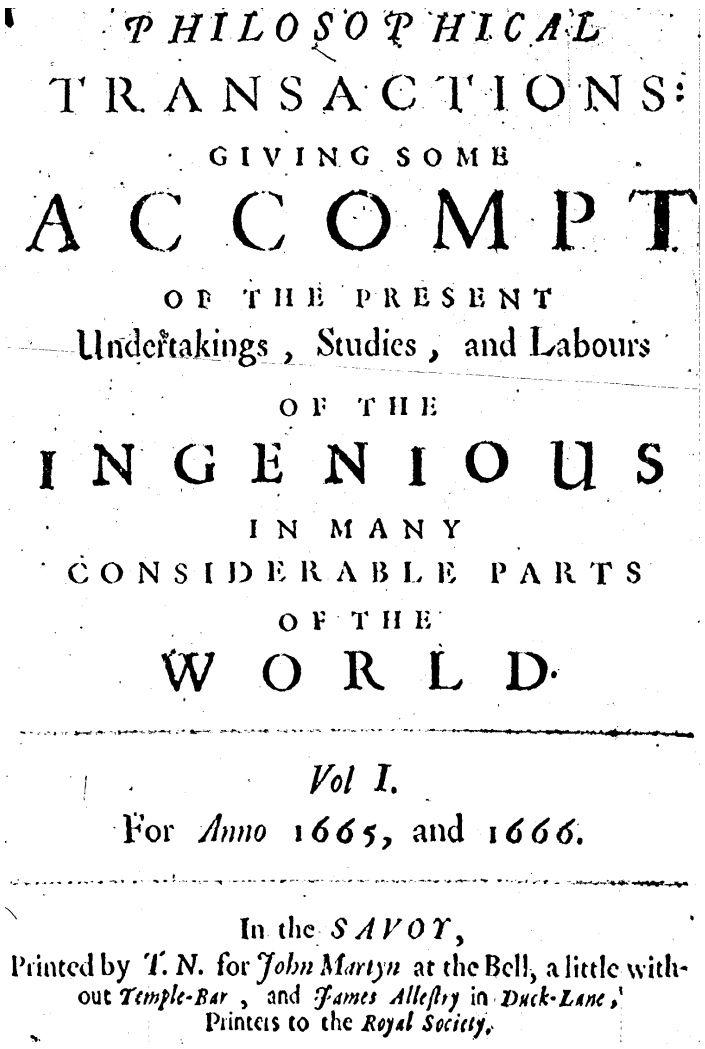
- The elements of a good scientific paper
- Why scientists write papers and why they are published
- How scientific quality and productivity are assessed
- How the paradigm is changing

Why do we perform scientific investigations?

"But so soon as I had achieve the entire course of study at the close of which one is usually received into the ranks of the learned....I found myself embarrassed with so many doubts and errors that it seemed to me that the effort to instruct myself had no effect other than the increasing discovery of my own ignorance."

-Rene Descartes, 1637

The scientific paper has become the principle medium of scientific communications "of record"



PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For *Anno* 1665, and 1666.

In the *SAVOY*,
Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,
Printers to the *Royal Society*.

Scientific Publications

- **The scientific paper as *primary* source material**
- **Technical reports**
- **Reviews as *secondary* source material**
- ***Tertiary* sources such as encyclopedias**

What does a great scientific paper look like?

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[3,4] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

Catchy first paragraph

Develop characters

Point of view

Setting and context

Set up the plot

Set up tension

Climax

Resolution

What should a scientist think about in writing a paper?

- **First- is it worth knowing?**
- **Transforming the question into a specific hypothesis**
  - What makes a good hypothesis?
    - Can it be conclusively rejected? (Is it testable?)
    - Do you see a good story to make the hypothesis compelling?
  - How will testing your hypothesis have an impact?
    - How will it change what we do in the future or our view of this area of science?

How should a paper be constructed?

- What is the logical sequence of experiments (not necessarily the order in which they were done)?
- What were the key observations?
- What are the major confounds for each and how were they addressed?

