Modelling biological systems: a computational challenge

Parma, 8-13 September, 2008

G.C. Rossi

University of Rome Tor Vergata

INFN - Sezione di Tor Vergata

Acknowledgments --- Apologies

- I wish to thank Enrico for the opportunity he gave me to present this material
- and all the people of the Biophysics group of ToV (especially Silvia) for ∞-ly many discussions which are at the origin of these lectures

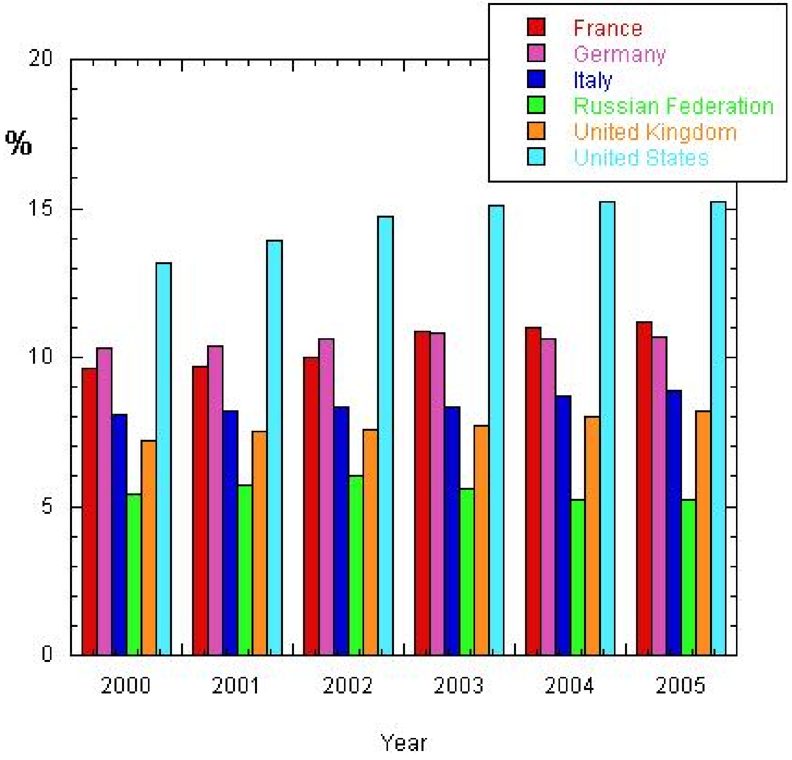
----

- Choice of arguments was made on the basis on my tastes, preferences and incompetence
- The amount of underlying biological knowledge behind most of the arguments I will touch is essentially unlimited and well beyond my competence
- Thus, I will try to convey you rather than a fully detailed biological information, some general description of certain broad classes of systems and problems on which one can probably say something interesting and useful
- I hope you'll find some of these problems intellectually appealing and exciting, not less than High Energy Physics (**HEP**) or Astrophysics, if not for their dramatic impact on our everyday life

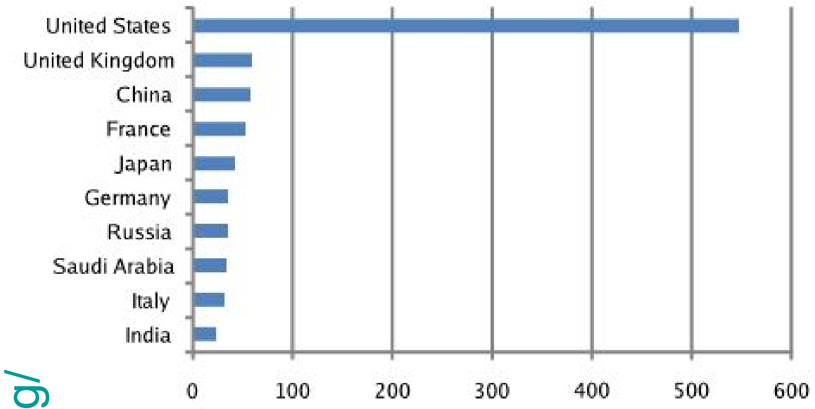
Outline?

The field of health care and biomedical sciences is where the **action** is (in developed countries)

Total expenditure on health as percentage of gross domestic product



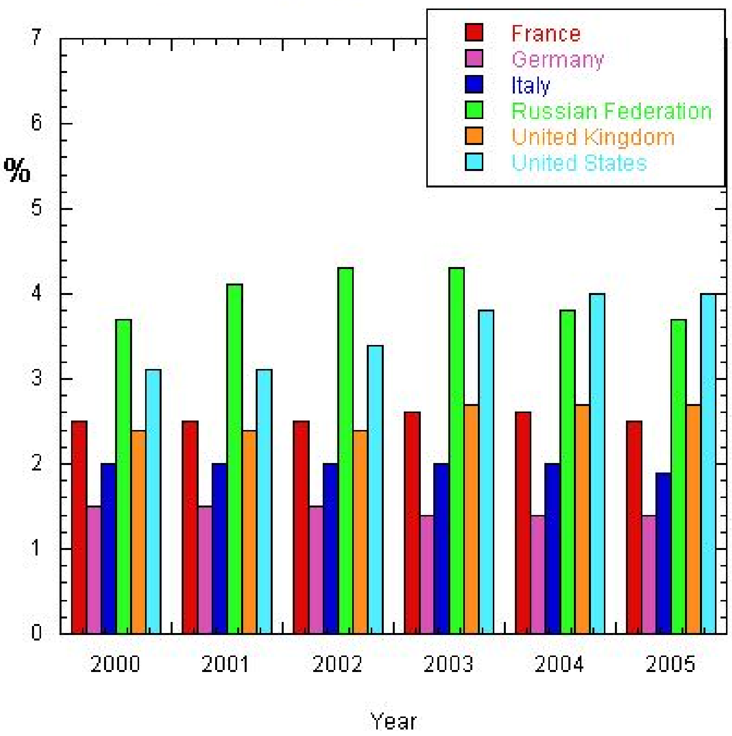
http://www.who.int/en/



Billions US $ in 2007

http://milexdata.sipri.org/

Military expenses as percentage of gross domestic product



Outline

I. Reductionism *vs* complexity

II. Data, (physical) models and (mathematical) tools

III. What we would like to know and/or to do

IV. What we can actually do and/or are really doing

V. Conclusions and outlook

From concepts to action

I. Reductionism *vs* complexity

- A bit of philosophy
- A bit of phenomenology

A bit of “philosophy”

Biology *vs* Physics

(the viewpoint of a theoretical physicist)

- Compare and contrast the situation in the study of
  Biological systems
  - “Complex” structures governed by (as yet) unknown macro-laws
  - Powerful and cheap experimental techniques
  - Huge amount of data
  - Inadequate models: poor understanding of “micro” to “macro” transition
- and, at the other extreme, of
  Elementary Particle Physics
  - Supposedly “simple” systems governed by “elegant” known micro-laws
  - Very complicated and expensive experiments
  - Very few new experimental data (LHC is coming!)
  - Rather good models (almost “theories”)

Physics (until very recently) has always found its way by progressively moving towards more and more elementary structures

matter → atoms → nucleons → quarks → ???

guided by the “radical reductionism” paradigm according to which

FUNDAMENTAL LAWS GOVERN ELEMENTARY OBJECTS

This attitude has been very fruitful in the “paradigmatic” case of **HEP**, but it is not obviously being employed in other emerging fields of investigation

- **Dynamical** systems

  Weather forecasting
  Catalytic reactions
  Fluidodynamics (turbulence)

  key words: non-linearity, chaos

- **Disordered** systems   Glasses, Spin glasses

  key-words: frustration, disorder

- **Biological** systems

  key-words: complexity, and perhaps all of the above

1 - There are implications for the notion of modelling and the nature of physical laws

● Even in **Fundamental Physics** what we usually call

$$\left.\begin{array}{l}\text{Relativity}\\ \text{Field}\\ \text{String}\end{array}\right\}\ \text{Theories}$$

are actually Models, formulated in the language of Mathematics, from which they borrow the necessary internal logical consistency

● Complications of everyday life (like friction in Mechanics) are considered (conceptually) irrelevant (up to a certain point - airplanes, cars,...!)

● Theories become progressively simpler in the process of understanding

●● For **Biosystems**, Models (nobody would call them theories) tend to become more and more complicated, as they develop (not simpler!), with a limit: the model shouldn't become as complicated as the system itself!

●● The key questions about modelling in **Biology** are then

⇒ When do we decide that we have "understood"?

protein folding

functional behaviour of the cell

⇒ What kind of knowledge/predictions will we be happy with?

2 - There are implications for the notions of experiment and reproducibility

- The Central Dogma of Physics

  Theories (models) are validated through reproducible experiments

- In many biological instances the situation is somewhat more complicated. For instance, to put it in a provocative way

"The experiment of testing ***in vivo*** the effectiveness of a drug (working ***in vitro***), would certainly not be considered a failure if, say, only **30%** of ill people recover"

●● Can we somehow understand this situation?

1. Biological experiments may not give reproducible results because not all the relevant dof's are/can be kept under control ⇒ # dof's >> 1
2. On the other hand, in most cases (but, see later) it is not of any interest to be able to predict the properties of the final state of a biological system, or process, in its finest details ⇒ disorder & redundancy
3. Models are very crude (when they exist at all) and most often overwhelming complicated ⇒ need for some intrinsically new concept?

The systems of interest

Reductionism

- Elementary is an object characterized by a small # of properties
- All elementary objects of a given kind are alike (electrons)
- Simple physical laws (theories) apply to elementary objects
- Strict determinism and experimental reproducibility follow

Complexity

- Complex systems have many dof's and many functionally relevant components
- One should talk of classes of systems, e.g.
  - the class of nervous cells, the class of liver cells
  - or, more generally, the class of nucleated cells

  Classes are defined by identifying the common properties of the constituent systems
- Models yield a mathematical description of common features of systems belonging to a given class in terms of probability distribution functions (PDF)
- Class averages are computed and compared to results coming from averages over sets of experiments

3 - There are implications for the amount and the nature of the possible information output

Key point

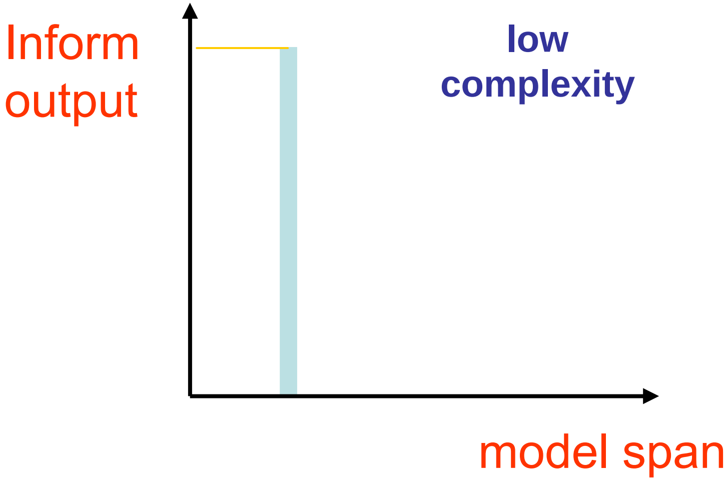
is the accuracy by which a class of homogeneous objects can be defined

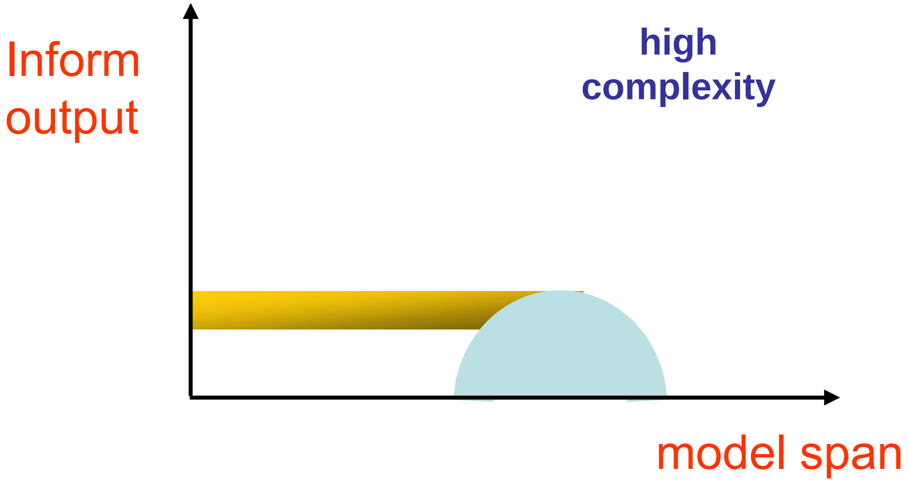
The more accurate (looser) the definition of the objects belonging to a certain class

the simpler (more complicated) the model

the sharper (more involved) its mathematical description

the more precise (fuzzier) the information output





Key questions at this point are

Q1: What is complexity?
A1: Its meaning is context dependent

Q2: Are biosystems complex objects?
A2: Looks like they are

1. Algorithmic Complexity of Kolmogorov and Chaitin

- Definition:

Given a string *S* of **N** symbols → **AC** = # of bits of a T.M. code that can produce *S* as an output

- Such a definition does not look interesting for us

$$\text{AC (random string)} >> \text{AC } (\pi)$$

$$\begin{cases} \text{AC (random string)} \sim \mathbf{N} \\ \text{AC } (\pi) \sim \log \mathbf{N} \end{cases}$$

[actually the digits of π are totally random]

2. Logical depth of Bennett

- Definition:

Given a string *S* of **N** symbols → **LD** = time (# of operation) for a T.M. to run the shortest code that can produce *S* as an output

- A somewhat more interesting definition

$$\begin{cases} \text{LD (random string)} \propto \text{time to read } S \sim \mathbf{N} \\ \text{LD } (\pi) \propto \text{time to generate } \pi \sim \mathbf{N} \end{cases}$$

Biological Complexity

- is not **randomness**

Box of molecules
with random velocities

S=large

- is not **entropy**

Box of molecules
with all parallel velocities

S=0

- is not **logical depth**

Life emerged from a very short (random) program, but it took 10^9 y to run the code: very high LD!
What about running the code today?

Then what is it?

Necessary conditions

- many variables
- many relevant dof's

Here a bit of “phenomenology” starts

Numerosity

	# of elementary constituents (atoms)
ATOM	1
AMINO ACID	10
PROTEIN	10^3-10^5
CELL	10^{10}
.	
.	
.	
HUMAN BODY	$5x10^{28}$ (nucleons)

Numerosity and Heterogeneity

- Proteins

  10^2 - 10^3 amino acids
  10^3 - 10^5 atoms → (only ~10^7 expressed)
  20^{300} different possible sequences!

- Immune system

  10^6 actual repertoire of Ab's
  10^7 available repertoire
  10^8 lymphocytes

- Brain

  10^{10} neurons
  x 10^3~10^4
  10^{13}-10^{14} synapses

- Genoma

  3x10^9 bases (human DNA)
  4^n with n = 3x10^9 possible genomes
  (only 10^{60} expressed @ 1 mut/sec) Eigen

2-3 nm helix x 2 m long
2x23 chromosomes V~(1.5 μm)^3

It is not so much the number of "elementary" objects that is important (gas), but rather the existence of a large number of "functionally" relevant distinct components

- There is a lot of disorder in Biosystems

They have (~ ∞-ly) many randomly distributed microscopic variables and few (still very many!) mesoscopic variables controlling the system

Not every detail can be encoded in DNA, nor every Genoma has been tried

No optimal evolution

- There is a lot of redundancy in Biosystems

They can exist in very many “equilibrium/metastable” states

Individuals
Organs
Immune system states
Proteins

Microscopically different organs (harts, brains,…) equally well accomplish their task

High degeneracy

Complexity:

here is a sort of “phenomenological” definition

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

Complexity is complexity of classification

1. Sequences of random numbers

Not much can be said

all instances belong to the same class

⇒ It is a very simple class of systems

2. Equilibrium states of a system of spins at $H = 0$, $T \sim 0$

Only two states: spin up, spin down

⇒ It is a simple system

3. Class of sequences of symbols giving rise to “books”

Many things can be said

Language	⇒	English, Italian, German, ...
Style	⇒	Poem, Tragedy, ...
Plot	⇒	Love story, Detective story, ...
...	⇒	...

