CHARITÉ

Systems Biology of Single Cells

H. Holzhütter

hergo@charite.de

http://www.charite.de/sysbio

COMPUTATIONAL SYSTEMS BIOCHEMISTRY

Systems Biology – a New Branch of Life Sciences ?

ecology

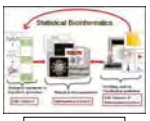
molecular biology

neurobiology

entomology

systems biology ?

plant biology



bioinformatics

physiology

microbiology

Hierarchical Organization of Living Systems

Rudolf Virchow
(1821 – 1920
German pathologist

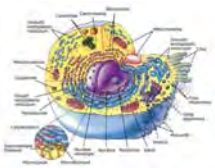
omnis cellulae cellula - every cell originates from another existing cell like it

Theodor Schwann
(1810 – 1882)
German physiologist

all living things are composed of cells, and cells can only arise from pre-existing cells

cell

basic functional unit of all living species



organism

unicellular

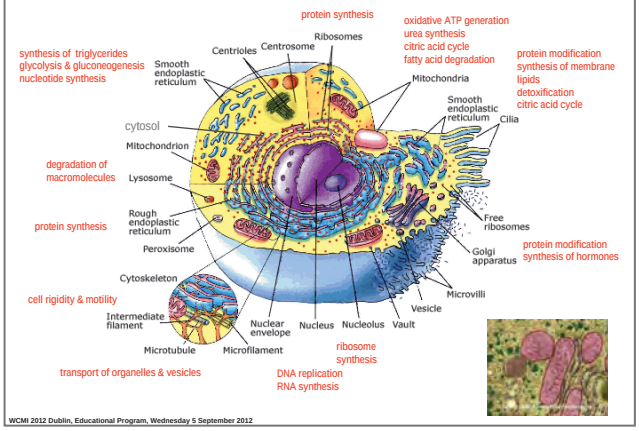
bacteria (e. coli)
(1 cell)

multi-cellular

Caenorhabditis elegans
(959 somatic cells)

Human
(46 – 68 x 10^{12} somatic cells)

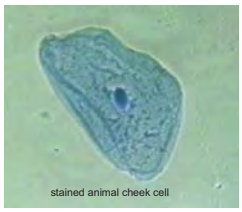
Eukaryotic Cells are Compartmentalized into a Variety of Organelles



Microscopy has Revolutionized Structural Biology

Zacharias Jansen
1580–1638
Dutch spectacle-maker
inventor of first compound microscope

Antoni van Leeuwenhoek
1632 – 1723
Dutch scientist
improved the light microscope
first „microbiologist"

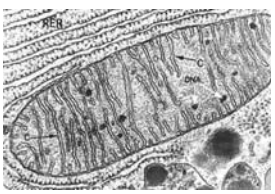


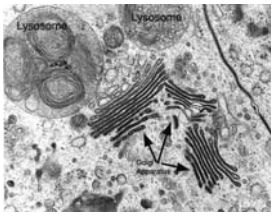
stained animal cheek cell

detailed stages in meiotic cell division

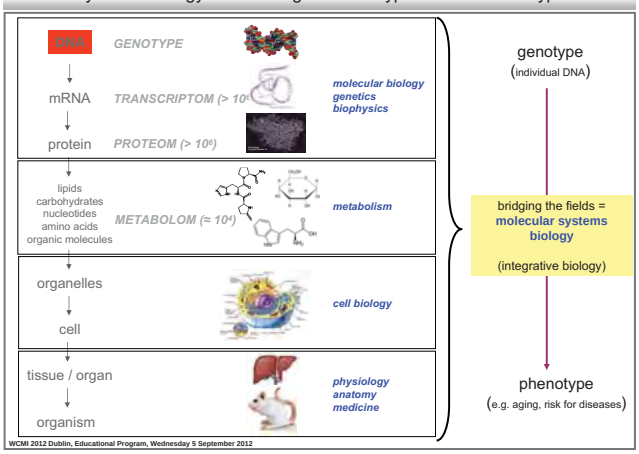
Microscopy has Revolutionized Structural Biology