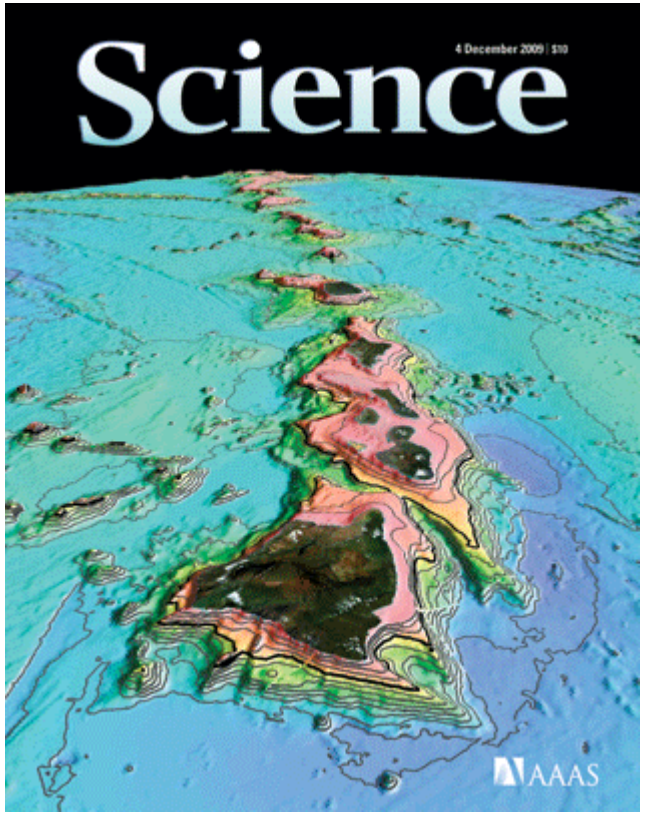
How to think about the range of papers you will read or write

- **Impact**
  - Short term: what is "hot" and what is not
  - Long term: a new view
- **Variations on the narrative form**
  - *Science, Nature, Cell*- the general science journals- make the story more interesting by linking motivation, result and interpretation, point by point
  - Rapid publication journals often demand highly concise, fact-filled reports with brief context-setting and minimal discussion
- **Research reports are not reviews and reviews are rarely for new results**
  - Read primary reports critically- observations depend on the observer
  - Read reviews more critically- they are opinions
- **Innovative experiments are the substance of reports, creative speculation belongs in reviews**

Scientific papers in their social and economic context

- **Science knowledge turned into an economic transaction thanks to print.**
- **Printing was expensive- the scientific paper became a knowledge-compression algorithm**
  - Not just about being right, but about being right in the smallest number of words possible.
- **The medium is the message**
  - With the advent of scientific publishing, we've connected the idea of "knowledge" to "words printed on paper"

The scientific paper as “fact” rather than news



Why do scientists publish?- I

- **Sharing results of their work and the ideas they generate**
- **Testing their ideas**
- **Expanding knowledge**
  - And the “balloon” of ignorance

Why do scientists publish?- II

- **Demonstrating competence: needed for grant funding**
- **Demonstrating productivity: needed for continued grant funding**
- **Demonstrating mastery and reputation: scientific advancement**
- **Supporting their institution's claim for excellence**

Metrics of quality
Impact factors

A measure of the frequency with which the "average article" in a journal has been cited in a particular year or period.