Many “description levels” or tasks ⇒ Various possible “types of classification”

⇒ It is a complex class of systems

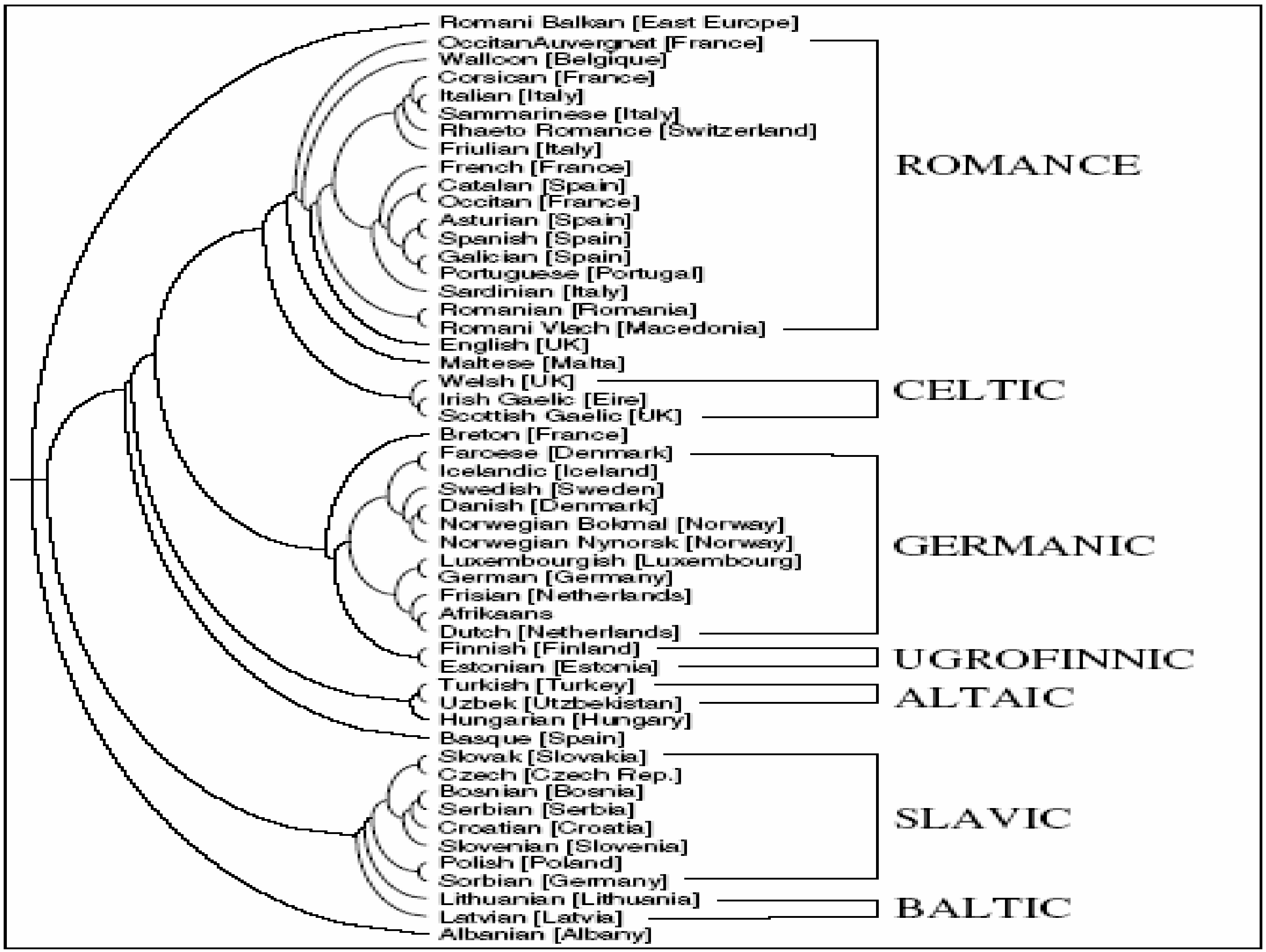
4. Set of painters

We could learn a lot, if we could establish

When they were active	⇒	Date of birth
Where they were active	⇒	Place of birth
Their style	⇒	Relative influence
...	⇒	...
Many “description levels” or tasks	⇒	Various possible “types of classification”

⇒ It is a complex class of systems

5. The class of human languages is a complex system



Evolutive tree

Correlating Genetic Tree and Linguistic Phyla

Cavalli Sforza
Piazza Menozzi
Mountain

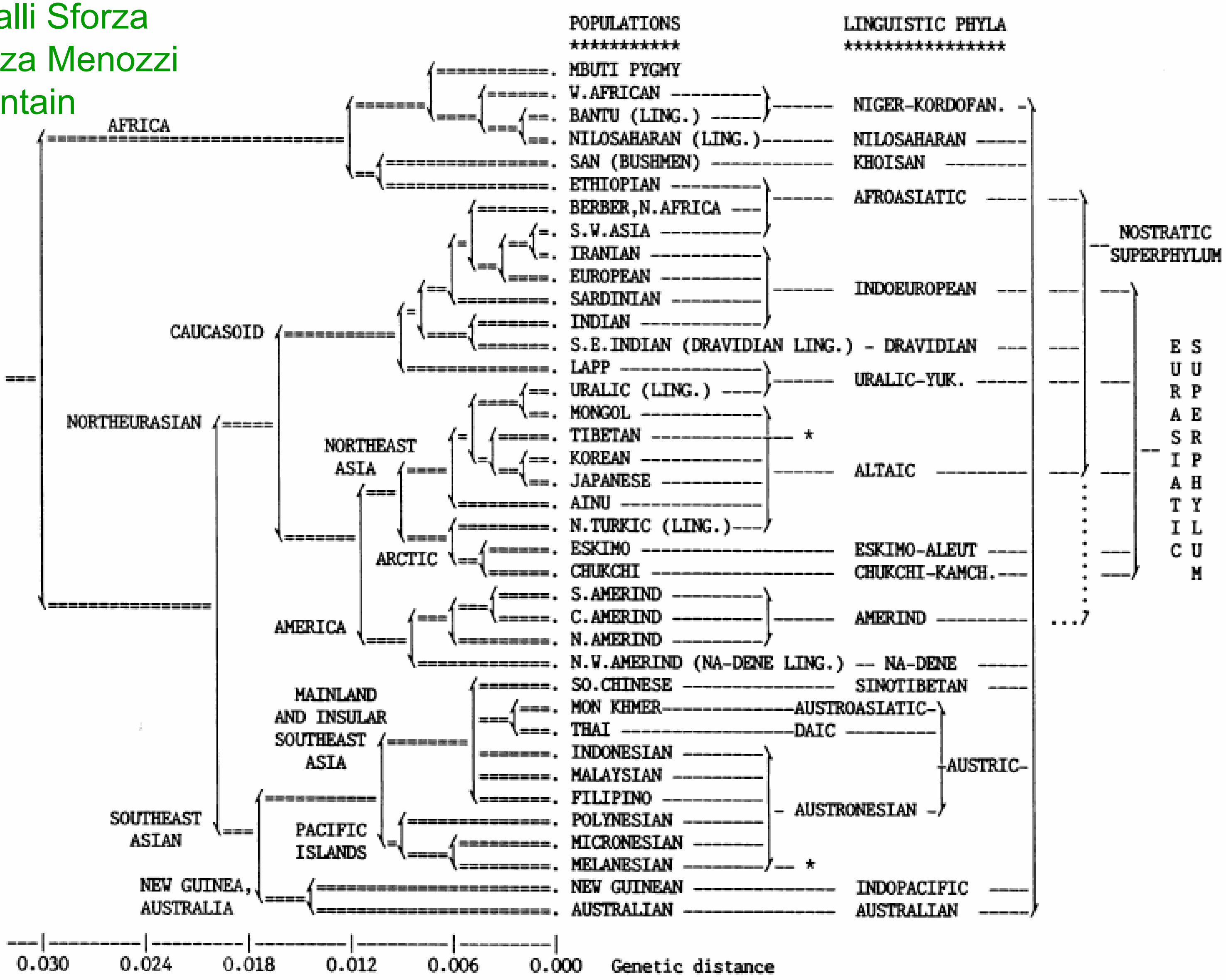
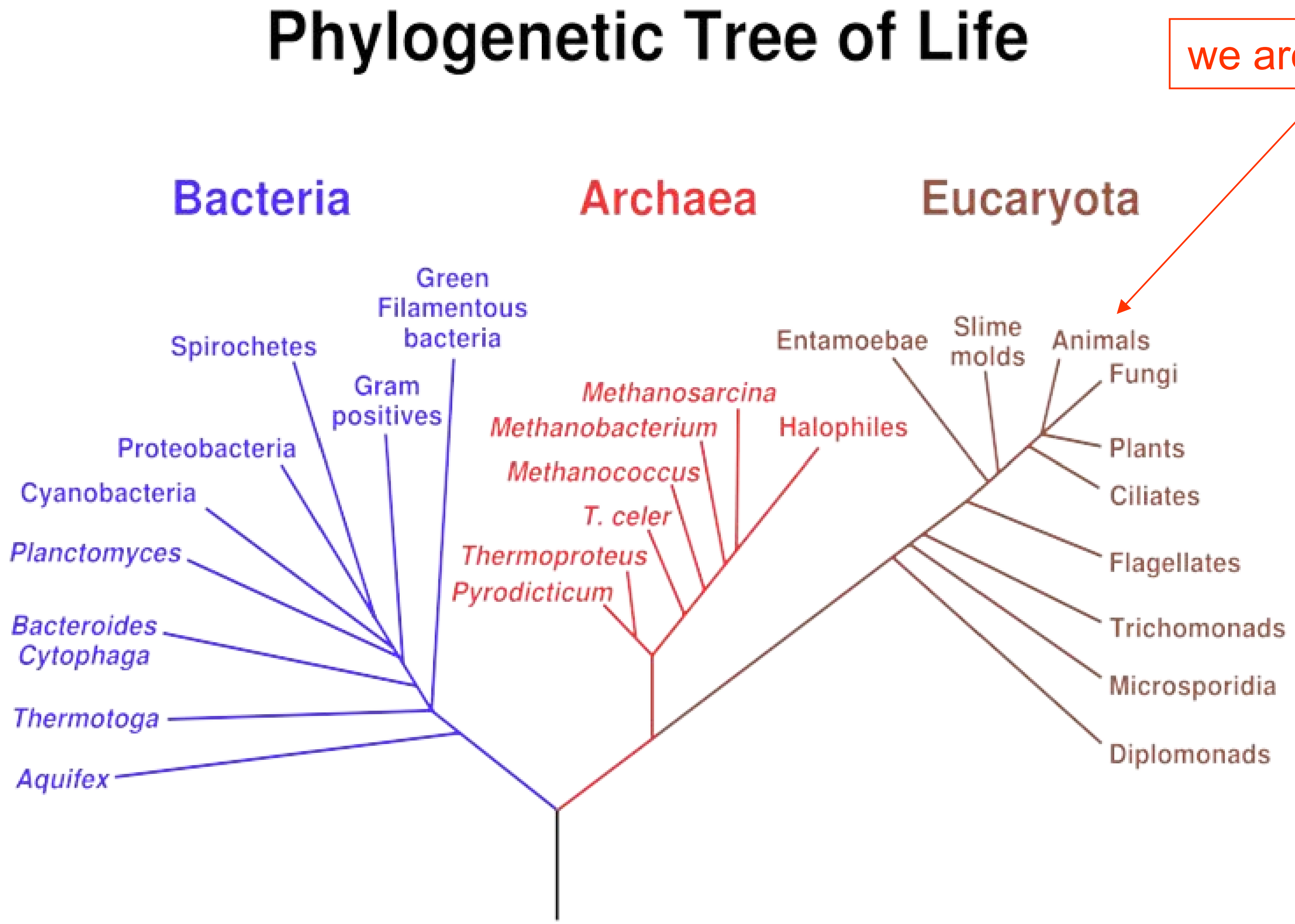


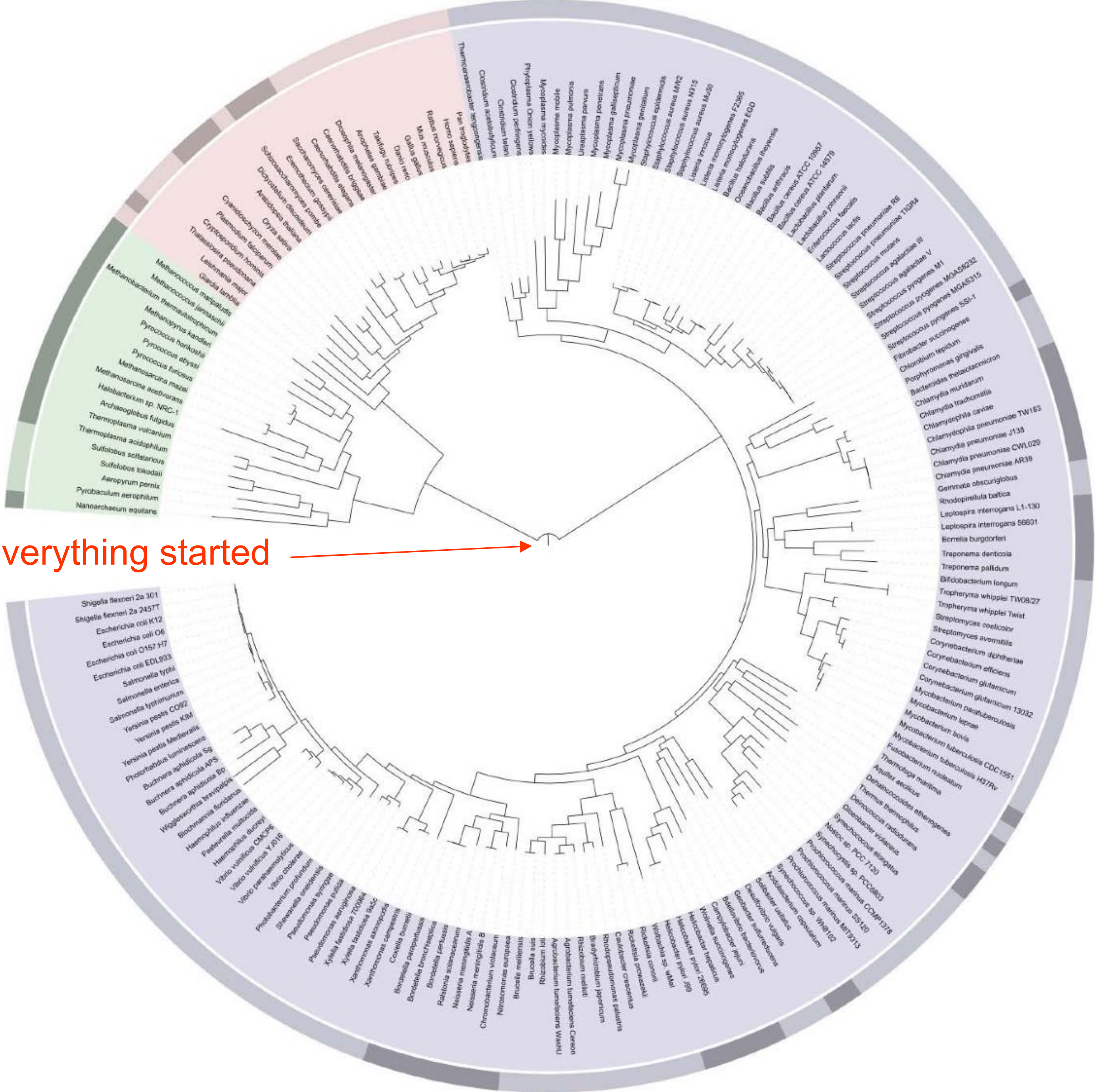
FIG. 1. Comparison of genetic tree and linguistic phyla. See text for details. (Ling.) indicates populations pooled on the basis of linguistic classification. The tree was constructed by average linkage analysis of Nei's genetic distances. Distances were calculated based on 120 allele frequencies from the following systems: *A1A2BO*, *MNS*, *RH*, *P*, *LU*, *K*, *FY*, *JK*, *DI*, *HP*, *TF*, *GC*, *LE*, *LP*, *PEPA*, *PEPB*, *PEPC*, *AG*, *HLAA* (12 alleles), *HLAB* (17 alleles), *PI*, *CP*, *ACP*, *PGD*, *PGM1*, *MDH*, *ADA*, *PTC*, *E1*, *SODA*, *GPT*, *PGK*, *C3*, *SE*, *ESD*, *GLO*, *KM*, *BF*, *LAD*, *E2*, *GM*, and *PG*.