Ernst Ruska
(1906 – 1988)
German engineer
inventor of the electron microscope

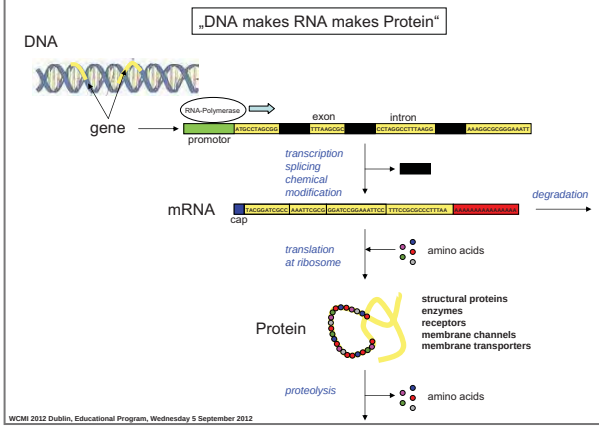




Systems Biology: Translating the Genotype into the Phenotype



Central Dogma of Molecular Biology



Step 1: Determine the Genotype – The Human Genome Project

„the ultimate goal of this initiative is to understand the human genome“

„knowledge of the human genome is as necessary to the continuing progress of medicine and other health sciences as knowledge of human anatomy has been for the present state of medicine

two competing projects („race for the genome')

publicly funded international project comprising the United States, United Kingdom, France, Australia, Japan and a myriad of other spontaneous relationships (F. Colins)

private initiative of Celera Genomics (C. Venter)

draft published in 2000
essentially complete genome published in April 2003

Key findings:

- there are **approximately 23,000 genes** in human beings, the same range as in mice and roundworms
- the human genome has significantly more nearly identical, repeated sections of DNA than other mammalian genomes
- less than 7% of protein families appear to be vertebrate specific

Exploting Genomic Information for the Identification of Genetic Markers

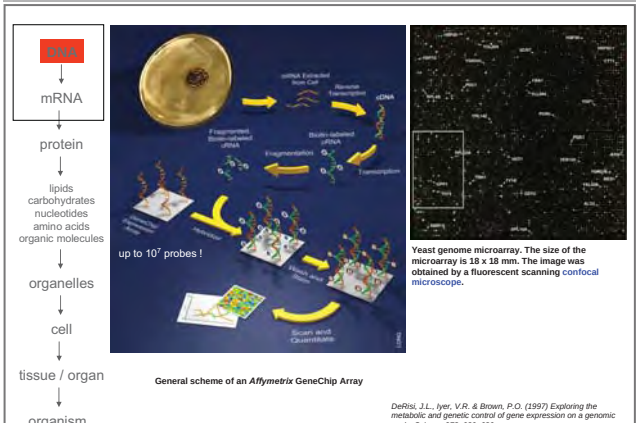
A **genetic marker** is any alteration in your DNA that may indicate an increased risk of developing a specific disease or disorder. Because SNP (Single Nucleotide Polymorphism)s are, by their very definition, variations in DNA, they can be used as flags or markers for nearby DNA that affects your health

Sequence Data are not Sufficient to Understand Diseases

*It is fair to say that **the Human Genome Project has not yet directly affected the health care of most individuals.**"*

*However, when all this sequence data got into computers, it became obvious **that the genetic blueprints by themselves tell us very little about the functional behavior of cells and multicellular organisms**; that is, about what we really want to know about biological systems. In this way, the human genome project, which is perhaps the most spectacular success of molecular biology, also meant that a vast space of future research of a radically different kind became visible.*

Quantification of Gene Activities in Different Tissues

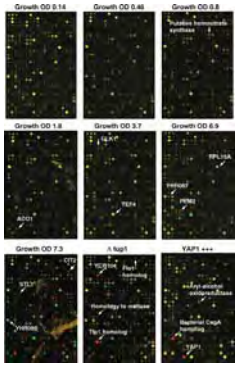


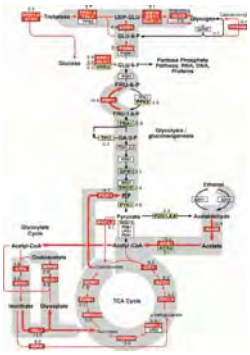
General scheme of an *Affymetrix* GeneChip Array

Yeast genome microarray. The size of the microarray is 18 x 18 mm. The image was obtained by a fluorescent scanning confocal microscope.

