Predictability and Complex Systems: the Notion of Scientific Knowledge in Biology

Silvia Morante
Department of Physics
University of Rome "Tor Vergata"

Paris 22 June 2011
Premise

The “scientific method” seems to move in an area framed by the following two crucial issues

Falsifiability and Reductionism

The falsifiability paradigm

A good scientific theory

✓ Should incorporate (or suggest) the experiment(s) that could possibly falsify it

but it can never fulfill the request of being “proved” once forever

“No amount of experimentation can ever prove me right; a single experiment can prove me wrong.”
(Albert Einstein, lettera a Max Born del 4 dicembre 1926)
The reductionist paradigm

In order to be able to answer the questions that are still waiting for a solution, it is "only" necessary to collect a "sufficiently large" amount of experimental information.

"There is still much lack of knowledge about this advanced level of organization of matter, we call 'life', and its novel non-material consequences. However, at this stage it is simply 'lack of knowledge', and not 'discrepancy' with the present concepts of physics."

Living systems (apparently) violate both paradigms

- very few good models (none of them can be called “theory”)
- low reproducibility

Can they be falsified?

Many irreducible “single” cases

Biology is not a Science?
Reductionism *versus* complexity

**Biology vs Physics**
*(the physicist point of view)*

Compare the situation we face in the investigation of Biological Systems:

- “*Complex*” structures ruled by (to date) unknown *macro-laws*
- Powerful and not so expensive experimental techniques
- Huge amount of data
- Inadequate models: not clear boundary between “*micro*” and “*macro*” systems

With, at the other extreme end what we find in Physics of elementary particles:

- Theoretically “*simple*” systems ruled by “*elegant*” *micro-laws*
- Extremely expensive and very difficult experiments
- Very few new experimental data (LHC is at work!)
- Very good models (often confidentially called “*theories*”)
Physics has found a (successfully) way by moving from complicated toward more and more elementary structures

matter → atoms → nucleons → quarks → ???

under the guidance of a “radical reductionism” paradigm according to which

FUNDAMENTAL LAWS are at the basis of ELEMENTARY PHENOMENA

An attitude that, has been very effective and very successful, think only to the field of high energy physics.
This is apparently not the case in new emerging fields like for example:

- **Dynamical Systems**
  - weather forecasting
  - enzymatic reactions
  - fluid-dynamics (turbulence)
  - keywords: “non-linearity” and “chaos”.

- **Disordered Systems**
  - glasses, spin-glasses
  - keywords: “frustration” and “disorder”

- **Biological Systems**
  - keywords: “complexity” (and maybe also all the others)
“Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe sind blind”

I. Kant “Kritik der reinen Vernunft"

“Ideas without content are empty, experiences without the abstraction are blind”

“Theory without experiment is useless, Experiments without theory do not give any knowledge”
Implications of the concept of reductionism

In Fundamental Physics

- In Physics one (improperly) talks about Relativity Theories, like Field String

- They are indeed Models, formulated in the language of Mathematics, thus inheriting from it the necessary internal logical consistency

- Everyday phenomena, like friction in Mechanics, are considered (conceptually) irrelevant (up to a certain point: without dry friction a wheel can't roll and a car can't move, without fluid friction a plane can't fly and a sailboat can't sail,...!)

- The deeper is the comprehension the simpler the theories become.
Models (nobody would dare to call them theories) show the tendency of becoming more and more complicated as they evolve (not simpler!)

with the obvious limitation that a model to be useful should be fairly simpler than the system it wants to model!

Two crucial questions then arise

- When do we have the right to consider a given biological phenomenon or system, as being understood?
- Which level of knowledge (which in turn implies a given level of prediction ability) can be considered as satisfactory?
Implications of the concept of reproducibility

Central dogma of Physics

Theories (better models) must be validated by reproducible experiments.

But in Biology

Reproducibility is severely compromised (at risk):

- In biological systems we have too many (much greater than 1) relevant degrees of freedom (dof's) to keep under control;
- In many cases a too detailed description is not of interest (disorder and redundancy)

Models, if any, are (very) approximated and often (unacceptably) complicated.
Elementary objects are characterized by a small # of properties

All the elementary objects of a given type are similar (electrons)

Simple Laws (theories) apply to elementary objects

Experimental reproducibility is required and a rigid determinism follows

Complex systems have many dof’s and many relevant components

Classes of systems (e.g. the class of neuronal cells, the class of liver cells, or rather the larger class of all nucleate cells). Systems of the same class share common features

Models should provide a mathematical description of the features common to all the systems of the class.