6. The set of living organisms on the Earth is a complex system

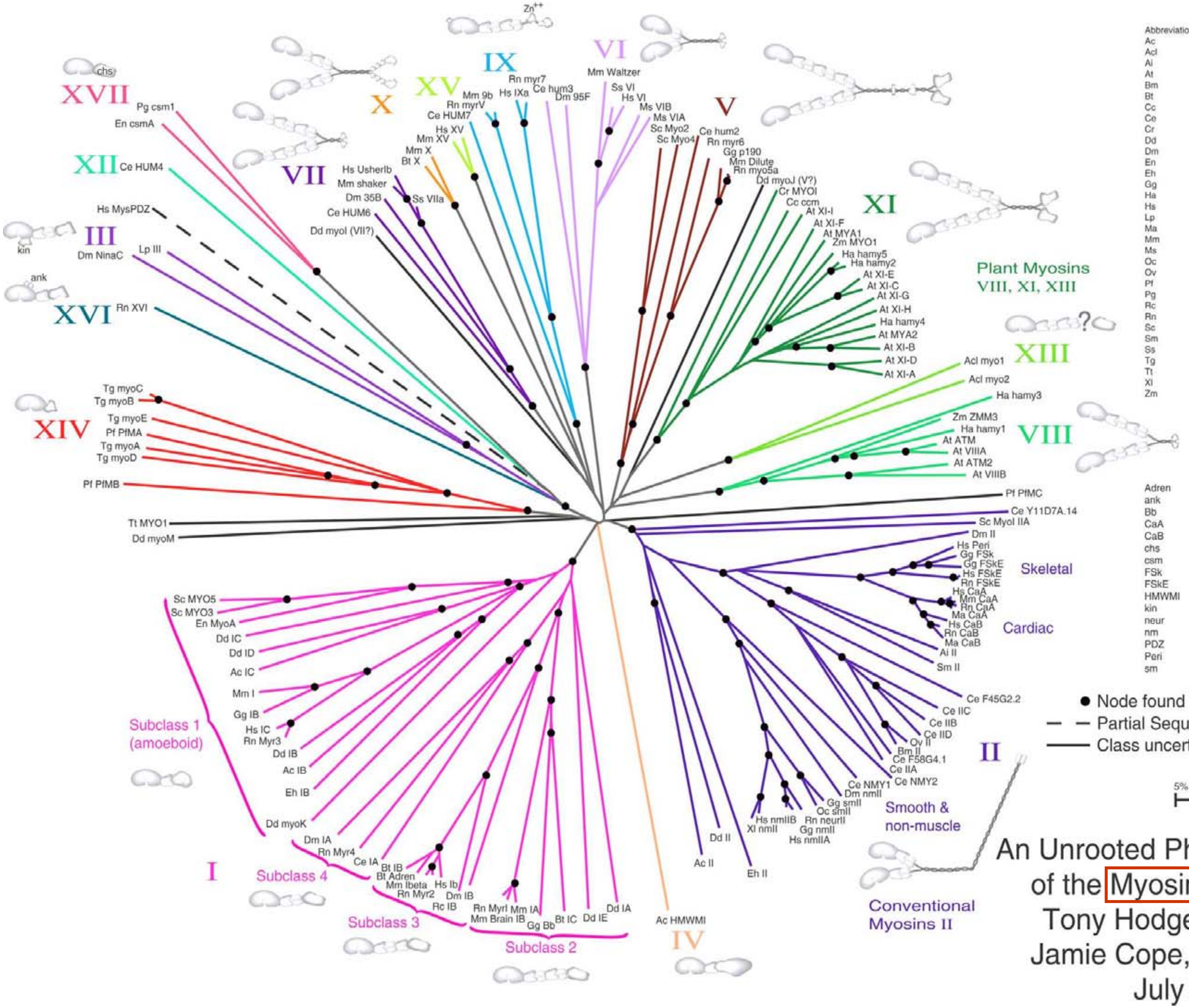
temporal evolutive tree



we are here



from where everything started



Abbreviation	Full Terms
Ac	Acanthamoeba castellanii
Acl	Acetabularia cliftonii
Ai	Aequipecten irradians (scallop)
At	Arabidopsis thaliana (thale cress)
Bm	Brugia malayi
Bt	Bos taurus (cow)
Cc	Chara corallina
Ce	Caenorhabditis elegans
Cr	Chlamydomonas reinhardtii
Dd	Dictyostelium discoidium
Dm	Drosophila melanogaster
En	Emiricella nidulans (Aspergillus)
Eh	Entamoeba histolytica
Gg	Gallus gallus (chicken)
Ha	Helianthus annus (sunflower)
Hs	Homo sapiens (human)
Lp	Limulus polyphemus (horseshoe crab)
Ma	Mesocricetus auratus (hamster)
Mm	Mus musculus (mouse)
Ms	Morone saxatilis (striped bass)
Oc	Oryctolagus cuniculus (rabbit)
Ov	Onchocerca volvulus (a nematode)
Pf	Plasmodium falciparum (malaria)
Pg	Pyricularia grisea (rice blast fungus)
Rc	Rana catesbeiana (bullfrog)
Rn	Rattus norvegicus (rat)
Sc	Saccharomyces cerevisiae (yeast)
Sm	Schistosoma mansoni
Ss	Sus scrofa (domestic pig)
Tg	Toxoplasma gondii
Tt	Tetrahymena thermophila
Xl	Xenopus laevis (clawed toad)
Zm	Zea mays (maize)

Adren	Bovine Adrenal (myosin I)
ank	Ankyrin like repeats
Bb	Brush Border Myosin I
CaA	Cardiac alpha (myosin II)
CaB	Cardiac beta (myosin II)
chs	Chitin synthase type V homology
csm	Chitin synthase-myosin
FSk	Fast Skeletal (myosin II) = striated
FSkE	Embryonic Fast Skeletal (myosin II)
HMWMI	High Molecular Weight Myosin I
kin	Kinase domain
neur	Neuronal (myosin II)
nm	Non-muscle (myosin II)
PDZ	Myosin like protein with a PDZ domain.
Peri	Perinatal (myosin II)
sm	Smooth muscle (myosin II)

An Unrooted Phylogenetic Tree
of the Myosin Superfamily
Tony Hodge, MRC-LMB
Jamie Cope, UC Berkeley
July 2000

Biological systems and Spin glasses

Biosystems

Disorder

very many random variables,
few dynamical (relevant) dof’s

Degeneracy

can exist in very many “equilibrium” states

Spin glasses

Disorder

random coupling among spins

Frustration

within triplets of spins

Complexity of classification

Spin glasses: a suggestive paradigm for biosystems

Protein folding (see below)	Iori Marinari Parisi
Associative memory	Hopfield
Scaling laws in taxonomy	Mezard Parisi Virasoro
Immune system memory and stability	Parisi
…	

A Spin glass Primer

- N individuals interacts pairwise with couplings

$J_{AB}=+1$	if	A likes B
$J_{AB}=-1$	if	A dislikes B

- Given 3 individuals, there is frustration if

$$J_{AB}\,J_{BC}\,J_{CA}=-1$$

- The N individuals are asked to separate in 2 fields so as to minimize in each field the number of pairs of "enemies"
- Given a J-PDF and an initial subdivision, "equilibrium" is reached by asking each individual to decide to change field if the move lowers the frustration
- If many pairs are frustrated
  - system is highly unstable

    ↓
  - many possible equally good subdivisions

A locally optimal state is reached in polynomial time

A globally optimal state (if it can be reached at all)
generically requires an exponential time (NP-problem)

An illuminating example

- M likes M   W likes W
  M dislikes W   W dislikes M
  → For any triplet $J^3=+1$
  No frustration

$\Rightarrow$ Optimal state: 2 separate groups, [M] and [W]

- M dislikes M   W dislikes W
  M likes W   W likes M
  → For any triplet $J^3=-1$
  Maximal frustration

$\Rightarrow$ Optimal state: any subdivision with equal number of M and W

Further examples of interesting physical systems

- Alloys, like $Fe_x\ Au_{100-x}$, with small x % → $H = \Sigma_{ik}\ \sigma_i\ J(|x_i-x_k|)\ \sigma_k$
  $J(|x_i-x_k|)$ very rapidly oscillating with $|x_i-x_k|$, almost a random function

- Electrons moving in a metallic glass, containing various types of atoms, located at fixed but random positions

$\Rightarrow$ We expect the electron conducibility not to depend on the detailed positions of the impurities (for not too small samples)

$$H_{SG} = \Sigma_{ik}\ \sigma_i\ J_{ik}\ \sigma_k \text{ , with some PDF for the } J_{ik}$$

Basic Mathematics

Sherrington
Kirkpatrick
Parisi

- Hamiltonian

$$H_J\,[\sigma] = \Sigma_{ik}\ \sigma_i\ J_{ik}\ \sigma_k \qquad\qquad J_{ik} = J_{ki}\ ,\ J_{ii} = 0$$

- J_{ik} are random variables with PDF $\Rightarrow$ P(J)

- Partition Function and Free Energy at fixed P(J)

$$Z_J = \Sigma_{[\sigma]}\ \exp -\beta\ H_J\,[\sigma] \qquad\qquad \beta=1/KT$$

$$F_J = -\frac{1}{\beta N}\log Z_J$$

- N is the number of spins

- We want to compute the quenched average

$$F = \Sigma_J\, P(J)\ F_J = -\frac{1}{\beta N}\Sigma_J\, P(J)\log Z_J$$

and not the annealed average

$$F_{An} = -\frac{1}{\beta N}\log Z_{An} \qquad\qquad Z_{An} = \Sigma_J\ P(J)\ \Sigma_{[\sigma]}\ \exp -\beta\ H_J\,[\sigma]$$

- time scale of J-dynamics >> time scale of σ-dynamics

The Replica Method

$$Z_n \equiv \Sigma_J P(J)\,(Z_J)^n$$

$$F_n = -\frac{1}{\beta N}\,\frac{1}{n}\log Z_n \qquad \Rightarrow \qquad \lim_{n\to 0} F_n = F$$

the replica index

A simple proof

$$\lim_{n\to 0} -\frac{1}{\beta N}\,\frac{1}{n}\log Z_n = \lim_{n\to 0}\boxed{-\frac{1}{\beta N}\,\frac{1}{n}\log\,[\Sigma_J P(J)\,(Z_J)^n]} =$$

$$= \lim_{n\to 0} -\frac{1}{\beta N}\,\frac{1}{n}\log\,[\Sigma_J P(J)\,(1+n\log Z_J + \ldots)] =$$

$$= \lim_{n\to 0} -\frac{1}{\beta N}\,\frac{1}{n}\log\,[1+n\,\Sigma_J P(J)\log Z_J + \ldots)] =$$

$$= -\frac{1}{\beta N}\,\Sigma_J P(J)\log Z_J = F$$

looks OK, except that n is an integer…

Typical P(J)'s

Gaussian: $P(J) \propto \exp[-(J-J_0)^2/2\sigma_J^2]$

Uniform: $P(J=+1) = P(J=-1) = 1/2$

Phase structure

Edwards
Anderson

$$m_i(J) = < \sigma_i > = \Sigma_{[\sigma]}\, \sigma_i \exp -\beta\, H_J\, [\sigma]$$

$$q(J) = \frac{1}{N} \Sigma_i\, [m_i(J)]^2 = \Sigma_J\, P(J)\, [m_i(J)]^2 = q$$

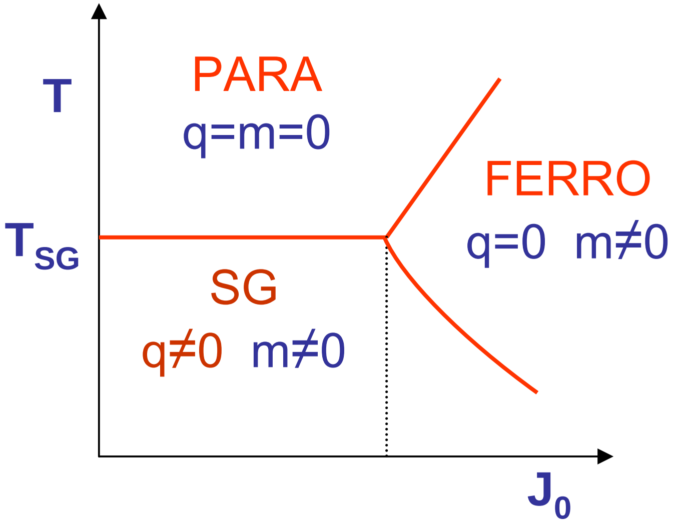
self-averaging

High temperature $m_i(J) = 0 \Rightarrow q = 0$

Low temperature $\left\{ \begin{array}{l} m_i(J) \neq 0 \text{ for some } i \\ \text{with } \Sigma_i\, [m_i(J)] = 0, \text{ but} \\ q(J) = \dfrac{1}{N} \Sigma_i\, [m_i(J)]^2 \neq 0 \end{array} \right.$

self-averaging

Order parameters $\left\{ \begin{array}{l} q = \dfrac{1}{N} \Sigma_i\, [m_i(J)]^2 \\ m = \dfrac{1}{N} \Sigma_i\, [m_i(J)] \end{array} \right.$



The whole game is to compute P(q)

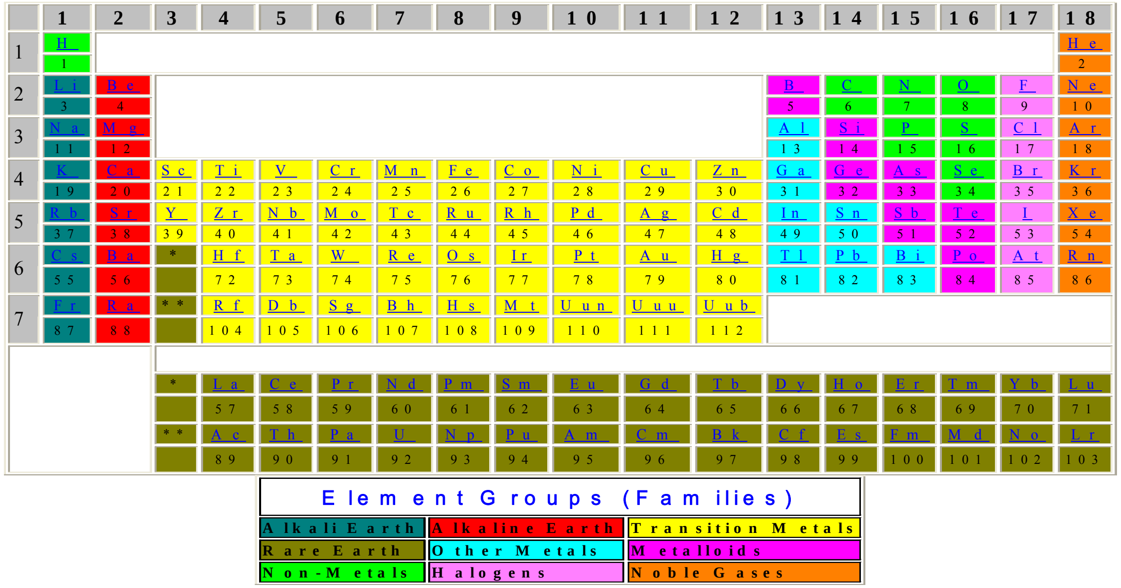
Few further numbers

dimensions
times
weights
chemical events

Human body: ~7 x 10^{27} atoms:
99% C, H, O and N; 87% are either H or O;
but 41 different elements

Estimated Atomic Composition of a lean 70 kg Male Human Body

Element	Sym		# Atoms	Element	Sym		# Atoms	Element	Sym		# Atoms
Hydrogen	H	1	4.22 x 10^{27}	Rubidium	Rb	37	2.2 x 10^{21}	Zirconium	Zr	40	2 x 10^{19}
Oxygen	O	8	1.61 x 10^{27}	Strontium	Sr	38	2.2 x 10^{21}	Cobalt	Co	27	2 x 10^{19}
Carbon	C	6	8.03 x 10^{26}	Bromine	Br	35	2 x 10^{21}	Cesium	Cs	55	7 x 10^{18}
Nitrogen	N	7	3.9 x 10^{25}	Aluminum	Al	13	1 x 10^{21}	Mercury	Hg	80	6 x 10^{18}
Calcium	Ca	20	1.6 x 10^{25}	Copper	Cu	29	7 x 10^{20}	Arsenic	As	33	6 x 10^{18}
Phosphorus	P	15	9.6 x 10^{24}	Lead	Pb	82	3 x 10^{20}	Chromium	Cr	24	6 x 10^{18}
Sulfur	S	16	2.6 x 10^{24}	Cadmium	Cd	48	3 x 10^{20}	Molybdenum	Mo	42	3 x 10^{18}
Sodium	Na	11	2.5 x 10^{24}	Boron	B	5	2 x 10^{20}	Selenium	Se	34	3 x 10^{18}
Potassium	K	19	2.2 x 10^{24}	Manganese	Mn	25	1 x 10^{20}	Beryllium	Be	4	3 x 10^{18}
Chlorine	Cl	17	1.6 x 10^{24}	Nickel	Ni	28	1 x 10^{20}	Vanadium	V	23	8 x 10^{17}
Magnesium	Mg	12	4.7 x 10^{23}	Lithium	Li	3	1 x 10^{20}	Uranium	U	92	2 x 10^{17}
Silicium	Si	14	3.9 x 10^{23}	Barium	Ba	56	8 x 10^{19}	Radium	Ra	88	8 x 10^{10}
Fluorine	F	9	8.3 x 10^{22}	Iodine	I	53	5 x 10^{19}				
Iron	Fe	26	4.5 x 10^{22}	Tin	Sn	50	4 x 10^{19}				
Zinc	Zn	30	2.1 x 10^{22}	Gold	Au	79	2 x 10^{19}	TOTAL			6.71x10^{27}