DeRisi, J.L., Iyer, V.R. & Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

Temporal Gene Expression of Metabolic Enzymes in Yeast



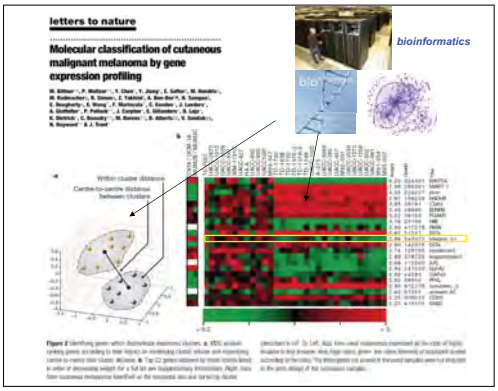


DeRisi, J.L., Iyer, V.R. & Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

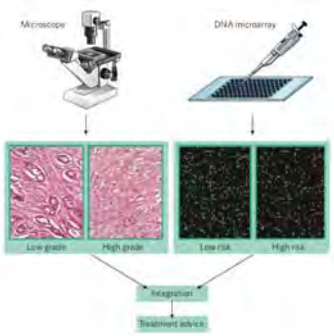
Molecular Classification of Cells Based on mRNA Profiles

DNA
↓
mRNA
↓
protein
↓
lipids
carbohydrates
nucleotides
amino acids
organic molecules
↓
organelles
↓
cell
↓
tissue / organ
↓
organism

does the mRNA profile allows predictions of the phenotype ?



Gene Expression Patterns as Diagnostic Tools

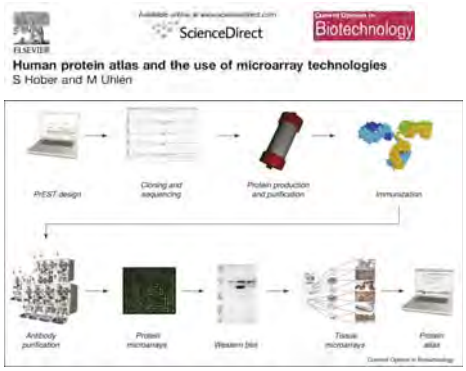


Conventional diagnostic tests rely heavily on **morphological criteria** (that is, properties of cells that can be observed microscopically in tissue sections from a tumour) to judge the aggressiveness of cancer

Multigene-expression tests (also known as *in vitro* diagnostic multivariate index assays, IVDMIAs) have been shown to be powerful tools for predicting disease outcome.

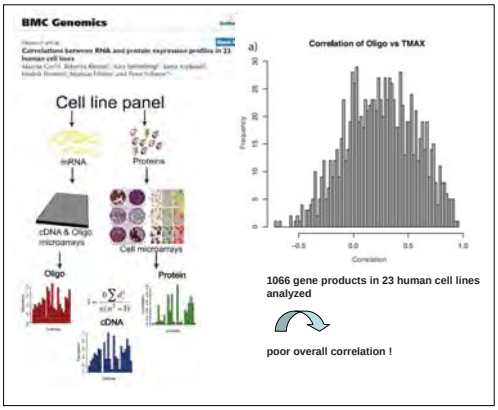
Quantification of Protein Levels

DNA
↓
mRNA
↓
protein
↓
lipids
carbohydrates
nucleotides
amino acids
organic molecules
↓
organelles
↓
cell
↓
tissue / organ
↓
organism



