Quantitative Systems Biology:

New Opportunities, Challenges and Directions

Maria I. Klapa

Foundation of Research and Technology-Hellas, Institute of Chemical Engineering and High Temperature Chemical Processes (FORTH-ICEHT), Patra, Greece

The Genomic Revolution 1953-2003

No. 4356 April 25, 1953 NATURE 737

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).

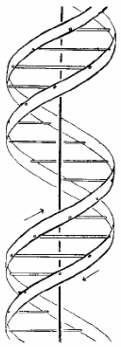
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

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.

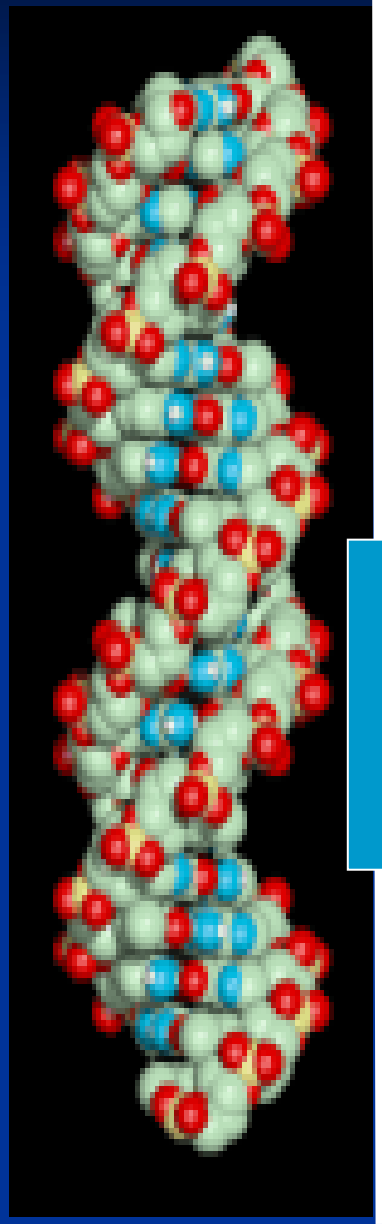
The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these 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