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	H 1																	He 2
2	Li 3	Be 4											B 5	C 6	N 7	O 8	F 9	Ne 10
3	Na 11	Mg 12											Al 13	Si 14	P 15	S 16	Cl 17	Ar 18
4	K 19	Ca 20	Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28	Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36
5	Rb 37	Sr 38	Y 39	Zr 40	Nb 41	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54
6	Cs 55	Ba 56	*	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	At 85	Rn 86
7	Fr 87	Ra 88	**	Rf 104	Db 105	Sg 106	Bh 107	Hs 108	Mt 109	Uun 110	Uuu 111	Uub 112						
			*	La 57	Ce 58	Pr 59	Nd 60	Pm 61	Sm 62	Eu 63	Gd 64	Tb 65	Dy 66	Ho 67	Er 68	Tm 69	Yb 70	Lu 71
			**	Ac 89	Th 90	Pa 91	U 92	Np 93	Pu 94	Am 95	Cm 96	Bk 97	Cf 98	Es 99	Fm 100	Md 101	No 102	Lr 103

Element Groups (Families)

Alkali Earth	Alkaline Earth	Transition Metals
Rare Earth	Other Metals	Metalloids
Non-Metals	Halogens	Noble Gases

Estimated Molecular Content of a Typical 20-micron Human Cell

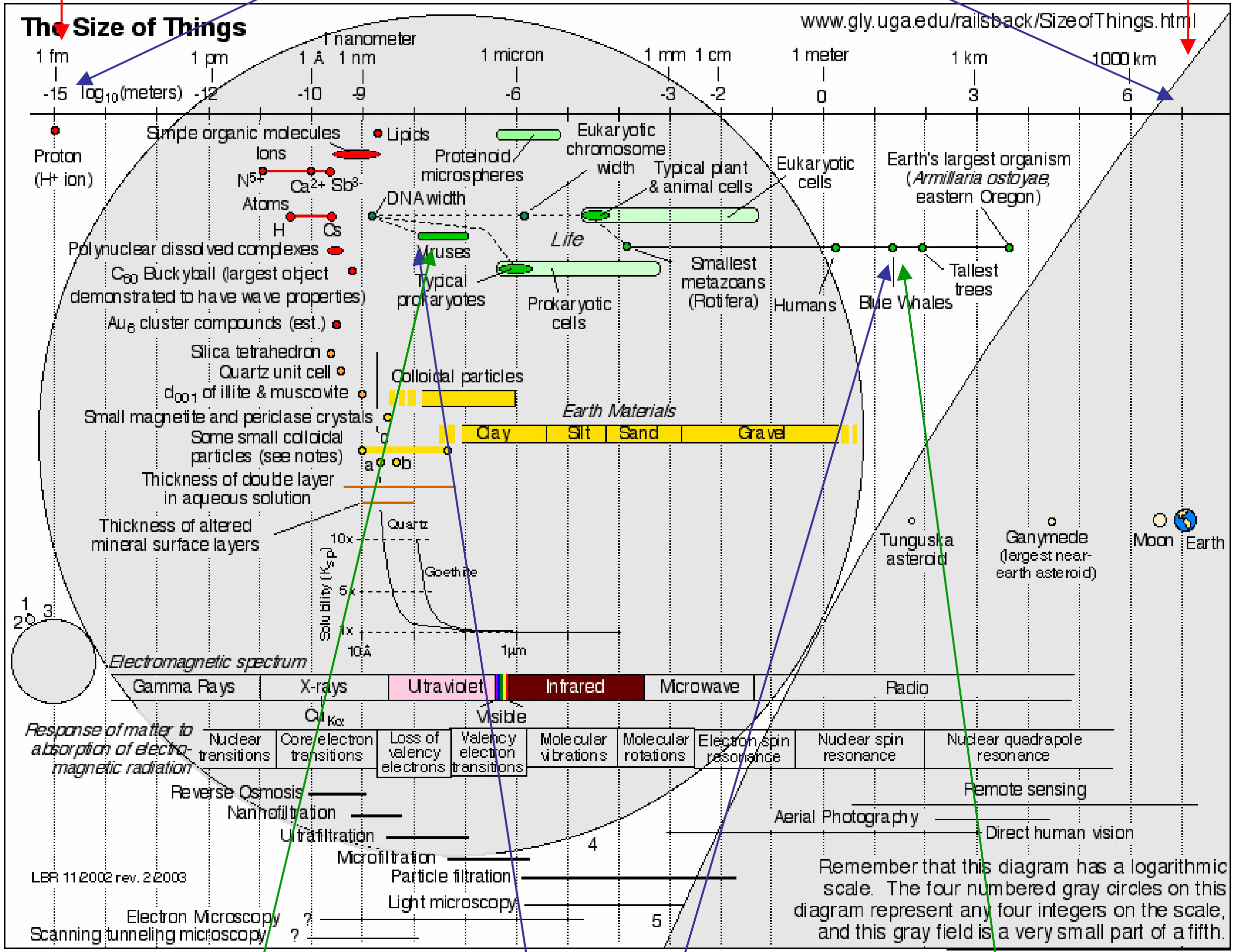
Molecule	Mass %	<MW> (Daltons)	# Molecules	Molecule %	# of Types
Water	65%	18	1.74×10^{14}	98.73 %	1
Other Inorganic	1.5%	55	1.31×10^{12}	0.74 %	20
Lipid	12%	700	8.4×10^{11}	0.475 %	50
Other Organic	0.4%	250	7.7×10^{10}	0.044 %	~200
Protein	20%	50,000	1.9×10^{10}	0.011 %	~5,000
RNA	1.0%	1×10^{6}	5×10^{7}	3×10^{-5} %	----
DNA	0.1%	1×10^{11}	46	3×10^{-11} %	----
TOTALS	100%	----	1.76×10^{14}	100%	----

1 Da (Dalton) = 1 atomic unit = $m_a(^{12}C)/(12 \times 1{,}660540\ 10^{-27}$ kg ~ hydrogen mass)
dimensionless unit

Proton

10^{22}

Earth

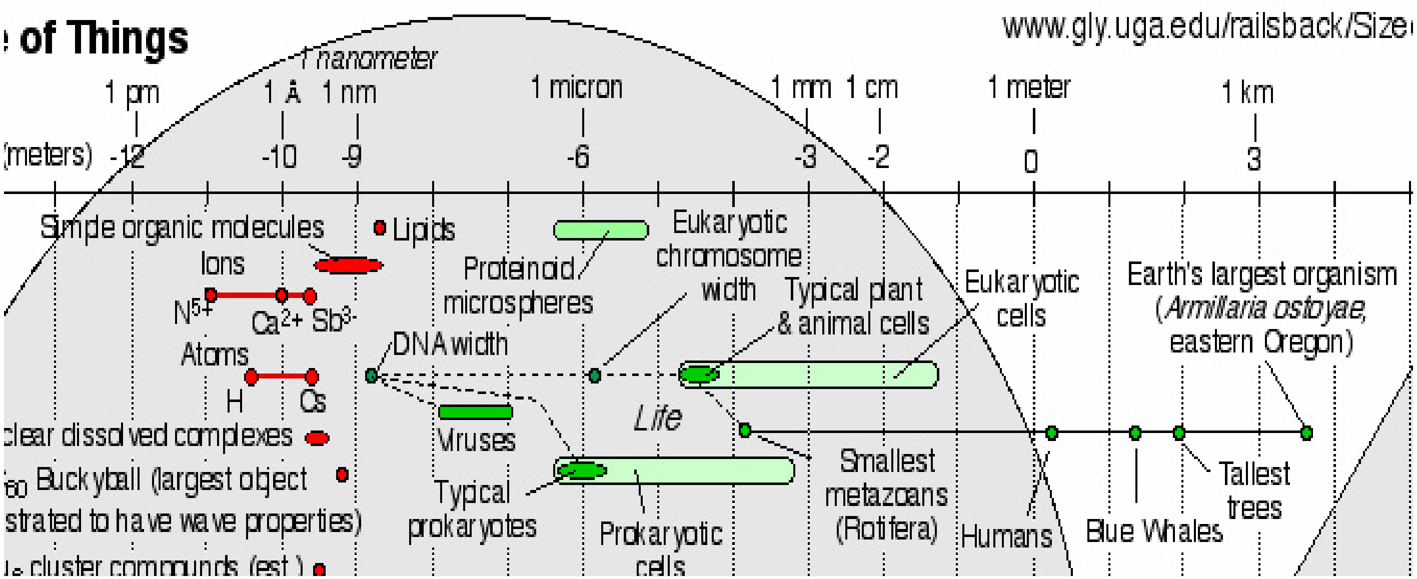


Virus

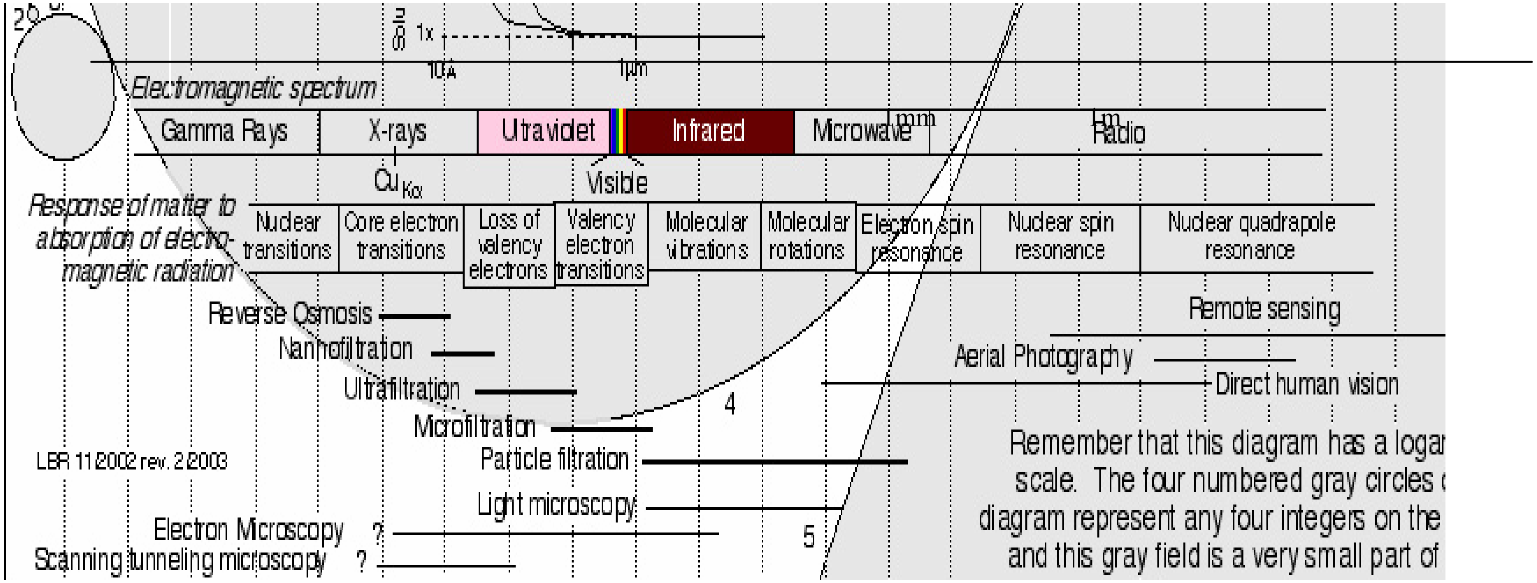
10^{9}

Blue whale

A blow up



of Things
www.gly.uga.edu/railsback/Size
1 nanometer
1 pm
1 Å
1 nm
1 micron
1 mm
1 cm
1 meter
1 km
(meters)
-12
-10
-9
-6
-3
-2
0
3
Simple organic molecules
Lipids
Ions
Proteinoid microspheres
Eukaryotic chromosome width
N5+
Ca2+
Sb3-
Typical plant & animal cells
Eukaryotic cells
Earth's largest organism (Armillaria ostoyae, eastern Oregon)
Atoms
H
Cs
DNA width
clear dissolved complexes
Viruses
Life
Buckyball (largest object
strated to have wave properties)
cluster compounds (est.)
Typical prokaryotes
Prokaryotic cells
Smallest metazoans (Rotifera)
Humans
Blue Whales
Tallest trees



1x
10Å
1μm
Electromagnetic spectrum
Gamma Rays
X-rays
Ultraviolet
Infrared
Microwave
mm
km
Radio
Cu Kα
Visible
Response of matter to absorption of electro-magnetic radiation
Nuclear transitions
Core electron transitions
Loss of valency electrons
Valency electron transitions
Molecular vibrations
Molecular rotations
Electron spin resonance
Nuclear spin resonance
Nuclear quadrapole resonance
Reverse Osmosis
Nanofiltration
Ultrafiltration
Microfiltration
Particle filtration
Light microscopy
Electron Microscopy ?
Scanning tunneling microscopy ?
Remote sensing
Aerial Photography
Direct human vision
4
5
LBR 11/2002 rev. 2/2003
Remember that this diagram has a logar
scale. The four numbered gray circles
diagram represent any four integers on the
and this gray field is a very small part of

The largest and smallest cells in the human body are the gametes or the sex cells

♀ female = oocyte: Ø ≈ 35 μm (almost visible with the naked eye)

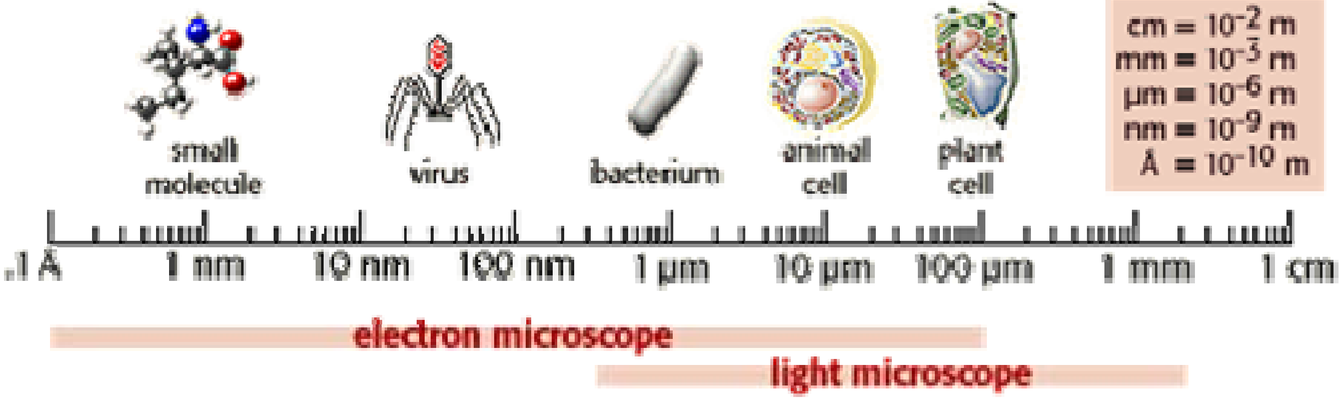
♂ male = spermatozoon: Ø ≈ 3 μm

The smallest known organism capable of independent growth and reproduction

Mycoplasma genitalium: Ø ≈ 0.2 - 0.3 μm

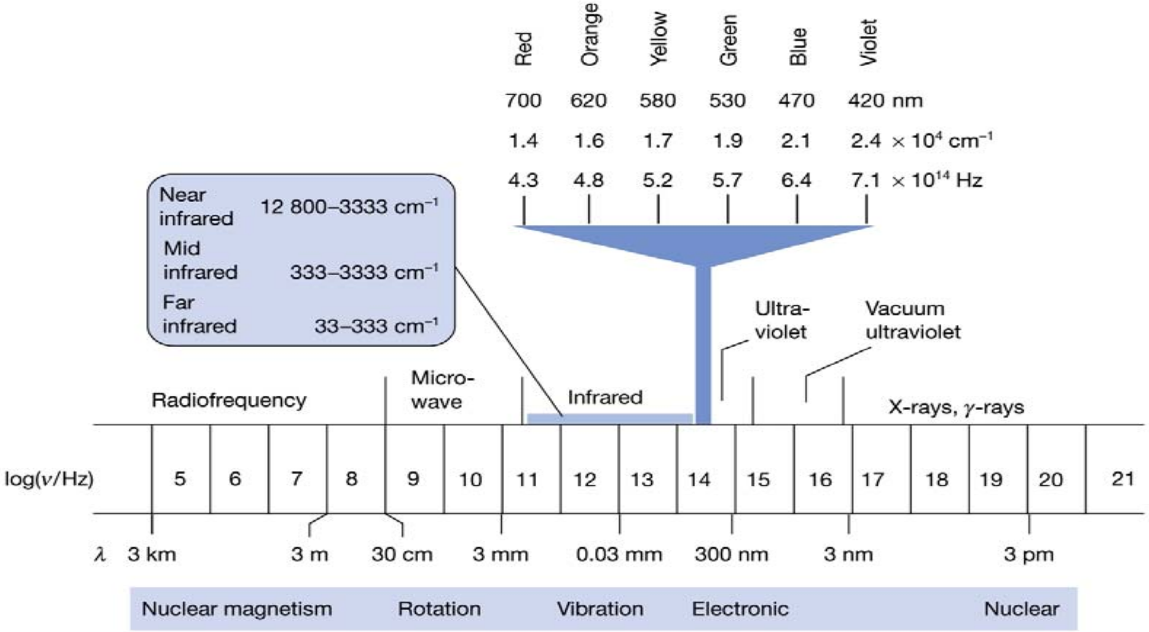
The smallest “theoretical” bacterium: Ø ≈ 0.17 μm