The mRNA Level is a Poor Predictor of the Protein Level

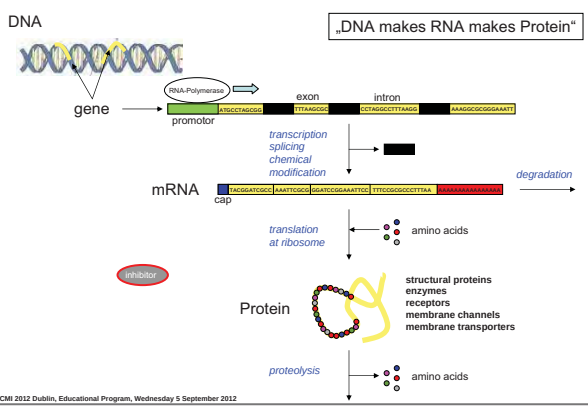
DNA
↓
mRNA
↓
protein
↓
lipids
carbohydrates
nucleotides
amino acids
organic molecules
↓
organelles
↓
cell
↓
tissue / organ
↓
organism



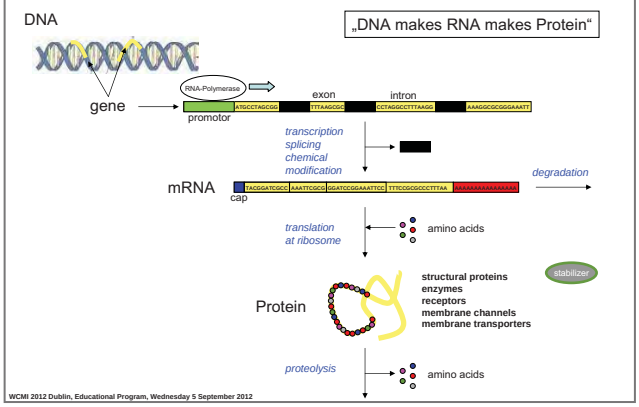
1066 gene products in 23 human cell lines analyzed

poor overall correlation !

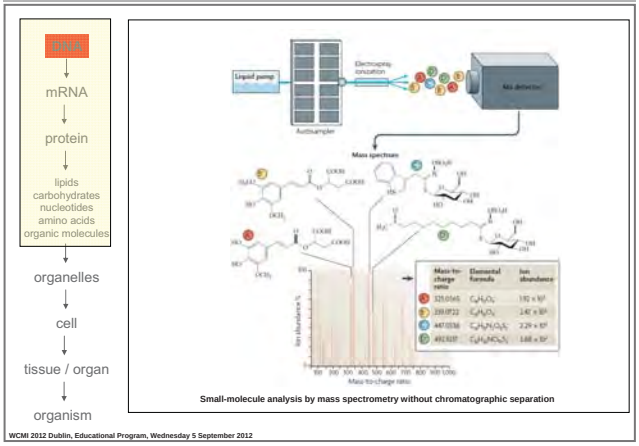
The Importance of Post-Transcriptional Control of Gene Expression



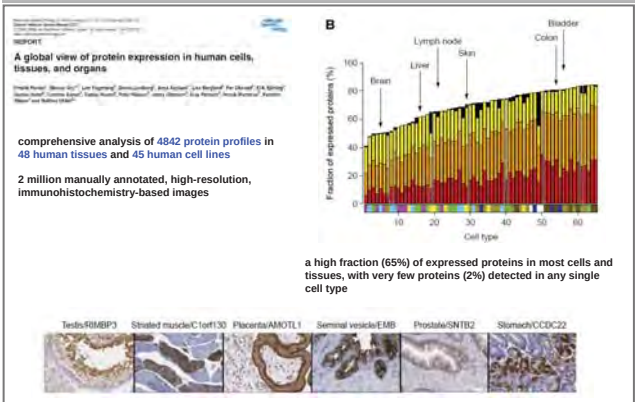
The Importance of Post-Transcriptional Control of Gene Expression



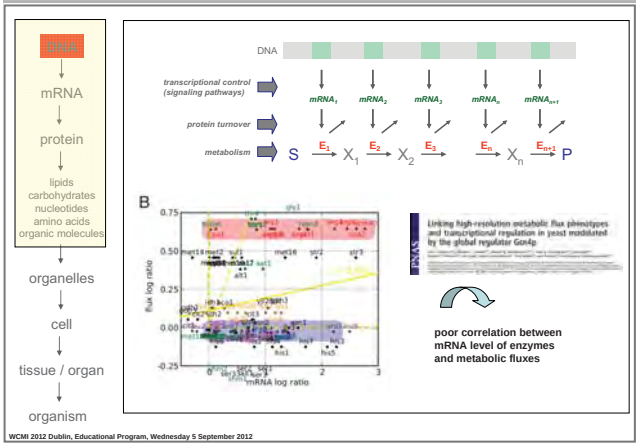
Quantification of Metabolites and Metabolic Fluxes