Models are expected to provide average information in terms of probability distribution.

Such expectation values should be compared with the average of many experiments carried out on systems of the class.
The fundamental point is the level of homogeneity among the objects belonging to the same class.

- the better (worse) is the accuracy
- the simpler (more complex) is the model
- the cleaner (more involved) is its mathematical description
- the more precise (the more vague) is the information one can get.

Pictorially:

- Low level of complexity
- High level of complexity

Model adaptability

**Amount of information one can get when dealing with complex systems.**
Does the definition of complexity unmistakably fit with biological systems?

The “numerosity” (many dof’s) is for sure fulfilled. As an example a 70 kg human being is made of approximately $7 \times 10^{27}$ atoms\(^(*)\), that is to say, 7 followed by 27 zeros $7,000,000,000,000,000,000,000,000,000,000$.

Numerosity is necessary, but not sufficient and it is shared by also very simple systems. Ex. a $1m^3$ box of a gas at $25 \, ^{\circ}C$ and 1 atm contains $\sim 2.5 \times 10^{25}$ molecules.

What exactly (mathematically) means “complexity”? Certainly not entropy.

\(^(*)\) $4.7 \times 10^{27}$ H (1p, 1e) + $1.8 \times 10^{27}$ O (8p, 8n, 8e) + $7.0 \times 10^{26}$ C (6p, 6n, 6e). Adding them up we obtain a total of $2.3 \times 10^{28}$ p + $1.8 \times 10^{28}$ n (+ $2.3 \times 10^{28}$ e).
But...

Complexity needs a certain amount of “organization” ...

...“ceteris paribus” organization tends to decrease entropy

the entropy of a complex system cannot be as large as it could be (given the number of its degrees of freedom).
When a system owns the second condition necessary for complexity, namely heterogeneity?

This happens when its elements may be subdivided in many different classes each made of many elements.

For this condition to be fulfilled numerosity is not sufficient, but still necessary

A good example is the cell, i.e. the minimal system able to autonomously carrying on its self-replication.
Several hundred of distinct human cell types

One of the many possible type of classification

Cells
- Prokariotic
- Eukaryotic

Plants
Animals

Human

(50 ÷ 75) x 10^{12} cells

Human cells

- endoderm
- epithelial
- nervous
- mesoderm

Blood and immune system

(≈ 10^6 antibodies, ≈ 10^8 lymphocytes)

B-cell monocye

20 classes
Workable definition of complexity

“the more one can say about a class of systems, the more the systems of that class are complex”

1. The ensemble of random strings of characters

Once the ensemble is given nothing more can be said about it and its elements cannot be further subdivided.

all cases belong to the same class

→ very simple (class of) systems

2. The set of quantum states of an ensemble of spins at $H = 0$, $T \sim 0$

Two classes: [all particles with spin up] [all particles with spin down]

→ very simple (class of) systems
3. The ensemble of strings of characters each representing one of the written book

Numerosity: all the books published in modern history ~ $1.3 \times 10^8$(*)

[Note that having at disposal ~ 30 different characters one can “build” $30^{100 \times 3000}$ “books” of 100 pages, each made of ~ 3000 characters. This means ~ $10^{400,000}$ “books”!!]

By far the largest part of them are of course strings of character with no meaning whatsoever

But among them there are all the books of 100 pages ever written in human history in all languages (that employ the same set of characters), all those that will be written in the future and even all those that will never be written.