Relative sizes of cells and their components

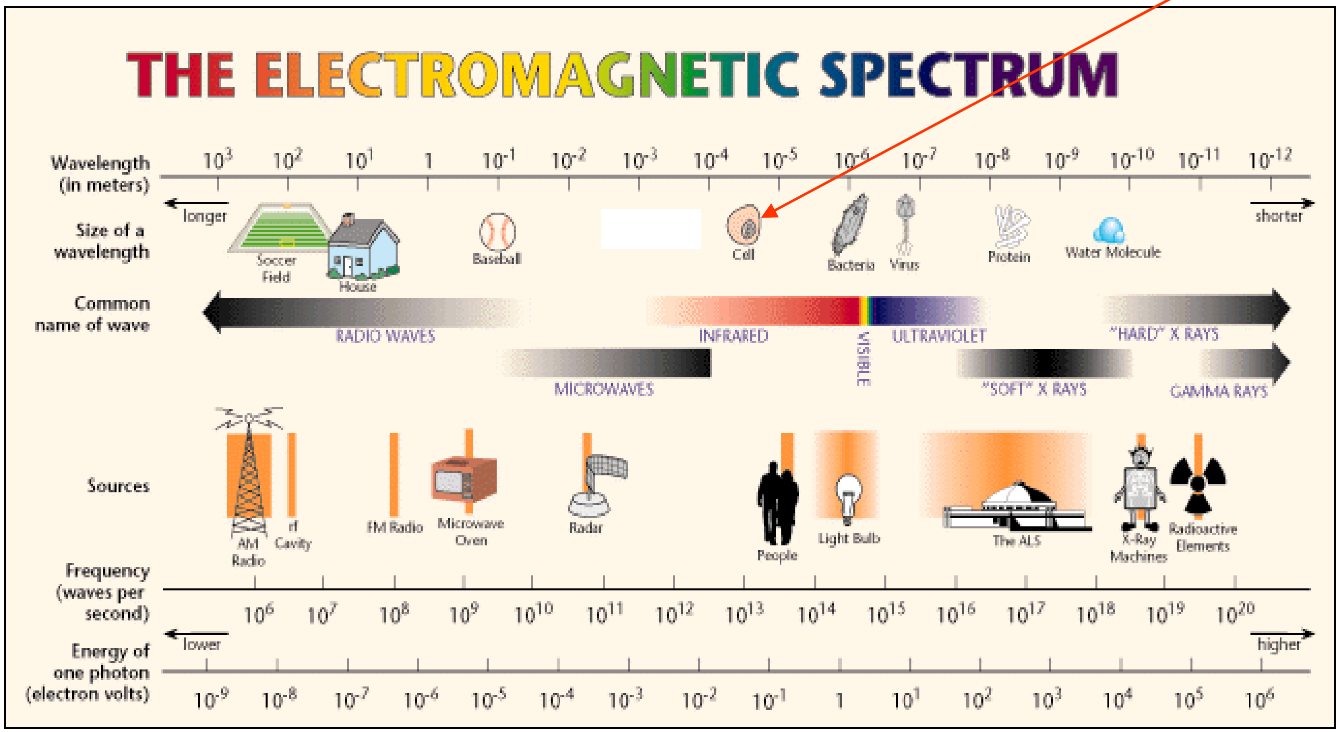


<Average bacterium>: rod shape V ≈ 1 μm^2 x 3 μm

<Average human cell>: spherical shape Ø ≈ 25 μm

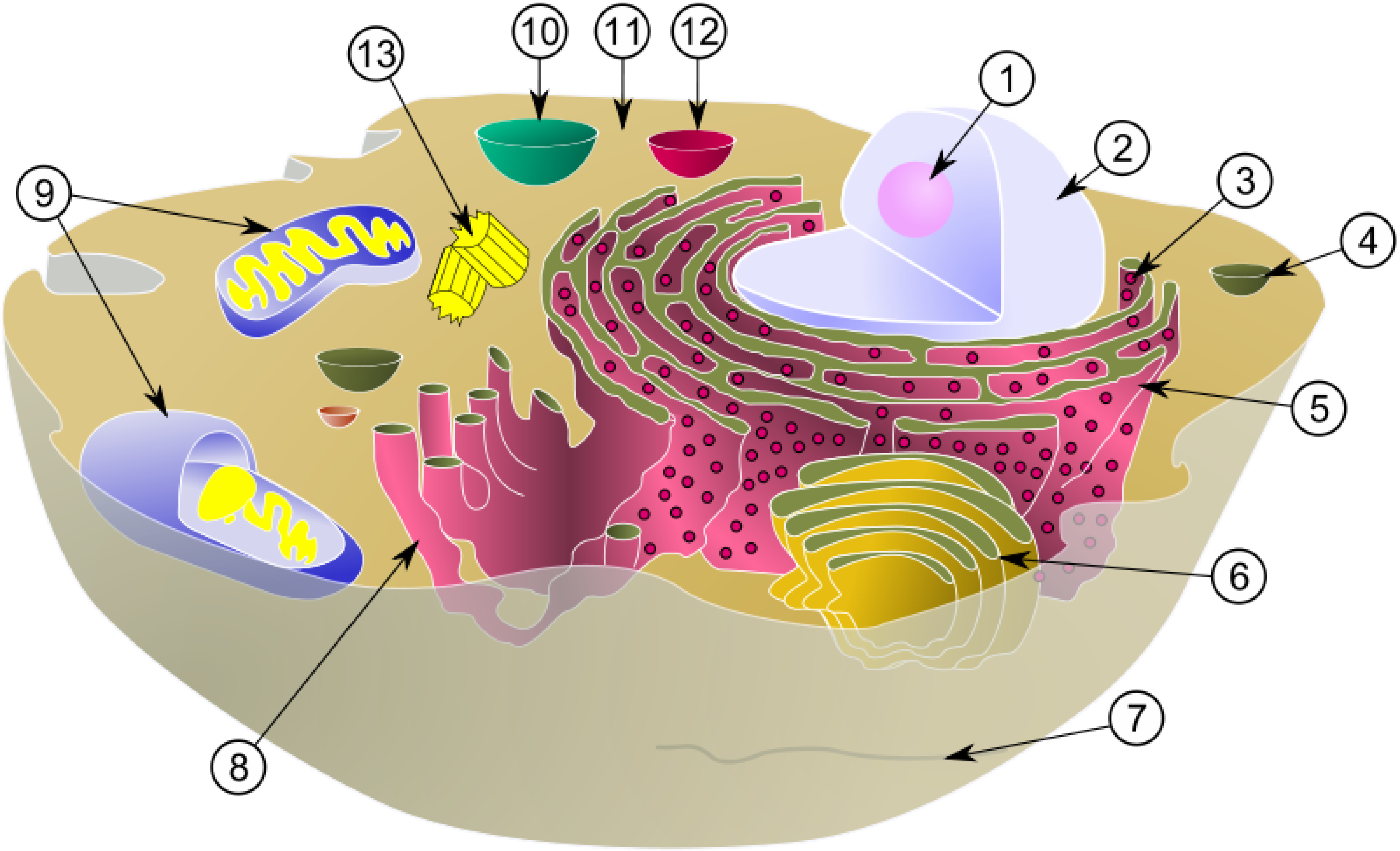


Red
Orange
Yellow
Green
Blue
Violet
700 620 580 530 470 420 nm
1.4 1.6 1.7 1.9 2.1 2.4 × 10⁴ cm⁻¹
4.3 4.8 5.2 5.7 6.4 7.1 × 10¹⁴ Hz
Near infrared 12 800–3333 cm⁻¹
Mid infrared 333–3333 cm⁻¹
Far infrared 33–333 cm⁻¹
Ultra-violet
Vacuum ultraviolet
Radiofrequency
Micro-wave
Infrared
X-rays, γ-rays
log(ν/Hz)
5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
λ 3 km 3 m 30 cm 3 mm 0.03 mm 300 nm 3 nm 3 pm
Nuclear magnetism
Rotation
Vibration
Electronic
Nuclear



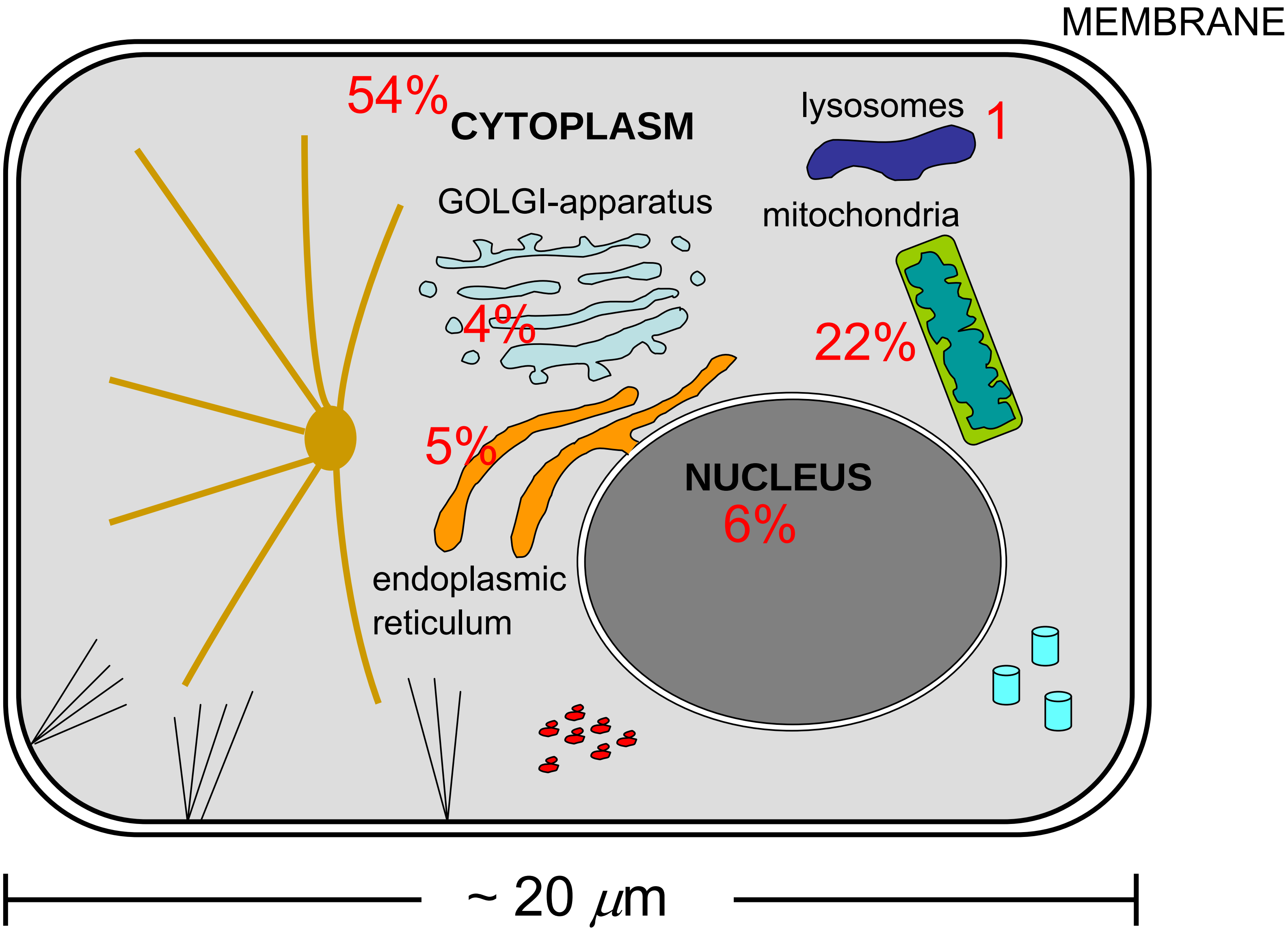
THE ELECTROMAGNETIC SPECTRUM
Wavelength (in meters)
10³ 10² 10¹ 1 10⁻¹ 10⁻² 10⁻³ 10⁻⁴ 10⁻⁵ 10⁻⁶ 10⁻⁷ 10⁻⁸ 10⁻⁹ 10⁻¹⁰ 10⁻¹¹ 10⁻¹²
longer
shorter
Size of a wavelength
Soccer Field
House
Baseball
Cell
Bacteria
Virus
Protein
Water Molecule
Common name of wave
RADIO WAVES
INFRARED
VISIBLE
ULTRAVIOLET
"HARD" X RAYS
MICROWAVES
"SOFT" X RAYS
GAMMA RAYS
Sources
AM Radio
rf Cavity
FM Radio
Microwave Oven
Radar
People
Light Bulb
The ALS
X-Ray Machines
Radioactive Elements
Frequency (waves per second)
10⁶ 10⁷ 10⁸ 10⁹ 10¹⁰ 10¹¹ 10¹² 10¹³ 10¹⁴ 10¹⁵ 10¹⁶ 10¹⁷ 10¹⁸ 10¹⁹ 10²⁰
lower
higher
Energy of one photon (electron volts)
10⁻⁹ 10⁻⁸ 10⁻⁷ 10⁻⁶ 10⁻⁵ 10⁻⁴ 10⁻³ 10⁻² 10⁻¹ 1 10¹ 10² 10³ 10⁴ 10⁵ 10⁶

A eukaryotic cell artistic view

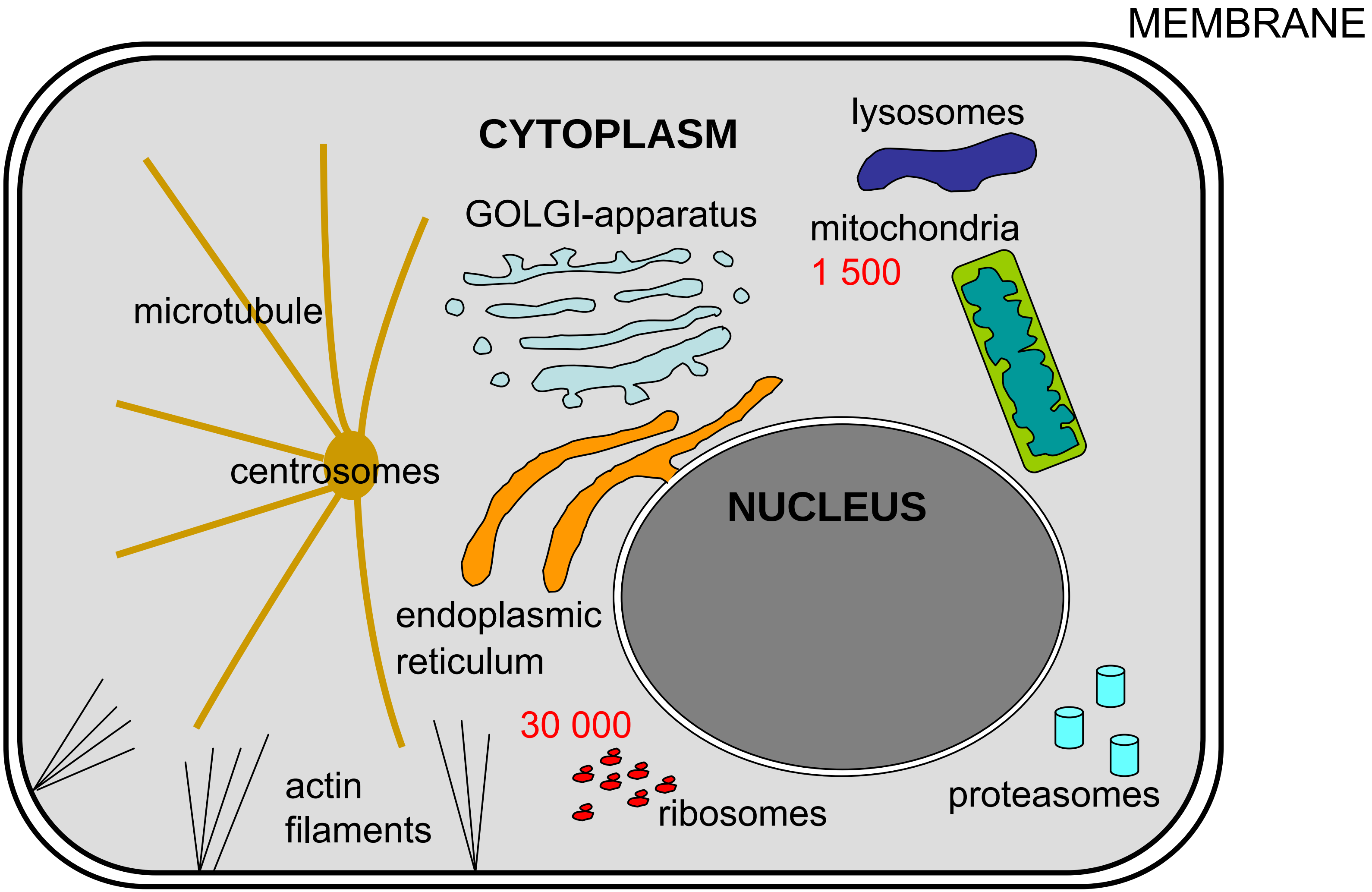


1. Nucleolus
2. Nucleus
3. Ribosome
4. Vesicle
5. Rough endoplasmic reticulum
6. Golgi apparatus
7. Cytoskeleton
8. Smooth endoplasmic reticulum
9. Mitochondrion
10. Vacuole
11. Cytosol
12. Lysosome
13. Centriole