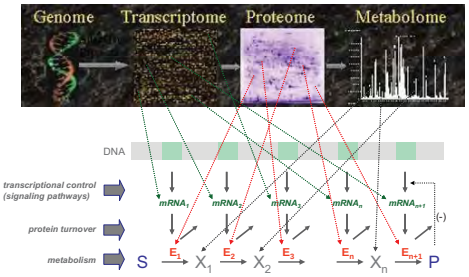
Striking Similarities Among Proteomes of Different Celly Types



Poor Correlation between TranscriptionLevels of Enzymes and Flux Rates



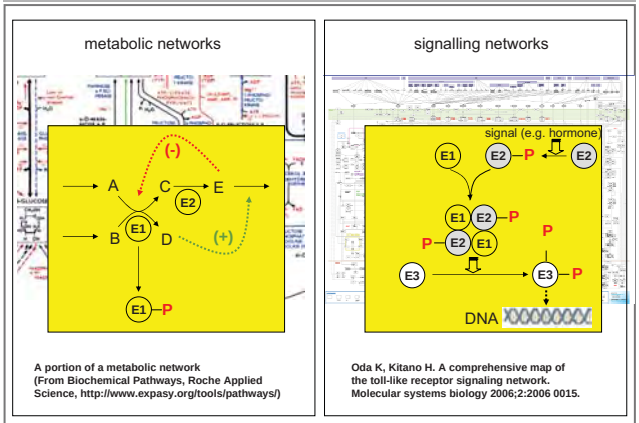
The Limitation of High-Throughput “Top-Down” Approaches



High-throughput methods (microarrays, mass spectrometry…) in combination with high-resolution imaging techniques (confocal microscopy) enable large-scale molecular snapshots of the cell – but the dynamic linkage between the snapshots remains unsettled

we need to know the molecular interactions (processes) governing the observed snaphots

Interacting Molecules Form Dynamic Networks



Identification and Quantification of Molecular Interactions

example: protein – RNA - binding

mRNA

stabilizer

formal notation of process

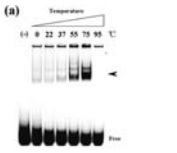
$$mRNA + S \rightleftharpoons mRNA{\cdot}S$$

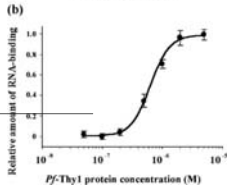
rate law (binding relation)

$$[mRNA \bullet S] = \frac{[mRNA_{total}]}{1+\left(\frac{S_{0.5}}{[S]}\right)^{n}}$$

determination of parameters from experimental binding curve

gel-shift-assay

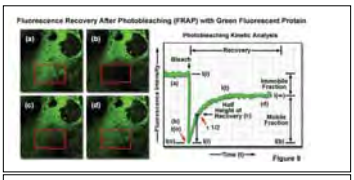


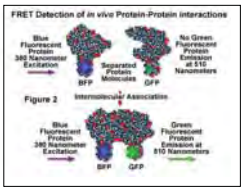


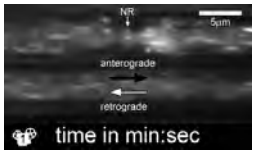
Kanai A, Sato A, Imoto J, Tomita M. Archaeal Pyrococcus furiosus thymidylate synthase 1 is an RNA-binding protein. The Biochemical journal 2006;393(Pt 1):373-9.

The Eminent Role of Confocal Fluorescence Microscopy in Systems Biology

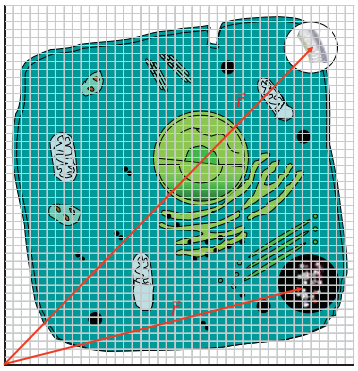
A high-speed multispectral confocal microscope can simultaneously acquire two-color images at speeds up to 30 frames per second.