Timetable of the Genomic Revolution



1944 Avery et. al. provide evidence that DNA is the hereditary material

1953 Watson & Crick publish the double helical structure of DNA

1963 Nirenberg elucidates the genetic code

1964 Perutz connects the 3-D structure of proteins to the amino acid sequence

1972-74 Development of recombinant DNA technology

1975-85 Development of techniques for rapid DNA sequencing

1990 Initiation of the Human Genome Project

1995 First complete genome published

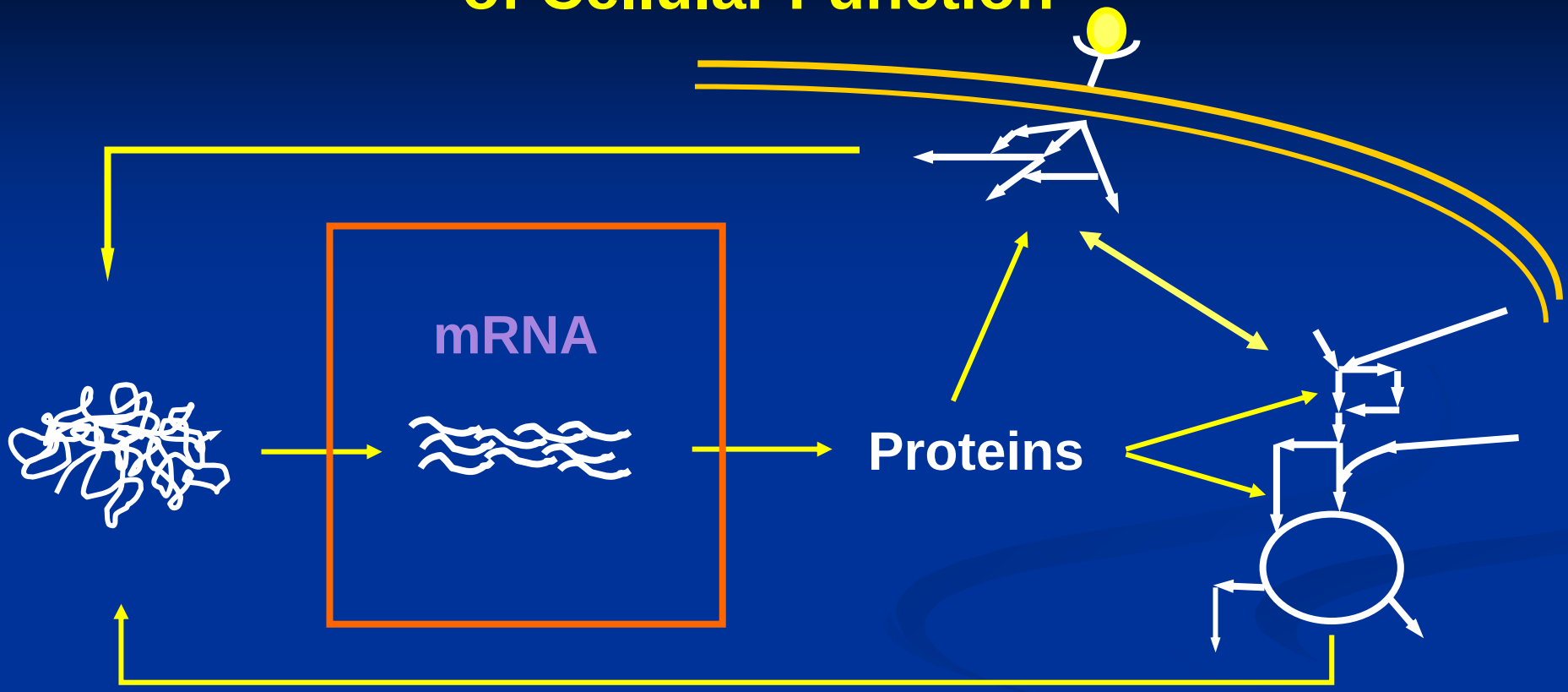
1996 Affymetrix publishes about GeneChip

2003 Human Genome Project completed

Viva la revolution!

We can now begin to look at gene expression and protein activity in the context of networks and systems of interacting genes and gene products

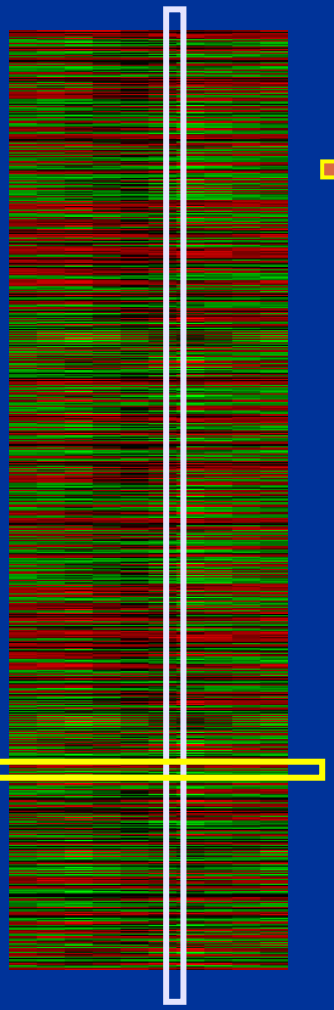
DNA microarray: The Genomic Fingerprint of Cellular Function