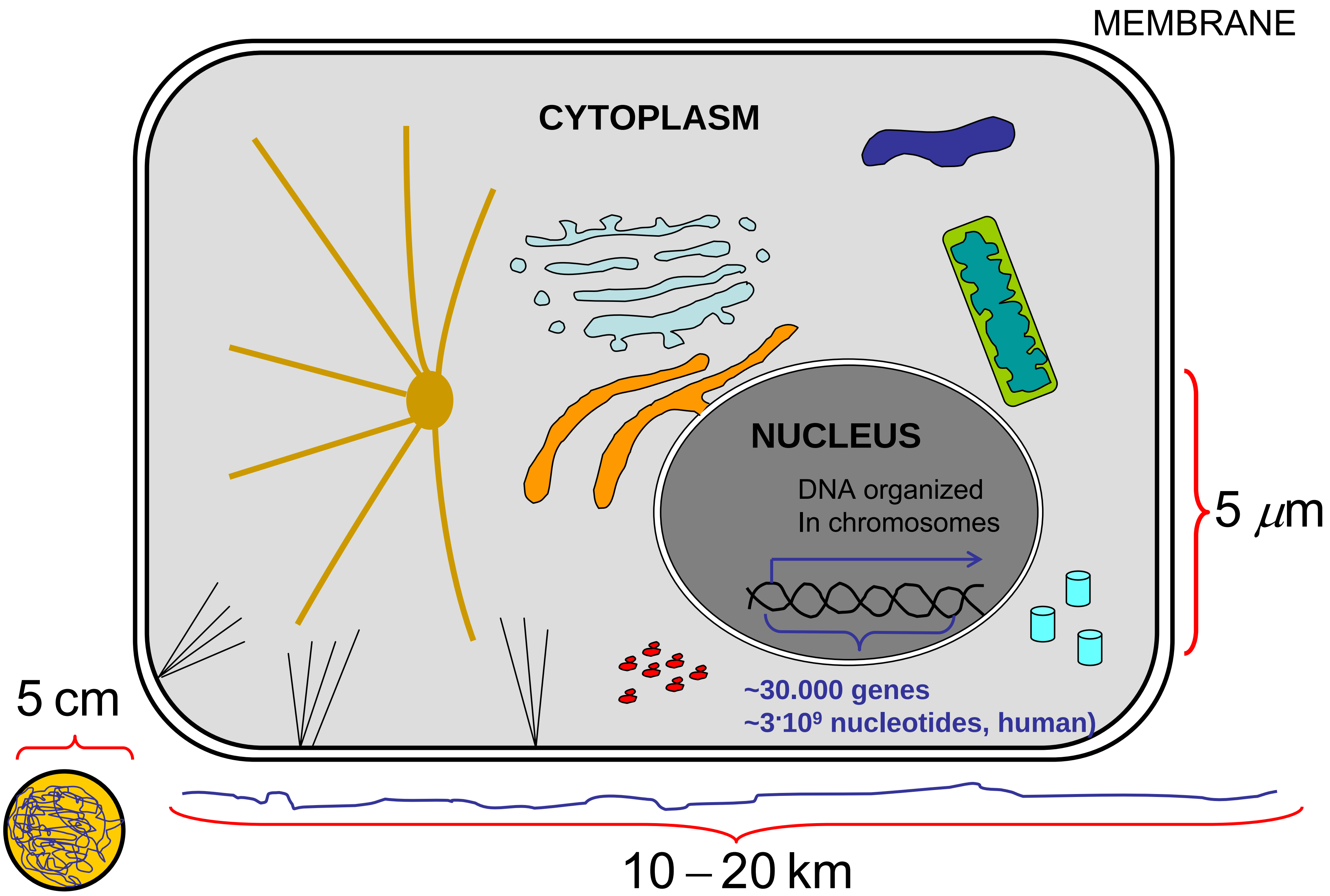
SCHEMATIC STRUCTURE OF LIVING EUCARYOTIC CELLS



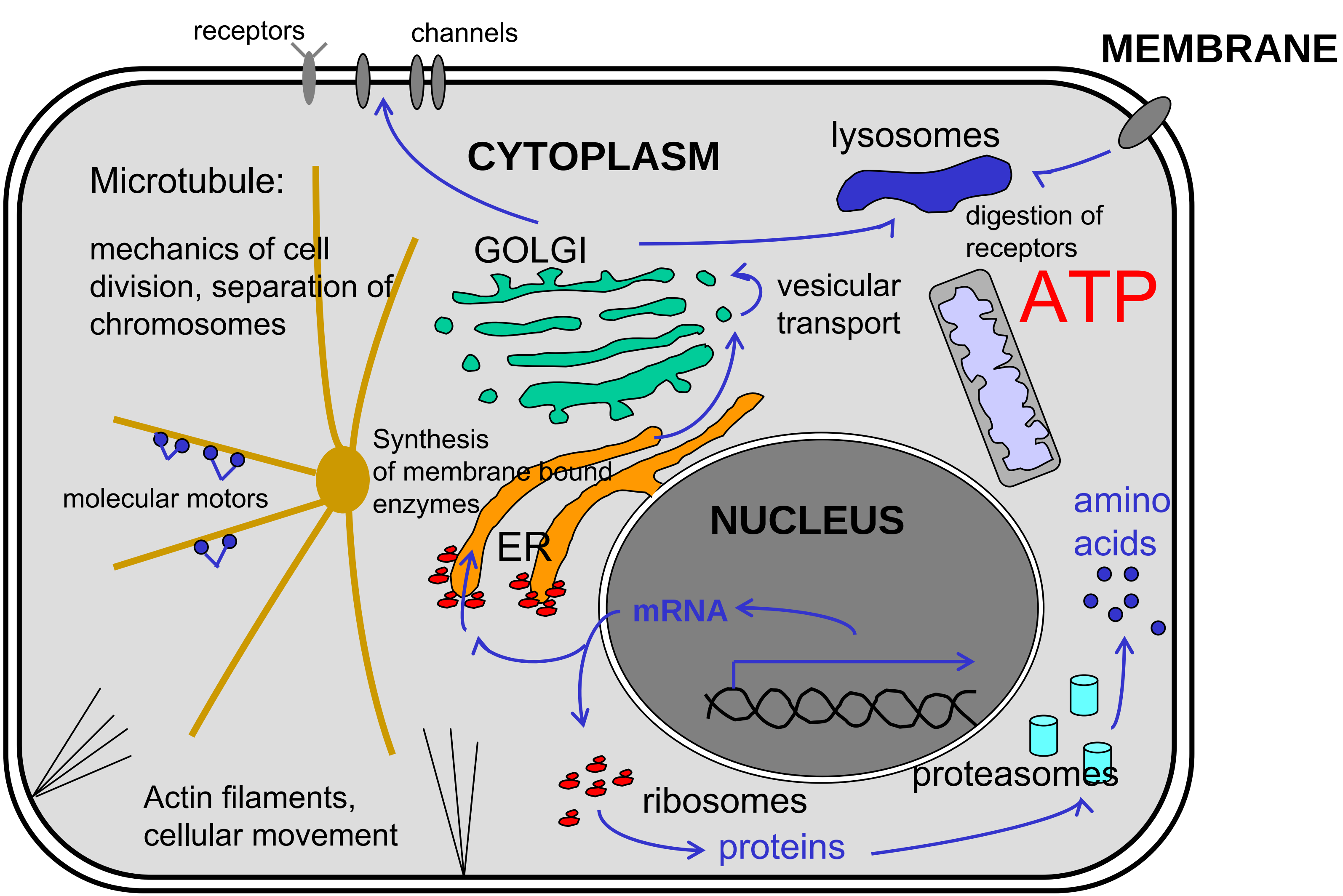
SCHEMATIC STRUCTURE OF LIVING EUCARYOTIC CELLS



SCHEMATIC STRUCTURE OF LIVING EUCARYOTIC CELLS



SCHEMATIC STRUCTURE OF LIVING EUCARYOTIC CELLS



Comparison of features of prokaryotic and eukaryotic cells

	Prokaryotes	Eukaryotes
Typical organisms	bacteria, archaea	protists, fungi, plants, animals
Typical size	~ 1-10 µm	~ 10-100 µm (sperm cells, apart from the tail, are smaller)
Type of nucleus	nucleoid region; no real nucleus	real nucleus with double membrane
DNA	circular (usually)	linear molecules (chromosomes) with histone proteins
RNA-/protein-synthesis	coupled in cytoplasm	RNA-synthesis inside the nucleus protein synthesis in cytoplasm
Ribosomes	50S+30S	60S+40S
Cytoplasmatic structure	very few structures	highly structured by endomembranes and a cytoskeleton
Cell movement	flagella made of flagellin	flagella and cilia containing microtubules; lamellipodia and filopodia containing actin
Mitochondria	none	one to several thousand (though some lack mitochondria)
Chloroplasts	none	in algae and plants
Organization	usually single cells	single cells, colonies, higher multicellular organisms with specialized cells
Cell division	Binary fission (simple division)	Mitosis (fission or budding) Meiosis

II. Data, (physical) models and (mathematical) tools

II. Data, (physical) models and (mathematical) tools

Bio data

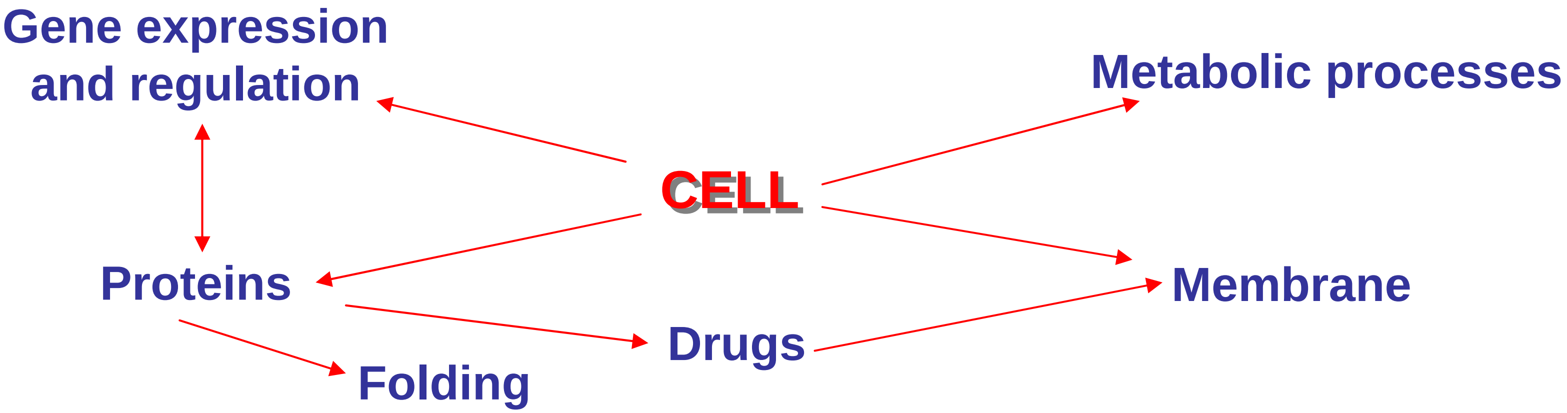
- Genome/Protein sequencing

  genome sequence reconstruction is an NP-hard problem
- Annotation

  elucidation and description of biologically relevant features in the sequence and relations with other data
- Identification of gene regulation and metabolic pathways

  reaction constants and chemical affinities
- ……………….

Ontologies

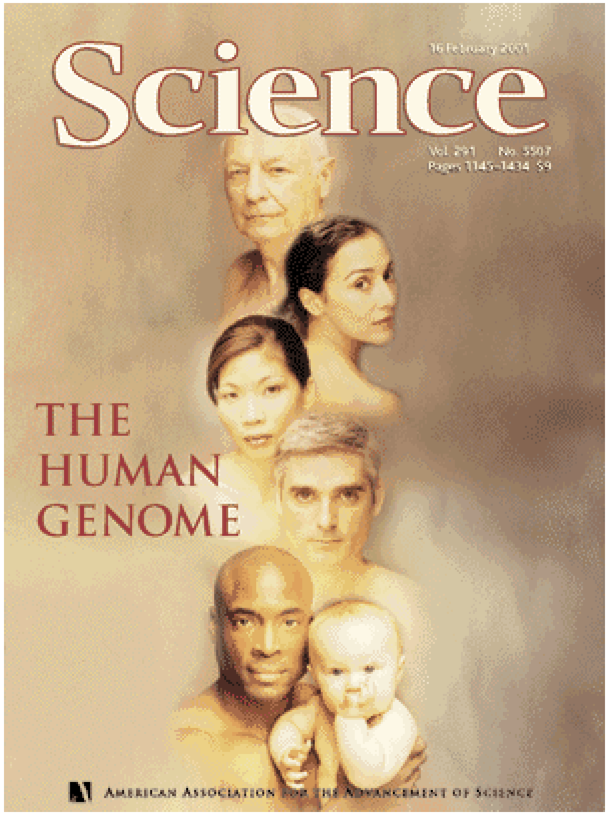


Modelling

- How to make a model
- Analytical methods
- Numerical approaches and simulations

D.D. Shoemaker et al.
15 February 2001
Vol. **409**, pp. 745-964

Human Genome Project
HGP



J. Craig Venter et al.
16 February 2001
Vol. **291**, pp. 1304-1351

CELERA

Genome Overview

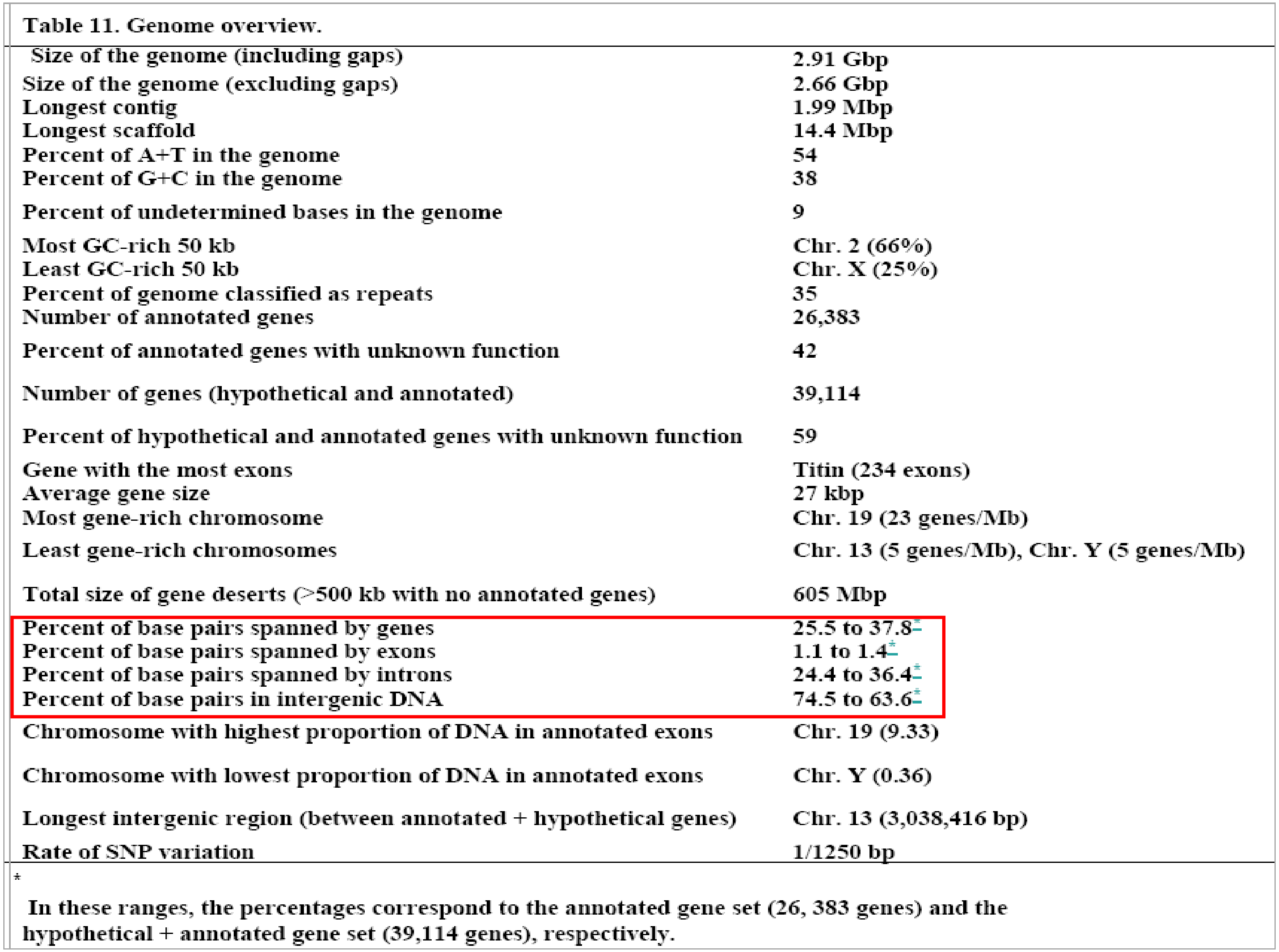


Table 11. Genome overview.

