An Introduction to Biological Information

Dave Ussery
Comparative Microbial Genomics Workshop
Pathumthani, Thailand
7 March, 2005
Outline

• Historical background.
• Biological sequences as information.
• The speed of sequencing.
• On the sizes of genomes.
Sub-themes to look for:

Where does biological information come from?

Is life “designed”?

If so, who (or what) is doing the designing?
A Brief History of Biological Information

Aristotle ~350 B.C.

plants
animals
minerals
DNA Reveals Diatom’s Complexity

Diatoms are an enigma. Neither plant nor animal, they share biochemical features of both. Though simple single-celled algae, they are covered with elegant casings sculpted from silica.

Now a team of 45 biologists has taken a big step toward resolving the paradoxical nature of these odd microbes. They have sequenced the genome of *Thalassiosira pseudonana*, which lives in salt water and is a lab favorite among diatom experts. The
Gregor Mendel 1866 genes
T.H. Morgan  1919

Chromosomes
G. Beadle 1930s

one gene

↓

one enzyme
STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES

INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III

BY OSWALD T. AVERY, M.D., COLIN M. MACLEOD, M.D., AND MACLYN MCCARTY,* M.D.

(From the Hospital of The Rockefeller Institute for Medical Research)

PLATE 1

(Received for publication, November 1, 1943)

Biologists have long attempted by chemical means to induce in higher organisms predictable and specific changes which thereafter could be transmitted in series as hereditary characters. Among microorganisms the most striking example of inheritable and specific alterations in cell structure and function that can be experimentally induced and are reproducible under well defined and adequately controlled conditions is the transformation of specific types of Pneumococcus. This phenomenon was first described by Griffith (1) who succeeded in transforming an attenuated and non-encapsulated (R) variant derived from one specific type into fully encapsulated and virulent (S) cells of a heterologous specific type. A typical instance will suffice to illustrate the techniques originally used and serve to indicate the wide variety of trans-
"We believe a gene - or perhaps the whole chromosome fibre - to be an periodic solid."

"...For an illustration, think of Morse code..."

"What is Life?" by Erwin Schrödinger (Cambridge University Press, 1944)
Watson & Crick 1953

James Watson

Francis Crick
**DNA is like Coca-cola!**

<table>
<thead>
<tr>
<th>Coke</th>
<th>DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Water</td>
</tr>
<tr>
<td>Sugar (sucrose)</td>
<td>(deoxyribose)</td>
</tr>
<tr>
<td>Phosphate acid</td>
<td>(PO₄) backbone</td>
</tr>
<tr>
<td>Caffeine</td>
<td>Bases</td>
</tr>
</tbody>
</table>
Where do the DNA bases come from?
DNA bases will spontaneously stack on top of each other and form a helix!
Thermodynamically, the **very most important** interaction is base stacking.

Properties of individual dinucleotide base pairs

<table>
<thead>
<tr>
<th>Dinucleotide</th>
<th>Stacking energy (Kcal/mol bp)</th>
<th>twist angle (degrees)</th>
<th>helix pitch (bp/turn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GC).(GC)</td>
<td>-14.59</td>
<td>40.0</td>
<td>9.0</td>
</tr>
<tr>
<td>(AC).(GT)</td>
<td>-10.51</td>
<td>34.4</td>
<td>10.5</td>
</tr>
<tr>
<td>(TC).(GA)</td>
<td>-9.81</td>
<td>36.9</td>
<td>9.8</td>
</tr>
<tr>
<td>(CG).(CG)</td>
<td>-9.61</td>
<td>29.8</td>
<td>12.1</td>
</tr>
<tr>
<td>(GG).(CC)</td>
<td>-8.26</td>
<td>33.7</td>
<td>10.7</td>
</tr>
<tr>
<td>(AT).(AT)</td>
<td>-6.57</td>
<td>32.1</td>
<td>11.4</td>
</tr>
<tr>
<td>(TG).(CA)</td>
<td>-6.57</td>
<td>34.5</td>
<td>10.3</td>
</tr>
<tr>
<td>(AG).(CT)</td>
<td>-6.78</td>
<td>27.9</td>
<td>13.0</td>
</tr>
<tr>
<td>(AA).(TT)</td>
<td>-5.37</td>
<td>35.6</td>
<td>10.1</td>
</tr>
<tr>
<td>(TA).(TA)</td>
<td>-3.82</td>
<td>36.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>
360° = one helical turn

10.5 bp per turn

34.3° twist angle (rotation per residue)

35.7 Å Helix Pitch

34.3° Base Pair Tilt

3.4 Å Axial Rise

20 Å Helix Diameter
Cells Obey the Laws of Chemistry

Until recently, heredity has always seemed the most mysterious of life’s characteristics. The current realisation that the structure of DNA already allows us to understand practically all its fundamental features at the molecular level is thus most significant. We see not only that the laws of chemistry are sufficient for understanding protein structure, but also that they are consistent with all known hereditary phenomena.