(*) The number of books written from the origin of writing cannot be more than a factor 10 larger than this.
Heterogeneity → building up complicated classification trees

The set of books can be subdivided in classes according to any of the following characteristics:

- the **language**: English, Italian, French, Russian…
- the **genre**: poetry, novel, essay, …
- the **plot**: love literature, thrillers, …
- …
- the **weight**
- the **colour**
- …
- or any combination of them

- the California Department of Education ([http://www.bsfcs.org/forums/green/Literature/Genres.htm](http://www.bsfcs.org/forums/green/Literature/Genres.htm))
- Wikipedia
### Complexity of classification

<table>
<thead>
<tr>
<th>Classification Categories $T_k$</th>
<th>Classes $C_{k,l_k}$</th>
<th>Number of classes in each category $m_k$</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Italian</td>
<td>French</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td><em>poetry</em></td>
<td><em>novel</em></td>
<td><em>essay</em></td>
</tr>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plot</strong></td>
<td><em>love literature</em></td>
<td><em>thriller</em></td>
<td></td>
</tr>
</tbody>
</table>

$T_k$=$1,2,3$  \hspace{0.5cm} $l_1$=$1,2,3,4$  \hspace{0.5cm} $l_2$=$1,2,3$  \hspace{0.5cm} $l_3$=$1,2$
Total number of possible classes = $N_c = m_1 \times m_2 \times m_3 \times ...$

The final classification depends on the choice of hierarchy

Total number of possible classes = $N_c \neq m_1 + m_2 + m_3 + ...$
An intrinsic difficulty/opportunity is the choice of classification hierarchy

Solved/exploited by having a good prejudice about the property we want to let emerge from our classification

When dealing with tree classification the natural choice is often time distance

Natural languages
An intrinsic difficulty is the choice of classification hierarchy

Solved by having a **good prejudice** about the **property** we want to let **emerge** by means of our classification

When dealing with tree classification the natural choice is often that of distance in time.

**Phylogenetic tree of Life**

---

**Bacteria**
- Spirochetes
- Proteobacteria
- Cyanobacteria
- Planctomyces
- Bacteroides
- Cytophaga
- Thermotoga
- Aquifex

**Archaea**
- Green Filamentous bacteria
- Gram positives
- Methanosarcina
- Methanobacterium
- Methanococcus
- T. celer
- Thermoproteus
- Pyrodicticum

**Eucaryota**
- Entamoebae
- Slime molds
- Animals
  - Fungi
  - Plants
  - Ciliates
  - Flagellates
  - Trichomonads
  - Microsporidia
  - Diplomonads

---

**we are here**

**genotypic similarities.**
From where everything began

New tree structures appear
Blowing up the class of Myosins

An Unrooted Phylogenetic Tree of the Myosin Superfamily
Tony Hodge, MRC-LMB
Jamie Cope, UC Berkeley
July 2000
One among the ~ 30000 different human proteins

1935 aminoacidic residues ~ 20000 atoms

MGDSEMAVFGAAAPYLRKSEKERLEAQTPRPFDLKKDVFPDDKQEFVYKAVSKRSEKVGVT
AETEYGTKTVKQDQVMQNPQPKFDKIEDMAMLTFHLHPAVLYNLKDRYGWSMIYTSGL
FCVTNPYKWLPPVETYPEVAARYGKRSSEQAPPHFSISDNSAYQMLTDFRPQSLITGES
GAGKTNTKAVIQYFVAVIAIGDRSKKDQSGKTELDQIQANPALEAGFNAKTVRNNDN
SSRFKFR1HRFGATGKLASSADITYELLEKSRVFQKALAEYHIYIQILSNKPEFELLM
LILTNTNFYDAISQGETTVASIDDAEELMTINDATNFREVDFGTSSEEEKSNYKTLGA1MHFG
NMKFKLQREEAQAEPDGTEEADKSAYLMGLNSADLLKGLCCHRPRVKVQNYEVTQKGNVQQV
IYATGALAKAVYERMNCWVTRINATLETQPRQYFIGVLDIAIFDFNSFQLCINF
TNEKLIQQFNNHFVLEEQEYKKEGIEWTFIDFGMLAQICDLIEKFGIMSEELCECMF
PKATDMTFKAKLDFHNGKSANFQPRIKGKEAHFSLIHYAGIDVQNIIGWAQKNNKDP
LNETTVGLYQKSLKLSTLFLANYAGADAPIEKGKAGKQGSSQTVSALHRENKLMT
NLRSTPHFVRCIIIPNETKSPGMNPLVMHQLRCNGVLEGIRICRNCFPNRIYGDFRQ
RYRILNPAAPIPEQFIDSRKGAELIISSLDIDHNQYKGFHTKFKAGILGELMEERDRE
LSRITIITRAQSRSRQAMHRYKLLEERRDSSLVQWNIRAQMGVKNPMKLYFKIKPLL
KSAEREKEMASMKFERTKLKEALKESEARRKELEEMKVSSLQKNDLQVIQAEQDNLAD
AERERDQIKNQLEAKVKEFVEMDEEMEAMNLEAKTKRLEDECSELKIDDLTLE
LAKVEKEKHATENKVKNLTEEMAGLDEIAIATLTKKEKALKQEAHQQALQDDLQAEDEKVNTL
TAKAVKLYEQVDDLEGSLQEGBKVRMDLERRAKRLEQDDLQKEQVTOESMFLDEKQQLDLR
KKKFDALNNAIERDEQALGSLQKQKELQARIIEEELAEartaERAVKELRSSL
RELEKSERLEEAAGGATSVQLMNNKREAETFQKMRDLEELHGLNQMDMEAAARLKPHDVS
AELQGQIDNLQVRKQKLEKESEFKLLELDVTSNMEIQIKIKALENQCMRTEQDMNEHR
SKAETQSRVNDLTSQRAKLQITENGETLSRQDEKEALISQSLRKGTLTYQOLLEDKQLE
EEVAKAKNALAHALQASRHCDDLRQLEYEEEAEKAELQRVLKSAVESQAQWTQKETYDAI
QRTELEEAKKALQRLQAEAEAVEAVNACKSELTKHRQNEQIIIDLMVQVSNAAA
ALLDKQFRDKIIAEWIKQYEEESQSELESSQKARESLSLFLKKNAYESELEHLETFKR
ENKNLQEEISDLTEQLGSSGTKIHELEKVRQLEAEKMLQSALEEAASELEHEEGKILR