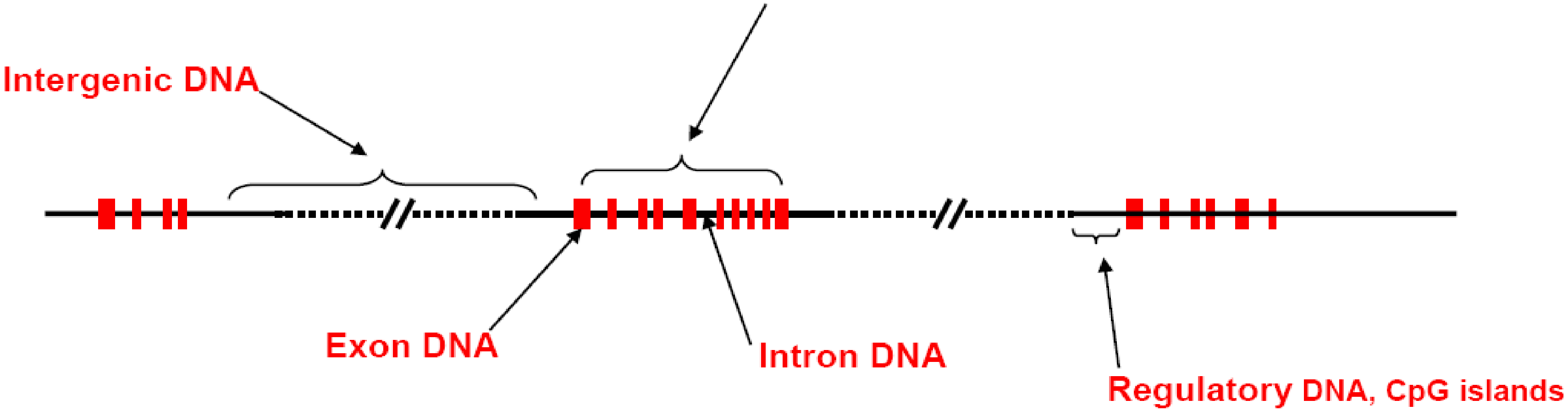
Size of the genome (including gaps)	2.91 Gbp
Size of the genome (excluding gaps)	2.66 Gbp
Longest contig	1.99 Mbp
Longest scaffold	14.4 Mbp
Percent of A+T in the genome	54
Percent of G+C in the genome	38
Percent of undetermined bases in the genome	9
Most GC-rich 50 kb	Chr. 2 (66%)
Least GC-rich 50 kb	Chr. X (25%)
Percent of genome classified as repeats	35
Number of annotated genes	26,383
Percent of annotated genes with unknown function	42
Number of genes (hypothetical and annotated)	39,114
Percent of hypothetical and annotated genes with unknown function	59
Gene with the most exons	Titin (234 exons)
Average gene size	27 kbp
Most gene-rich chromosome	Chr. 19 (23 genes/Mb)
Least gene-rich chromosomes	Chr. 13 (5 genes/Mb), Chr. Y (5 genes/Mb)
Total size of gene deserts (>500 kb with no annotated genes)	605 Mbp
Percent of base pairs spanned by genes	25.5 to 37.8[*]
Percent of base pairs spanned by exons	1.1 to 1.4[*]
Percent of base pairs spanned by introns	24.4 to 36.4[*]
Percent of base pairs in intergenic DNA	74.5 to 63.6[*]
Chromosome with highest proportion of DNA in annotated exons	Chr. 19 (9.33)
Chromosome with lowest proportion of DNA in annotated exons	Chr. Y (0.36)
Longest intergenic region (between annotated + hypothetical genes)	Chr. 13 (3,038,416 bp)
Rate of SNP variation	1/1250 bp

[*] In these ranges, the percentages correspond to the annotated gene set (26, 383 genes) and the hypothetical + annotated gene set (39,114 genes), respectively.

Genetic structure of human chromosomes

Each human chromosome contains a single long DNA molecule.

A gene is a locus of co-transcribed exons



Intergenic DNA

Exon DNA

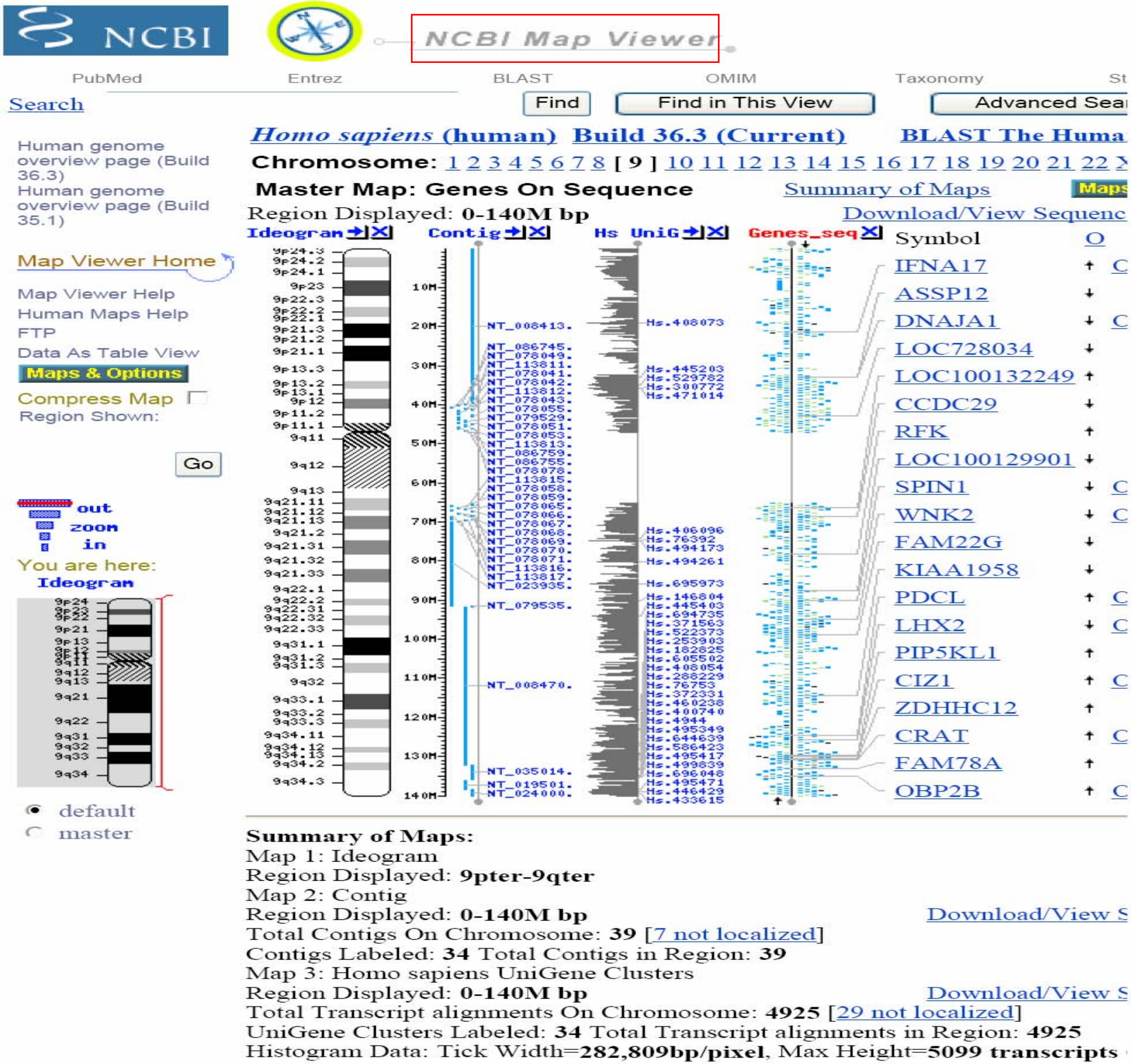
Intron DNA

Regulatory DNA, CpG islands

Intergenic DNA	75%
Intron DNA	24%
Exon DNA	1.1%

There are approximately **38,000 human genes** predicted from analysis of the human genome sequence

Human chromosome 9



Homo sapiens (human) Build 36.3 (Current) BLAST The Huma

Chromosome: 1 2 3 4 5 6 7 8 [9] 10 11 12 13 14 15 16 17 18 19 20 21 22 X

Master Map: Genes On Sequence

Summary of Maps

Maps

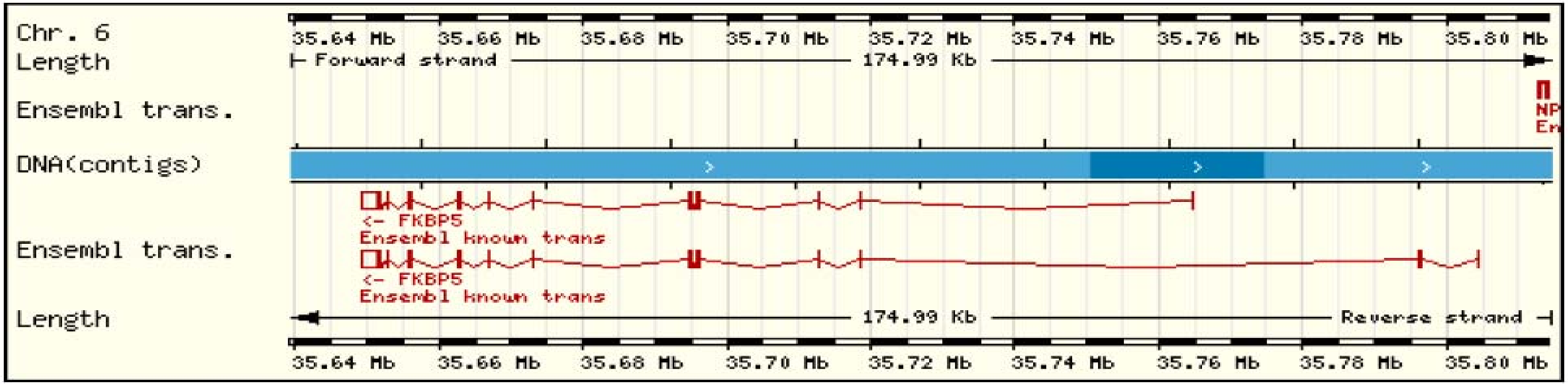
Region Displayed: 0-140M bp

Download/View Sequenc

Symbol		O
IFNA17	↑	C
ASSP12	↓	
DNAJA1	↓	C
LOC728034	↓	
LOC100132249	↑	
CCDC29	↓	
RFK	↑	
LOC100129901	↓	
SPIN1	↓	C
WNK2	↓	C
FAM22G	↓	
KIAA1958	↓	
PDCL	↑	C
LHX2	↓	C
PIP5KL1	↑	
CIZ1	↑	C
ZDHHC12	↑	
CRAT	↑	C
FAM78A	↑	
OBP2B	↑	C

Summary of Maps:
Map 1: Ideogram
Region Displayed: **9pter-9qter**
Map 2: Contig
Region Displayed: **0-140M bp** Download/View S
Total Contigs On Chromosome: **39** [7 not localized]
Contigs Labeled: **34** Total Contigs in Region: **39**
Map 3: Homo sapiens UniGene Clusters
Region Displayed: **0-140M bp** Download/View S
Total Transcript alignments On Chromosome: **4925** [29 not localized]
UniGene Clusters Labeled: **34** Total Transcript alignments in Region: **4925**
Histogram Data: Tick Width=**282,809bp/pixel**, Max Height=**5099 transcripts**