Computational Systems Biology: The Ultimate 'In Silico' Cell



ultimate **goal** of *in silico* cell biology: calculate (= predict)

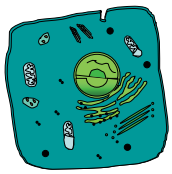
$$N_i(\vec{r},t)$$

= the number of molecules of sort (i) at time t in any infinitesimally small region at position $\vec{r}$

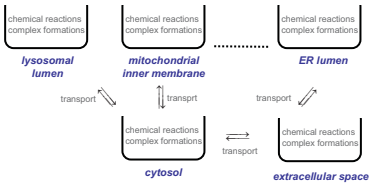
main issue:

molecular mechanisms of structure formation (e.g. genesis of membranes, self-assembly of multi-component macromolecular complexes) currently not amenable to mathematical modeling

Neglecting Spatial Structures: The Simplified 'In Silico' Cell



representation of the cell by various interconnected reaction compartments



simplifications:

- formation and dynamics of spatial structures (e.g. genesis and repair of organelles, mitotic cell division, movement of vesicles) not considered
- compartments treated as well-mixed reaction chambers
- copy numbers of molecules approximated by concentrations

goal:

calculate (= predict) $x_i(C,t)$

= concentration of molecules of sort (i) at time t in compartment (C)

'In Silico' Cell Models – Basic Ingredients

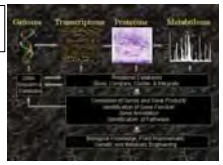
variables

data input thanks to high-troughput technologies

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

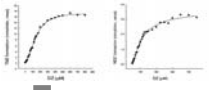
concentrations of cellular molecules

glucose, Na^+, VLDL, phosphatidylcholin, cystein, $NADH_2$, cAMP, hexokinase, $mRNA_{IL-1}$,…



enzyme (protein) preparation

in vitro kinetics



processes

small and further decreasing input

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \end{bmatrix}$$

rate equations for reactions and complex formations

Fru-1,6-P ↔ GAP + DHAP

free aa + $mRNA_{ApoB-100}$ → Apo_{B-100}

Apo_{B-100} + TG + Chol + PL ↔ VLDL…

$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_t \end{bmatrix}$$

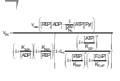
rate equations for trans-membrane transports

ATP_{mito} + ADP_{cyto} ↔ ATP_{cyto} + ADP_{mito},

Glu_{ext} ↔ Glu_{cyt}, …

mechanism

rate law



'In Silico' Cell Models – Basic Kinetic Equations

variables

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

concentrations of cellular molecules

glucose, Na^+, VLDL, phosphatidylcholin, cystein, $NADH_2$, cAMP, hexokinase, $mRNA_{IL-1}$,…

$$\frac{d}{dt}\vec{X} = \boldsymbol{N^R}\,\vec{v} + \boldsymbol{N^T}\,\vec{w}$$

stoichiometric matrix for reactions; stoichiometric matrix for transports

time-dependent concentration changes; generation and utilization by chemical processes; uptake and release by transport processes