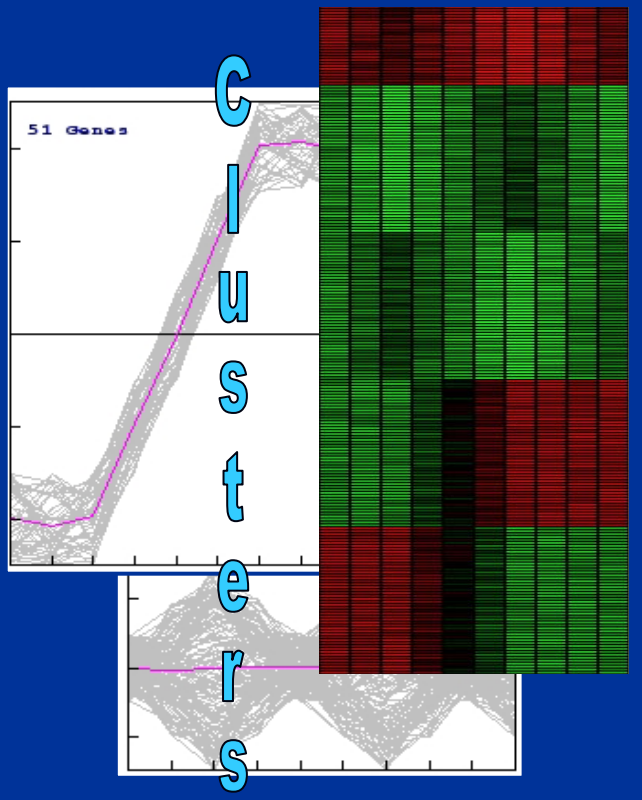
- Comparative genomics
- Functional genomics
- Characteristic genes
- Construction of metabolic network

Gene expression profiling

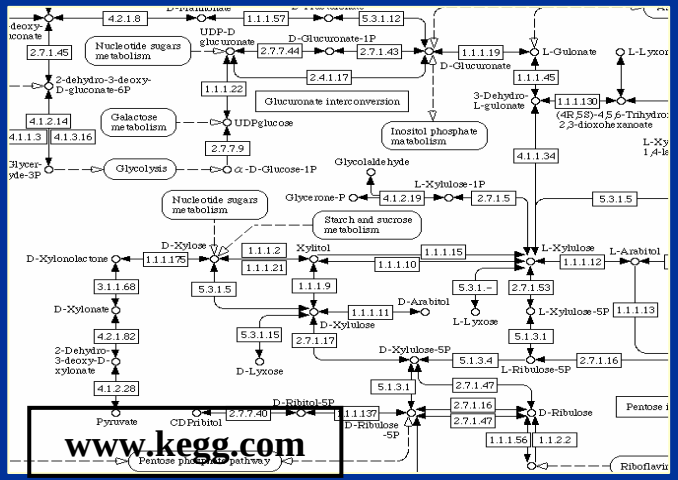
DNA Microarray



Clustering Algorithms



Pathway Window

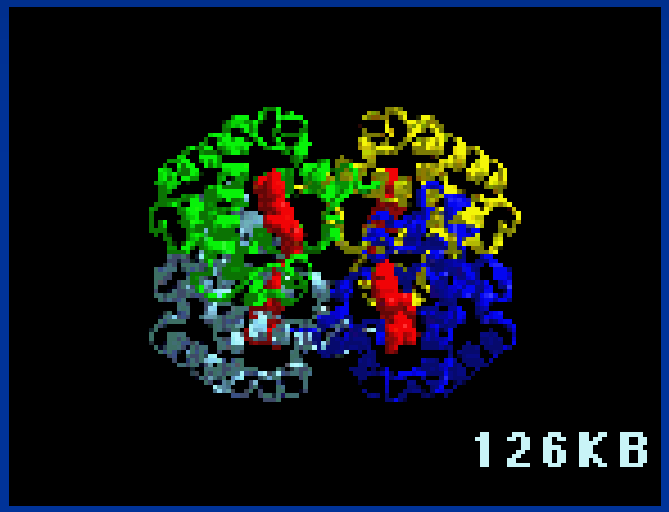


Challenges

- Development of clustering algorithms to better differentiate between gene clusters
- Clever experimental design
- Semi-quantitative experimental technique – determining “big” trends
- Gene networks are not available