ห้ามนำอาหาร และเครื่องดื่มเข้าห้องอบรม
DNA

5'...GATCTAGCGATGCCGATGAAACATGATCATG...3'
3'...CTAGATCGCTACGGCTACTTTGTACTAGTAC...5'

transcription

mRNA

5'...GAUCUAGCGAUGCCGAUGAAAACAUGGAUGCAUG...3'

translation

Protein

N met-pro-met-lys-his-his-his...C
The Central Dogma of Molecular Biology

DNA → Information → Replication (DNA passes coded information with replication)

RNA → Information → Transcription
RNA synthesis: coded information passed into RNA during transcription

RNA → Information → Translation
Messenger RNA carries coded information to ribosome during protein synthesis

Ribosome
Protein

The buck stops here. Proteins refuse to give away any information.

Proteins provide structure and help carry out almost all biological activity.
**Biological Sequences as Information**

**DNA sequences as information**

1. DNA sequence can code for amino acid sequences (mRNAs)

2. The DNA sequence can code for stable RNA sequences
   - tRNA
   - rRNA
   - snRNA
   - telomerase RNA

3. The DNA sequence can code for protein binding sites

4. The DNA can code for architectural information
   - intrinsic DNA curvature
   - nucleosome positioning

5. The DNA can code for structural / stability information
   - transcription initiation
   - origins of replication
   - mutational "hot spots"
bioinformatics, n.

The science of information and information flow in biological systems, esp. the use of computational methods in genetics and genomics.

1978 P. HOGEWEG in Simulation 31 90/1 Since 1970 she has been a staff member at the Subfaculty of Biology of the University of Utrecht, with her main field of research in bioinformatics. 1985 Jml. Theoret. Biol. 113 719 (heading) Tumor escape from immune elimination... R. J. De Beer, Bioinformatics Group, University of Utrecht. 1986 Philos. Trans. Royal Soc. A. 317 324 The area of modelling mutants from a known structure has been revolutionized by the latest tools of molecular graphics... This is a key element in the whole technology and has attracted much interest (for example, the recent E.E.C. ‘Bioinformatics’ programme). 1987 Science 4 Sept. 1108/3 One of the latest developments [at the European Molecular Biology Laboratory] has been the creation of a new research program in bioinformatics. This is intended to bring together research in computing science, structural biology, and molecular genetics. 1996 Fast Company Aug.-Sept. 32/3 A lot of breakthroughs in medicine will come out of the efforts of bioinformatics. 2001 N.Y. Times 4 Jan. B6/2 The hope...is to make New York a leader in cutting-edge fields like bioinformatics, in which computers are used to decipher genes and proteins.
**Biological Sequences as Information**

**RNA sequences as information**

1. The mRNAs can contain several different levels of information:
   - specifies amino acid sequence for proteins
   - localisation signals for WHERE the protein will be made
   - stability signals to determine HOW MUCH protein is made
   - splice sites
   - editing sites

2. The tRNAs code for the genetic code - same in all living organisms (n.b. diff. in mitochondria)

3. The rRNAs code for the structures of ribosomes

4. Other RNA/protein complexes
Biological Sequences as Information

Protein sequences as information

1. The protein sequence can code for an "active site" for enzymes

2. The protein sequence can code for structural roles:
   microtubules  myosin  collagen  etc.

3. The protein sequence can code for ion channels/pumps

4. The protein sequence can code for localisation information

5. The protein sequence can code for modification sites
Biological Sequences as Information
Telomeric repeat unit

Hairpin at telomere

Normal hydrogen bonds

Unusual G-to-G hydrogen bonds
DNA helix

Chromosomal DNA

Synthesis of complementary DNA strand

Internal guide RNA template

Telomere repeat

Telomerase

Telomerase elongates the template DNA strand at the 3' end
Summary (so far!)

DNA

RNA

Protein

Sequences ➔ Structure ➔ Function
AND NOW FOR SOMETHING COMPLETELY DIFFERENT!
Where does biological complexity come from?
E. Schrödinger 1943
Speculations on the physics of biology....

"What is Life?" by Erwin Schrödinger
(Cambridge University Press, 1944)

Where does order come from?

1. order ➔ order
   "... think of Morse code..."

2. disorder ➔ order
   "... organisms eat negative entropy..."

self organization - non-linear systems
(difficult to model), far from equilibria
Where does order come from?

2. disorder → order

“... organisms eat negative entropy...”

self organization - non-linear systems
(difficult to model), far from equilibria

Three examples of emergent properties:
1. The DNA helix

- **360°** = one helical turn
- **10.5 bp per turn**
- **34.3°** twist angle (rotation per residue)
- **Helix Pitch**: 35.7 Å
- **3.4 Å** Axial Rise
- **Helix Diameter**: 20 Å
- **Minor Groove**
- **Major Groove**
- **Base Pair Tilt - 6°**
- **Helix Diameter**: 20 Å
2. Hurricanes
3. Galaxies
Mathematics and Its Applications

Rainer Feistel and Werner Ebeling

Evolution of Complex Systems
Selforganization, Entropy and Development

Kluwer Academic Publishers
Information Theory, Evolution and the Origin of Life

HUBERT P. YOCKEY

CAMBRIDGE
William A. Dembski

NO FREE LUNCH

Why Specified Complexity Cannot Be Purchased without Intelligence
Discussion question:

Where does biological complexity come from?

Do cells obey the laws of chemistry?

The answer depends on where you live!

“Yes” if you live in Europe.
“No” if you live in the U.S.
??? if you live in the Thailand