processes

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \end{bmatrix}$$

rates (=fluxes) of reactions and complex formations within compartments

Fru-1,6-P ↔ GAP + DHAP

free aa + $mRNA_{ApoB-100}$ → Apo_{B-100}

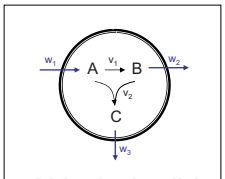
Apo_{B-100} + TG + Chol + PL ↔ VLDL…

$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_t \end{bmatrix}$$

rates (=fluxes) of trans-membrane transports

ATP_{mito} + ADP_{cyto} ↔ ATP_{cyto} + ADP_{mito},

Glu_{ext} ↔ Glu_{cyt}, …

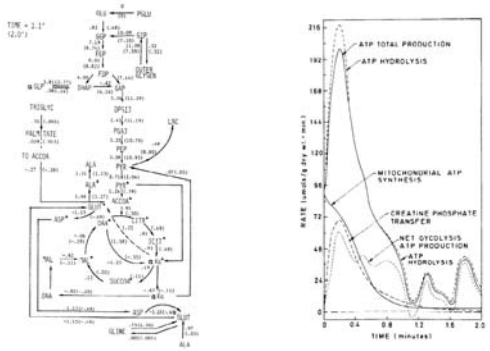


$$\frac{d}{dt}\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

$\boldsymbol{N^R}$ $\boldsymbol{N^T}$

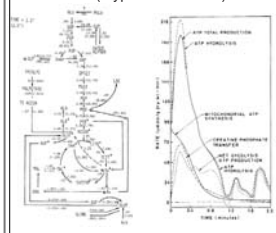
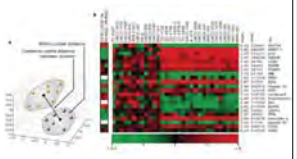
Mathematical Modelling of Cellular (Metabolic) Processes Has a Long History

energy metabolism of the heart during ischemia



Garfinkel D, Achs MJ (1979) Metabolism of the acutely ischemic dog heart. II. Interpretation of a model. *The American journal of physiology* **236:** R31-39.

„Reductionistic“ versus „High-Throughput“ Systems Biology

Reductionistic Approach (“hypothesis-driven”)	High-Throughput Approach (“data-driven”)
	
• **predictive**: provides a mechanistic explanation of systems behaviour • experimentally and computationally challenging • restricted to small cellular sub-systems	• **not predictive**: yields (almost complete)molecular photographs of the cell, only emenable to statisical analysis but not to causal modelling

?

„Reductionistic“ versus „High-Throughput“ Systems Biology

Point: Hypotheses first

There is little to show for all the time and money invested in genomic studies of cancer, says **Robert Weinberg** — and the approach is undermining tried-and-tested ways of doing, and of building, science.

During the twentieth century, biology — traditionally a descriptive science — became one of hypothesis-driven experimentation. Tightly coupled with this was the blossoming dominance of reductionism, the idea that complex biological systems can be understood by dismantling them into their constituent pieces and studying each in isolation. Implicit here was the tension that obsessive-compulsive disorder can be made to support or attack. Hypothesized mechanisms of action, and that single observation — phenomenology for its own sake — is of relatively little use.

These approaches served us well over the past half century: witness the revolutions in molecular and cellular biology, immunology, neurobiology and genetics. Our insights into pathogenesis and medicine exceed the wildest speculations of 50 years ago.

Now the dominant position of hypothesis-driven research is under threat. Many feel that traditional conceptual tools cannot map the enormous complexity that allows single cells and complex organisms to thrive, and that recent technological innovations have created a viable alternative. My students can gather certain types of information data 1,000 and even 10,000 times faster than I could 40 years ago.

biology that has since emerged. Alternating proportions of national research budgets are being diverted to them. But is it worth anticipating 20 or 30 small-scale, hypothesis-driven projects to make room for an attack at the systems-wide level?

From a cancer researcher's perspective, the successes of hypothesis-driven science are clear and undeniable. They stretch back over half a century and continue week after week, month after month, to yield new conceptual insights. By contrast, the new ways of doing biology are so untested that their long-term benefits are still hard to project. Nonetheless, it's useful to make comparisons, if only because economic success tries them to be made.

Analyses of expression arrays, which show which genes are active in a tumour sample, have shown that cancers previously viewed as a single entity have different pathogenetic mechanisms and respond differently to therapy. The use of genome-wide libraries of small interfering RNAs to inhibit large cohorts of genes, and so identify those behind cancer, is ahead of the old and the new. It lacks a clear preconception of what will eventually be found, but contains clear hypotheses about the biological phenotypes that will result. This approach, still in

— although this is likely to change as technology costs tumble. Meanwhile, countless studies of experimental research programmes — private sources, true-false fear of conceptual immaturity — have struggled to survive.