A= total citations in 1992

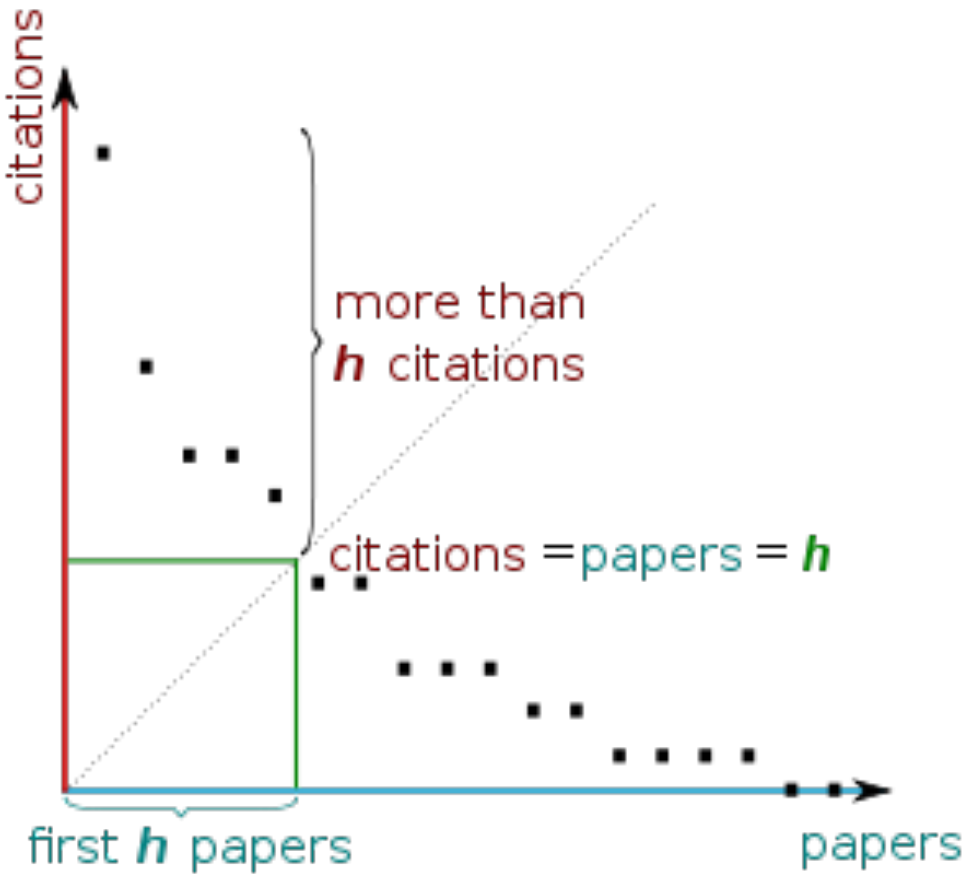
B= 1992 citations to articles published in 1990-91 (this is a subset of A)

C= number of articles published in 1990-91

D= B/C = 1992 impact factor

Assessing productivity and impact

Joel Hirsch's h-index



Sounds simple...but is it?

- The *h*-index does not account for the number of authors of a pape- fields with larger groups, e.g. experimental over theoretical.
- The *h*-index is bounded by the total number of publications- scientists with a short career are at an inherent disadvantage, regardless of the importance of their discoveries.
  - Had Albert Einstein died after publishing his four groundbreaking Annus Mirabilis papers in 1905, his *h*-index would be stuck at 4 or 5.
- The *h*-index does not consider the context of citations. The fastest way to a high h-index is to publish controversial and incorrect results!

The business of scientific publishing

- **A US$7 billion industry**
- **Scientific publishing is the fastest-growing media sub- sector of the past 15 years**
  - An inelastic market
  - Markets define themselves: the implicit compact between scientist and scientific publisher
    - Loyal readership
    - Name recognition
    - “Editorial Boards” that become themselves the currency of scientific advancement

Consequences of the potent mix of ambition and economics: fraud

A model in transition

- **Science output is growing faster than the funding to disseminate it**
- **Paper is an expensive medium to produce, archive and access**
  - Libraries spend US$1.5 on staff costs and other operating expenses for every US$1 on content
    - 1986, libraries spent 44% of their budgets on books compared with 56% on journals
    - 1998, the ratio had skewed to 28% and 72%.

What is the scientific paper of the future?

- **Removing the middle man**
  - Open archives for data
  - Ideas blogs
- **The journal as an aggregator of sub-disciplines**
- **The journal as a synthesizer**
- **Scientific journalism as polemic: creating or influencing a field**