The a.a. sequence of a given individual

VEEAVQECRNLTEAEKAKAITDAAMMAELKKEQDTSAHLERMKKNEQTKIDLOQHRLEDA
EQLAKGGKKSQKELQAREVLENELEAEQKRMNASEVKGMRKSKMRRIELTYQTEDRKN
LLRLQDLVDKLQLKVKAYKQRAAEAEQQANTNLKSFKRSKVQHELDEAERADIAESQVNKL
RAKSRDIGTKGLNEE
<table>
<thead>
<tr>
<th>Aminoacid</th>
<th>Chemical formula</th>
<th>Molecular weight, g/mol</th>
<th>C</th>
<th>H</th>
<th>N</th>
<th>O</th>
<th>S</th>
<th>Tot</th>
<th>Tot Lat</th>
<th>Tot -H</th>
<th>Tot-pept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoleucine</td>
<td>C₆H₁₃NO₂</td>
<td>131.1736</td>
<td>6</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td></td>
<td>22</td>
<td>13</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Leucine</td>
<td>C₆H₁₃NO₂</td>
<td>131.1736</td>
<td>6</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td></td>
<td>22</td>
<td>13</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Lysine</td>
<td>C₆H₁₄N₂O₂</td>
<td>146.1882</td>
<td>6</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td></td>
<td>24</td>
<td>15</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>Methionine</td>
<td>C₅H₁₁NO₂S</td>
<td>149.2124</td>
<td>5</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>23</td>
<td>14</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>C₅H₁₁NO₂</td>
<td>165.1900</td>
<td>9</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td></td>
<td>23</td>
<td>14</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>Threonine</td>
<td>C₄H₉NO₃</td>
<td>119.1197</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>16</td>
<td>7</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>C₁₁H₁₂N₂O₂</td>
<td>204.2262</td>
<td>11</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td></td>
<td>27</td>
<td>18</td>
<td>15</td>
<td>24</td>
</tr>
<tr>
<td>Valine</td>
<td>C₅H₁₁NO₂</td>
<td>117.1469</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td></td>
<td>19</td>
<td>10</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Arginine</td>
<td>C₅H₁₄N₂O₂</td>
<td>174.2017</td>
<td>6</td>
<td>14</td>
<td>4</td>
<td>2</td>
<td></td>
<td>26</td>
<td>17</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Histidine</td>
<td>C₅H₁₄N₂O₂</td>
<td>155.1552</td>
<td>6</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td></td>
<td>20</td>
<td>11</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Alanine</td>
<td>C₃H₇NO₂</td>
<td>89.0935</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td></td>
<td>13</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Asparagine</td>
<td>C₄H₉N₂O₃</td>
<td>132.1184</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td></td>
<td>17</td>
<td>8</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Aspartate</td>
<td>C₄H₇NO₄</td>
<td>133.1032</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td></td>
<td>16</td>
<td>7</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C₃H₇NO₂S</td>
<td>121.1590</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Glutamate</td>
<td>C₅H₁₀N₂O₄</td>
<td>147.1299</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td></td>
<td>19</td>
<td>10</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>Glutamine</td>
<td>C₅H₁₀N₂O₃</td>
<td>146.1451</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td></td>
<td>20</td>
<td>11</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Glycine</td>
<td>C₂H₅NO₂</td>
<td>75.0669</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td></td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Proline</td>
<td>C₅H₁₂NO₂</td>
<td>115.1310</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td></td>
<td>17</td>
<td>8</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Serine</td>
<td>C₃H₇NO₃</td>
<td>105.0930</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td></td>
<td>14</td>
<td>5</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>C₈H₁₁NO₃</td>
<td>181.1894</td>
<td>9</td>
<td>11</td>
<td>1</td>
<td>3</td>
<td></td>
<td>24</td>
<td>15</td>
<td>13</td>
<td>21</td>
</tr>
</tbody>
</table>