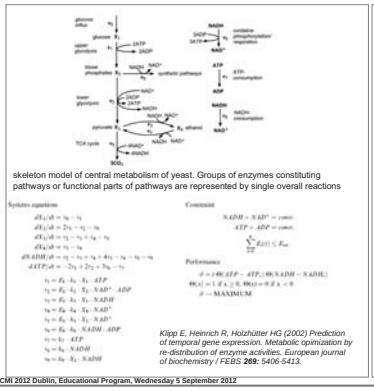
High stakes

A quality the most ambitious large-scale projects — those involving assembling the many interacting, signalling components within individual cells into wiring diagrams. These elaborate maps, sometimes termed 'manifolds', and the computer algorithms that model signal processing, could shed light on why and how individual cells respond to external signals and produce their various behaviour. Although aesthetically pleasing, they have yielded few conceptual insights into how and why cells and tissues behave the way they do. Some feel that a thorough understanding of individual signal-processing components is an essential prerequisite to predicting the behaviour of entire signalling circuits — a notion often dismissed as old-style

"Is it worth attempting 20 small-scale projects for an attack at the systems-wide level?"

Hypothesis-Driven Prediction of Temporal Gene Expression in Yeast

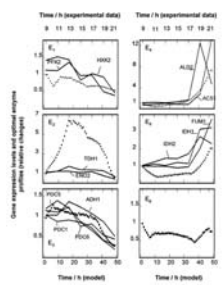
Hypothesis:
Gene regulation aims at economizing protein usage while stabolizing essential metabolites (ATP, NADH)



skeleton model of central metabolism of yeast. Groups of enzymes constituting pathways or functional parts of pathways are represented by single overall reactions

Klipp E, Heinrich R, Holzhütter HG (2002) Prediction of temporal gene expression. Metabolic optimization by re-distribution of enzyme activities. European journal of biochemistry / FEBS **269:** *5406-5413.*

dioxic shift (glucose depletion)

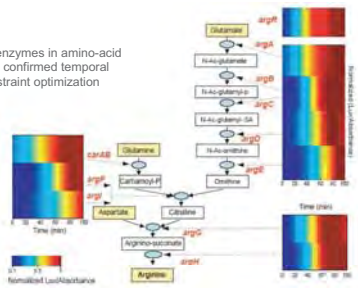


dotted curves: optimal model solution
solid lines: expression profiles taken from the Stanford Microarray database
http://cmgm.stanford.edu/pbrown/explore/array.txt

Experimental Verification of Predicted ‚Just-in-Time‘ Gene Expression

measured promotor strengths of enzymes in amino-acid biosynthesis pathways of *E. coli* [*)] confirmed temporal enzyme profiles predicted by constraint optimization

economization of enzyme usage is an important goal of natural evolution and can be achieved by optimal temporal gene expression !



Temporal promoter activity profiles of the nine arginine biosynthesis operons. *Lux reporter strains were grown in defined medium containing all amino acids and diluted into the same medium lacking arginine. Shown is the ratio of Lux activity to absorbance at each time point divided by the highest ratio for each strain.* ***Blue*** *and* ***red*** *indicate* ***low*** *and* ***high*** *expression, respectively. The pathway diagram illustrates the metabolic steps of the arginine biosynthesis pathway and the dashedarrows indicate the stages at which each gene product participates. Similar results were obtained using GFP reporter strains.*

[*)] Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G. & Alon,U. (2004) *Just-in-time transcription program in metabolic pathways. Nat Genet.36, 486-491*

Cellular Systems Biology in a Nutshell

Systems Biology an **interdisciplinary** research field which

aims at understanding how molecular interactions between DNA, mRNA and proteins result in phenotypic properties like growth, aging and malfunction

DNA → RNA → protein
↓
lipids
carbohydrates
nucleotides
amino acids
organic molecules
↓
organelles
↓
cellular function

applies (automated) high-throughput technologies (including imaging techniques) to quantify cell-wide dynamic changes in virtually all RNAs, proteins and small organic molecules

needs bioinformatics to retrieve relevant information from huge data sets

applies rigorously mathematical models to be quantitative and predictive

can only be succesfull if it integrates discovery-driven and hypothesis-driven (global or focused) approaches