Design of protein effectors

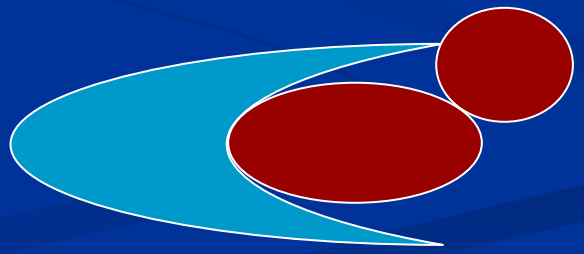
Knowing just the amino acid sequence is not sufficient to design a protein effector



Determine the 3-D structure of the protein

using X-Ray crystallography

Design the effector (ligand) that minimizes the energy level of the ligand – receptor complex



Challenges

- High- throughput analysis of protein 3D structure
- Development of algorithms (filters) to determine the relationship between amino acid sequence and 3D structure (of the active site)
- Development of high-throughput analysis of protein function and family
- High-throughput development of specific antibodies
- Development of protein microarrays
- High-throughput design and development of drugs

Viva la (Systems Biology) revolution!

It is becoming increasingly clear that a comprehensive analysis of biological systems requires the integration of all fingerprints of cellular function:

- gene sequence
- gene expression
- protein 3D structure
- total protein production
- metabolic output (macroscopic)
- *in vivo* enzymatic expression (activity)

Few integrated analyses so far…

… primarily because we lack both the conceptual understanding and the computational tools that would allow the identification of cause-effect relationships between the gene and protein regulation and phenotypic profiles

Algorithm development has to go hand in hand with experiments!

- It is very important to carefully design experiments that can provide **comparable** gene expression, protein production and metabolic function data that can lead to useful results
- Increase and improve the experimental techniques and methodologies for the **quantitative** measurement of the cellular physiological state at each level of cellular function

Integrated Databases

- Development of extended relational databases that can effectively capture and integrate genomic, proteomic and phenotypic data
- Development of protocols for submitting experimental data to public databases that will allow the integration of information from various labs

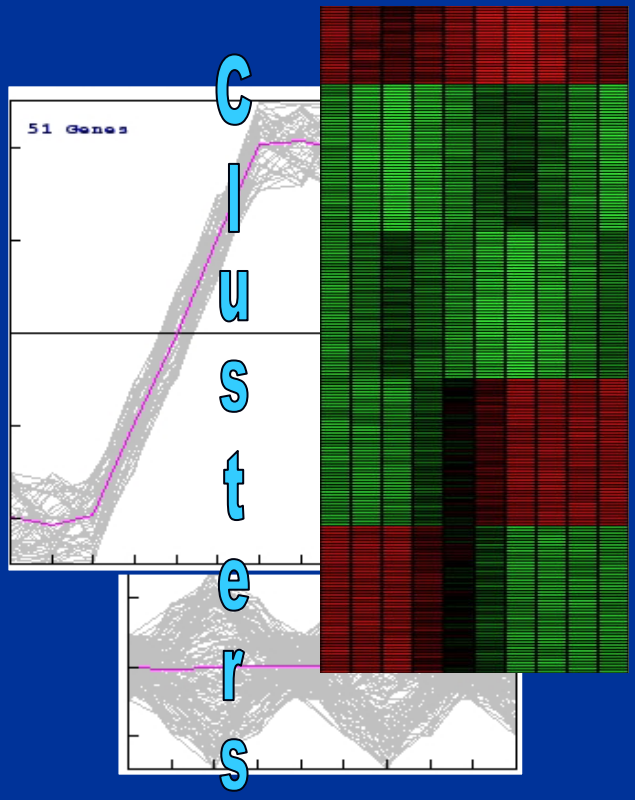
Integrated Data Visualization and Mining Software needed…

… to explore the already known relationships between physiological profiles and infer new interaction models