Average: 5.35, 10, 1.45, 2.4, 19.3, 10.3, 9.3, 16.3
Individuality (within a species) given by punctual genetic differences → **Polymorphisms**

**Snips** = mutation of a single base, Single Nucleotide Polymorphisms (SNPs), present in at least 1% of the population.

**Human species**: 99.9% identical bases - in the remaining 0.1% 90% are snips

**Mutations:**
- **harmless** (modify the phenotype, for example the color of eyes or hairs);
- **harmful** (even single mutations can cause severe diseases, ex. diabetes, some type of tumors, cardiac diseases, haemophilia, Huntington's disease etc.);
- **latent** (not directly harmful, but can enhance the individual tendency to develop certain pathologies)

**Snips frequency** $1/1000 \div 1/100$ bp.

**Myosin**: 2000 a.a.'s = 6000 bp’s = (about) 6 snips → $4^6 = 4096 \sim 4000$ classes of individuals with a different snips sequence in Myosin gene (which is only one out of the 30000 different genes in our organism!!)

$6 \times 10^9$ individuals in the World → (in average) $1.5 \times 10^6$ individuals /class
Snips mapping and localization

Classification according to the snips profile

Paris 22 June 2011
Non-pathological Myosins → almost the same structure (independently of their snips content)

Myosin Structure

Myosin function
The different (due to the snips occurrence) Myosin molecules should have a (substantially) identical structure in order to make a (substantially) identical job

→ in order for Myosin to be “understood” the knowledge of location and quality of snips is essentially unimportant

A single base mutation can give rise (or favor) the appearance of certain disease

→ search for snips can be very relevant (es: Hypertrophic cardiomyopathy is connected with the presence of a single site mutation in the Myosin gene)

⇒ When can one say that one has “understood” Myosin?
⇒ Which level of understanding is satisfying?
⇒ Should one know the relative position of all Myosin atoms in space (i.e. the three-dimensional structure)?
⇒ Should one know the way it works (as shown in the movie)?
⇒ Are both information equally necessary?
⇒ Maybe there is something else causing the disease without modification of the structure or of the way of functioning?
⇒ Am I more interested in the content, weight or color of books?)
Biosystems can be classified in many different ways, according to: their composition, their dimension,…, or even the number of chemical processes they can perform.

Human body: \(\sim 7 \times 10^{27}\)

99% C, H, O and N; 87% H and O;