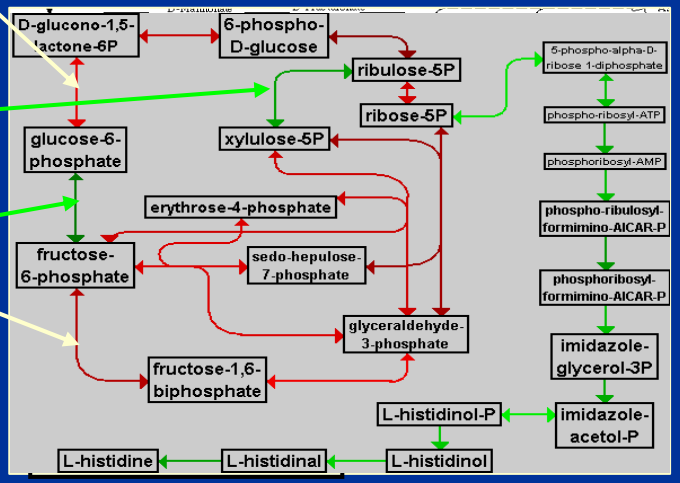
Visualizing the activity of metabolic pathways at the transcriptional level

DNA Microarray

Clustering Algorithms



Pathway Window



Viva la (Interdisciplinary) revolution!

- Holistic analyses of biological systems require a change in the way in which questions are approached in biological sciences
- Collecting, managing and analyzing data from various cellular profiles requires expertise from various fields that transcend traditional discipline boundaries: engineering and computer science, statistics and applied mathematics, and chemistry, physics, and biology
- Bioinformatics and Systems Biology will be the framework for the training of a new generation of researchers in the life sciences who will be able to work, interact and collaborate in a very diverse and highly interdisciplinary environment

Biology will be the foremost science of the 21st Century

“ Doubling in size in 10 years, the biotechnology industry generated 191,000 direct jobs and 535,000 indirect jobs in 2001. Revenues for that year totaled more than $20 billion directly and $28.5 billion indirectly ”

(U.S. Department of Energy)

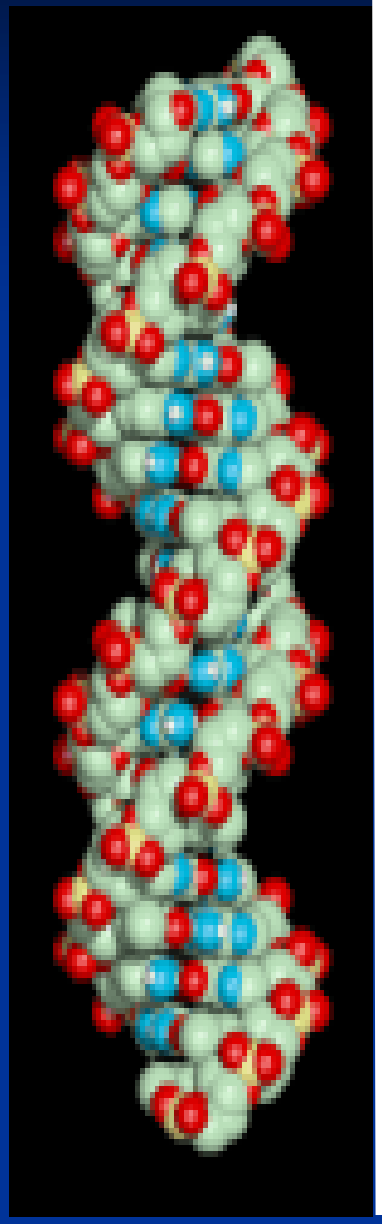
The researchers in systems biology possess for the first time the opportunity to unravel the mechanisms of life

The enormous impacts of these discoveries and the smaller ones along the way in diverse areas, such as:

Metabolic engineering, Strain selection, Drug screening and Development, Bioprocess development, Disease Prognosis and Diagnosis, Gene and Other Medical Therapies

is an obvious motivation for pursuing integrated analyses of cellular systems

2003-2053



2003

2053

Imagine…