41 different elements

Mean composition of a thin man of 70 Kg
## cell composition

**Molecular content of a human cell of 20 µm**

<table>
<thead>
<tr>
<th>Type of molecules</th>
<th>mass %</th>
<th>PM (Dalton)</th>
<th>molecules %</th>
<th>molecules #</th>
<th>Types #</th>
</tr>
</thead>
<tbody>
<tr>
<td>water</td>
<td>65</td>
<td>18</td>
<td>98.73</td>
<td>1.74x10^{14}</td>
<td>1</td>
</tr>
<tr>
<td>Inorganic</td>
<td>1.5</td>
<td>~200 types of cells</td>
<td>1.31x10^{12}</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lipids</td>
<td>12</td>
<td>700</td>
<td>0.475</td>
<td>8.4x10^{11}</td>
<td>50</td>
</tr>
<tr>
<td>Organic</td>
<td>0.4</td>
<td>250</td>
<td>0.044</td>
<td>7.7x10^{10}</td>
<td>~200</td>
</tr>
<tr>
<td>molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proteins</td>
<td>28</td>
<td>580000</td>
<td>0.011</td>
<td>1.9x10^{10}</td>
<td>~5000</td>
</tr>
<tr>
<td>RNA</td>
<td>1.0</td>
<td>1x10^{6}</td>
<td>3x10^{-5}</td>
<td>5x10^{7}</td>
<td>3</td>
</tr>
<tr>
<td>DNA</td>
<td>0.1</td>
<td>1x10^{11}</td>
<td>3x10^{-11}</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
<td>100</td>
<td>1.76x10^{14}</td>
<td></td>
</tr>
</tbody>
</table>

1 Da (Dalton) = 1 atomic unit = \( m_a^{(12C)/(12 \times 1.660540 \times 10^{-27} \text{ kg} \sim \text{hydrogen mass}) \text{ dimensionless unit} \)
their dimension

Proton

10^{22}

Earth

Virus

10^9

Blu Wales
Mitochondria + O₂

Citric acid cycle or Krebs cycle

Glyoxylate cycle

CO₂

COOH

C=O

CH₂

COOH
MITOCHONDRIUM

H\(^+\) gradient

NAD\(^+\) pump

NADH

e\(^-\)

Membrane proteins

fats and carbohydrate molecules

citric acid cycle

products

CO\(_2\)

H\(_2\)O

O\(_2\)

NAD\(^+\)
Step 1: Citrate synthase
Step 2: Aconitase
Step 3: Isocitrate dehydrogenase
Step 4: Alpha-ketoglutarate dehydrogenase complex
Step 5: Succinyl-CoA synthetase
Step 6: Succinate dehydrogenase
Step 7: Fumarase
Step 8: Malate dehydrogenase
Play in Concerto

most biological processes require interactions among several actors

interactions may require and/or promote the actors to modify their own structure
When can one say that he has “understood” the Krebs cycle (or any of the many biochemical processes displayed in the “messy” network chart above)?

What level and what kind of knowledge is required?

When this knowledge is sufficient to be predictive?

- Should one know the chemical and three-dimensional structure of all the molecules involved in the process?
- Should one know when, how and at which level they interact among themselves?
- Should one known the details of the surrounding environment?
- When, at what distance (think for instance at the gravity that has a great impact in many biological processes) the influence of the environment can be ignored?
- Or maybe the essential ingredient is something we have not yet recognized, neither the (mean) structure, nor the (mean) way of functioning of the involved molecules?
Growth of GenBank
(1982 - 2008)
All these collected data should be put in relation one with the other

• Inside the cell the gene is replicated, transcripted and translated into proteins (regulation mechanisms involve proteins)

• Into the cell all the methabolic processes necessary to its survival and replication happen

• The cell regulates its interaction with the surrounding environment

... In this complicated panorama, even the to imagine a model is a hard challenge

• All those ingredients that are essential must be included

A not too complicated, at least not as complicated as the system itself model must be reached
When do we decide that one has understood?

The knowledge of the structure and the chemistry of all the molecules involved (DNA, RNAs, ribosome, regulating, activating and inhibiting proteins) is enough to reproduce the phenomenon?

It is enough to predict the result of an experiment?

It depends on the type and level of knowledge (foreseeing) one will consider satisfying.

But it is quite evident that in the case of a complex system “something” different from the simple sum of the parts composing the system emerges.
Omic era • availability of complete genome sequences

complete set of transcripts
complete set of proteins
complete set of phenotypes
localization of all transcripts and proteins

Multidisciplinarity approach

Dealing with complex systems requires a revision/adaptation of

- **Concepts**
- **Methods**
- **Goals**

**Concepts** from radical reductionism to classification & self-organization (emergence of new features)

**Methods** from (strict) determinism to probability distributions & mean values

**Goals** from basic principles & theories to finalized actions & modelizations
"The emergence of complexity" Jochen Fromm