Gene Report for ENSG00000096060, FKBP5

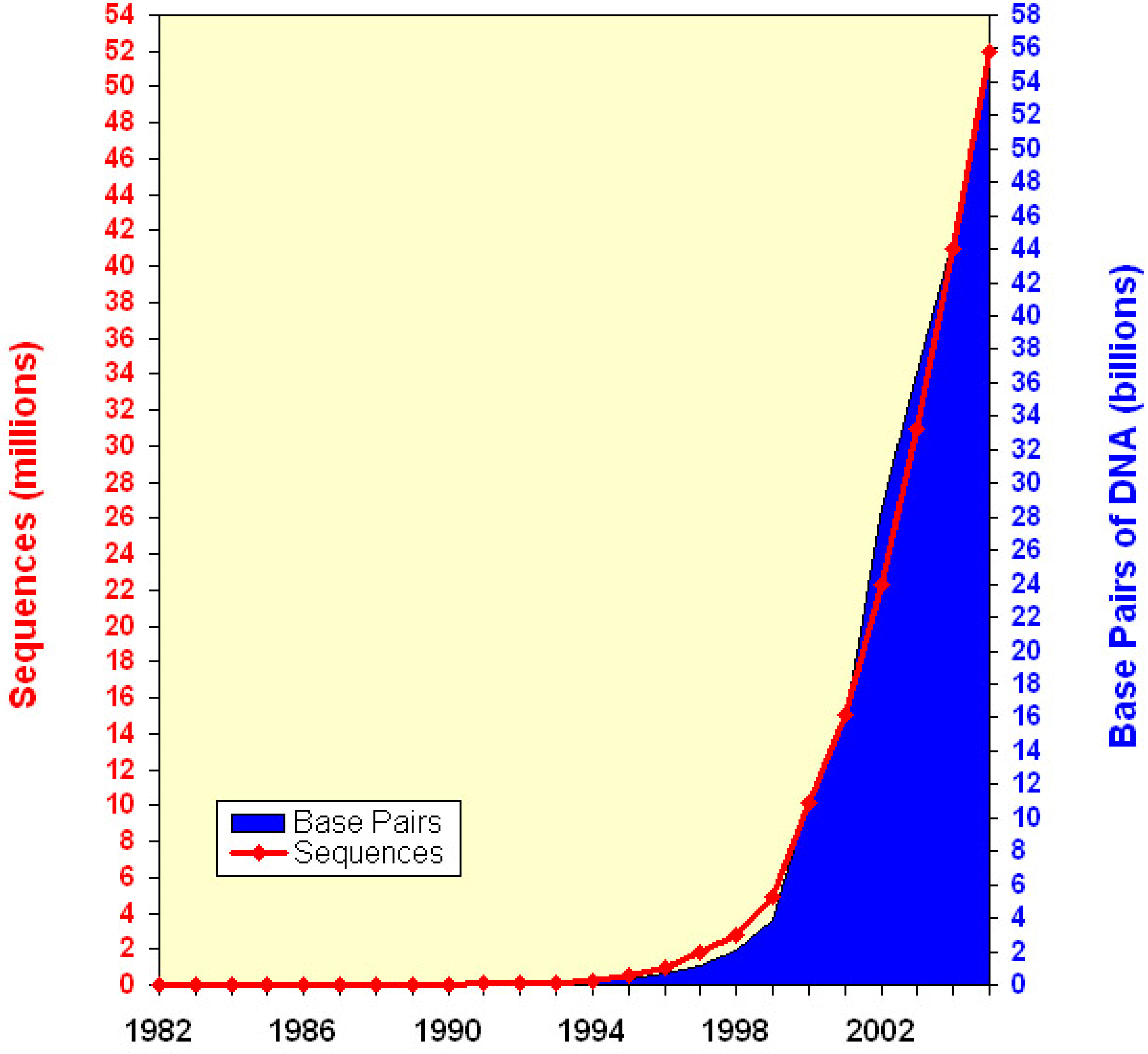


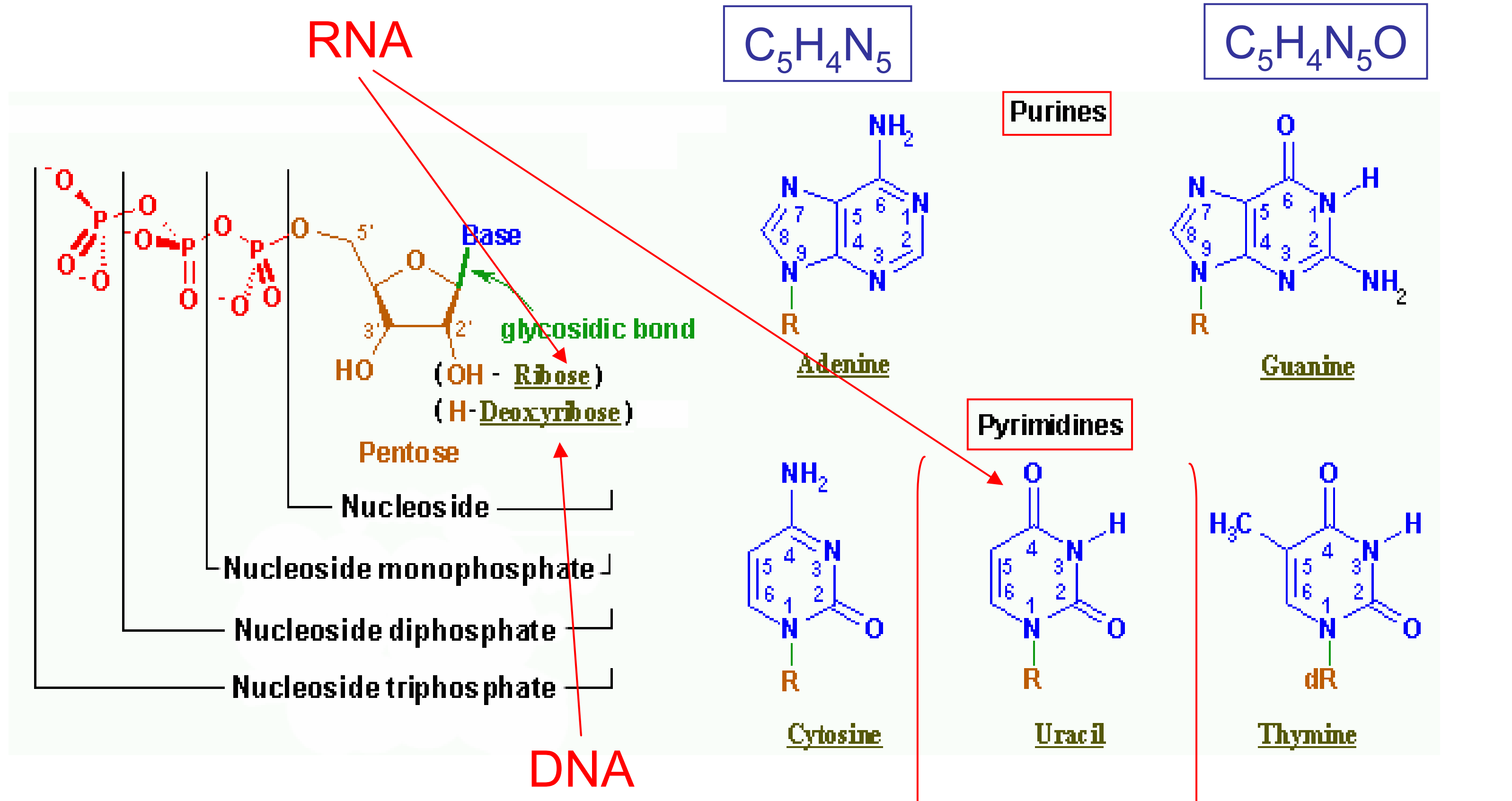
Nucleotide Sequence (1374 nt):

```
ATGACTACTGATGAAGGTGCCAAGAACAATGAAGAAAGCCCCACAGCCACTGTTGCTGAGCAGGGAGAGG
ATATTACCTCCAAAAAAGACAGGGGAGTATTAAAGATTGTCAAAAGAGTGGGGAATGGTGAGGAAACGCC
GATGATTGGAGACAAAGTTTATGTCCATTACAAAGGAAAATTGTCAAATGGAAAGAAGTTTGATTCCAGT
CATGATAGAAATGAACCATTTGTCTTTAGTCTTGGCAAAGGCCAAGTCATCAAGGCATGGGACATTGGGG
TGGCTACCATGAAGAAAGGAGAGATATGCCATTTACTGTGCAAACCAGAATATGCATATGGCTCGGCTGG
CAGTCTCCCTAAAATTCCCTCGAATGCAACTCTCTTTTTTGAGATTGAGCTCCTTGATTTCAAAGGAGAG
GATTTATTTGAAGATGGAGGCATTATCCGGAGAACCAAACGGAAAGGAGAGGGATATTCAAATCCAAACG
AAGGAGCAACAGTAGAAATCCACCTGGAAGGCCGCTGTGGTGGAAGGATGTTTGACTGCAGAGATGTGGC
ATTCACTGTGGGCGAAGGAGAAGACCACGACATTCCAATTGGAATTGACAAAGCTCTGGAGAAAATGCAG
CGGGAAGAACAATGTATTTTATATCTTGGACCAAGATATGGTTTTGGAGAGGCAGGGAAGCCTAAATTTG
GCATTGAACCTAATGCTGAGCTTATATATGAAGTTACACTTAAGAGCTTCGAAAAGGCCAAAGAATCCTG
GGAGATGGATACCAAAGAAAAATTGGAGCAGGCTGCCATTGTCAAAGAGAAGGGAACCGTATACTTCAAG
GGAGGCAAATACATGCAGGCGGTGATTCAGTATGGGAAGATAGTGTCCTGGTTAGAGATGGAATATGGTT
TATCAGAAAAGGAATCGAAAGCTTCTGAATCATTTCTCCTTGCTGCCTTTCTGAACCTGGCCATGTGCTA
CCTGAAGCTTAGAGAATACACCAAAGCTGTTGAATGCTGTGACAAGGCCCTTGGACTGGACAGTGCCAAT
GAGAAAGGCTTGTATAGGAGGGGTGAAGCCCAGCTGCTCATGAACGAGTTTGAGTCAGCCAAGGGTGACT
TTGAGAAAGTGCTGGAAGTAAACCCCCAGAATAAGGCTGCAAGACTGCAGATCTCCATGTGCCAGAAAAA
GGCCAAGGAGCACAACGAGCGGGACCGCAGGATATACGCCAACATGTTCAAGAAGTTTGCAGAGCAGGAT
GCCAAGGAAGAGGCCAATAAAGCAATGGGCAAGAAGACTTCAGAAGGGGTCACTAATGAAAAAGGAACAG
ACAGTCAAGCAATGGAAGAAGAGAAACCTGAGGGCCACGTATGA
```

Growth of GenBank

(1982 - 2005)

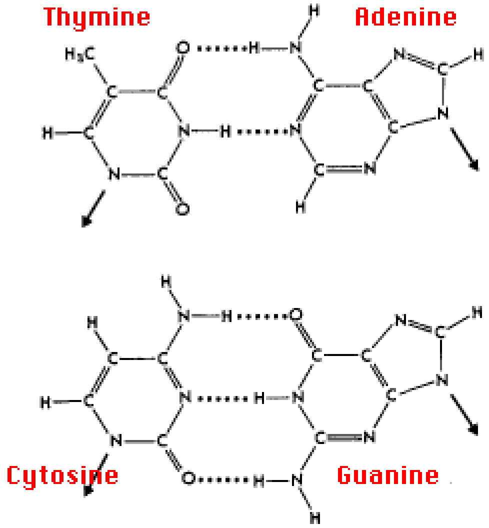




$C_4H_4N_3O$

$C_4H_3N_2O_2$

$C_5H_5N_2O_2$



Complementary bases, paired via **2** (T-A) or **3** (C-G) hydrogen bonds

PDB

DNA sequencing

- Cut long DNA strands in short fragments using Restriction Enzymes

- Expand short DNA fragments (e.g. by PRC) Polymerase Chain Reaction

- Two strategies

  1) Maxam-Gilbert method

  - mark (radioactively by ^{32}P) at the 5’ end DNA fragments
  - cut out chemically the 3’ end from each basis onward
  - keep only the marked fragments

  2) Sanger method

  - single strand fragments are let to polymerize in 4 kinds of different environments, i.e. in the presence of A, T, C, G, plus either A’, or T’, or C’, or G’, respectively
  - when either A’, or T’, or C’, or G’ is incorporated copying is blocked
  - primer that let the polymerization start or dideoxynucleotides A’, T’, C’, G’ are marked (e.g. with a fluorescent dye)

- Electrophoresis and radiography

From marked positions one gets all possible locations of each of the four bases along the fragment from which its sequence is reconstructed

The Maxam-Gilbert method

marked sequence ⇒ 5’-^{32}P-GCTACGTA-3’

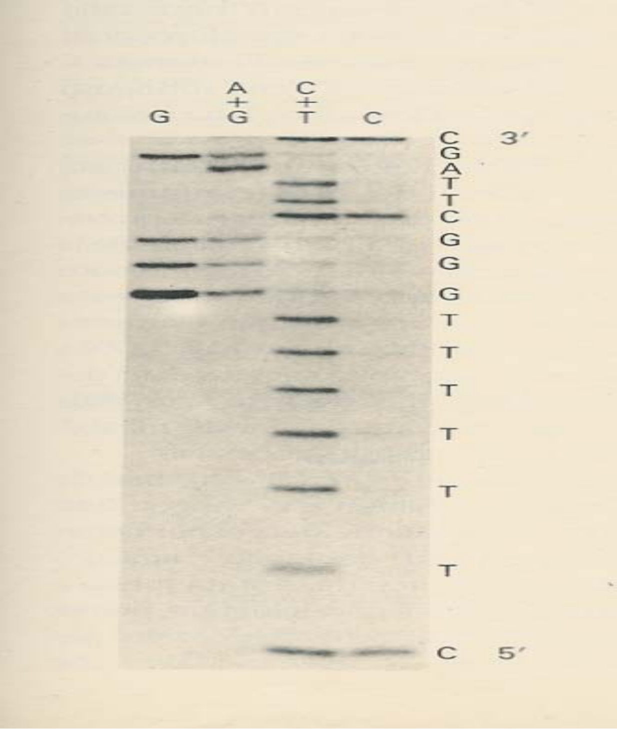
Resulting radioactive fragments

Cleavage @ A ^{32}P-GCT
^{32}P-GCTACGT

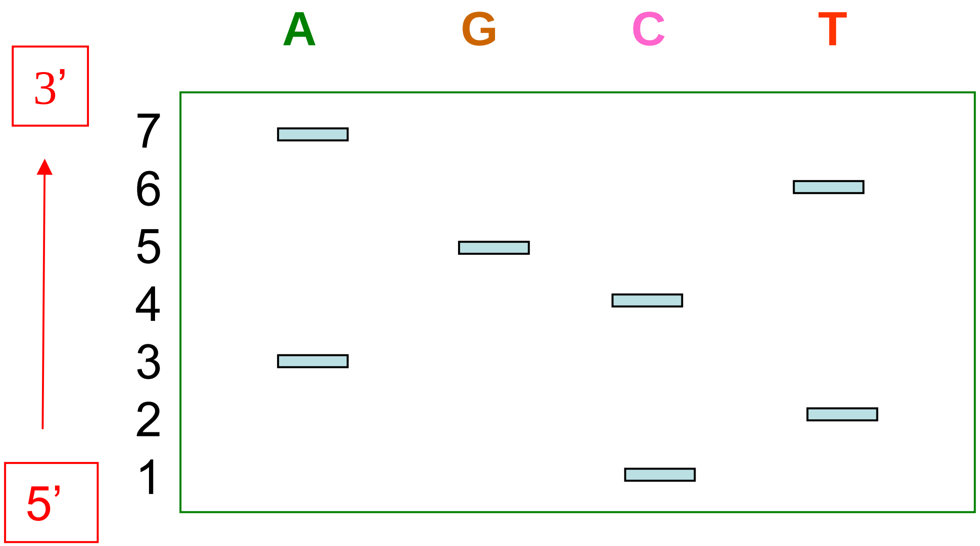
Cleavage @ G ^{32}P-GCTAC

Cleavage @ C ^{32}P-G
^{32}P-GCTA

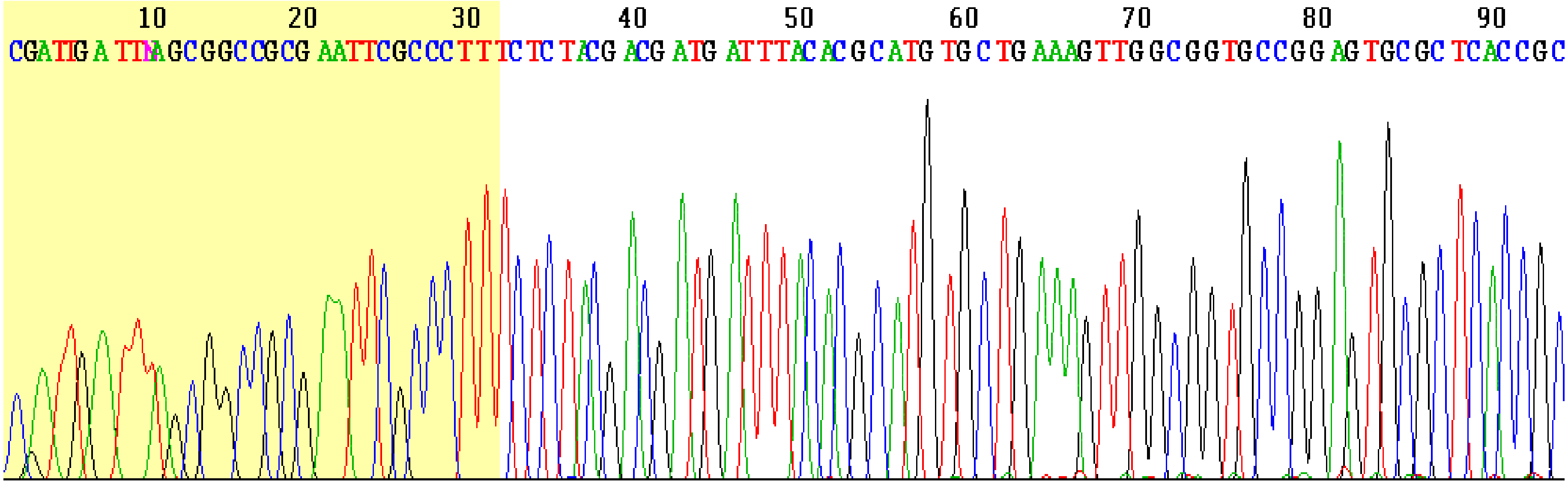
Cleavage @ T ^{32}P-GC
^{32}P-GCTACG

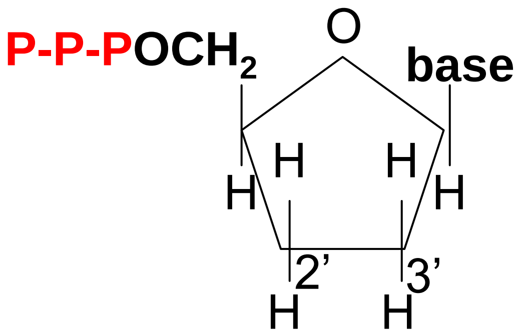


Radiography



The Sanger method

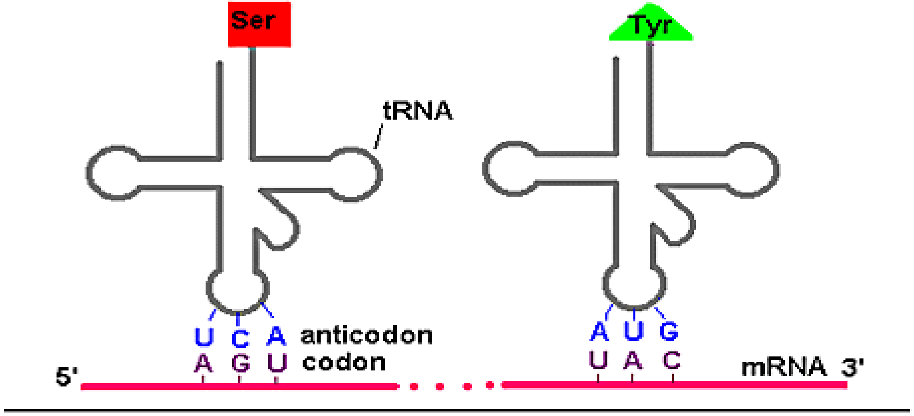




dideoxynucleotides

A' → ddATP
G' → ddGTP
C' → ddCTP
T' → ddTTP

chain-terminating nucleotides, lacking both 3'-OH groups required for the formation of a phosphodiester bond between two nucleotides during DNA strand elongation



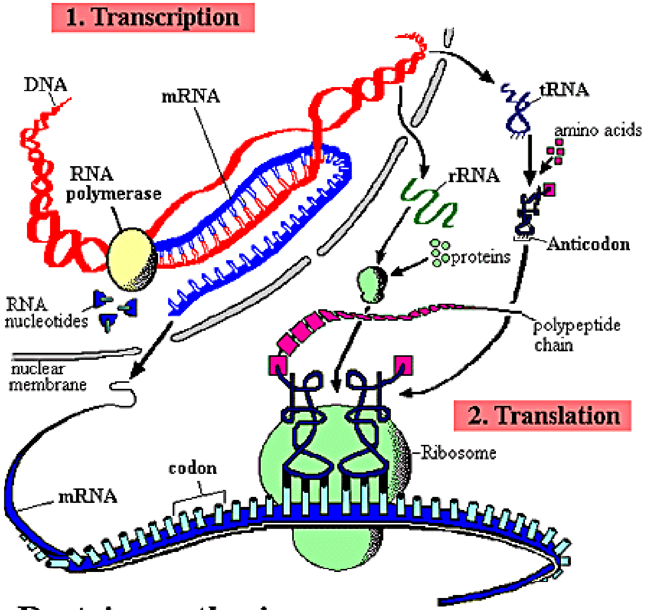
2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The Genetic Code

20 amino acids plus a stop (beginning is Methionine)

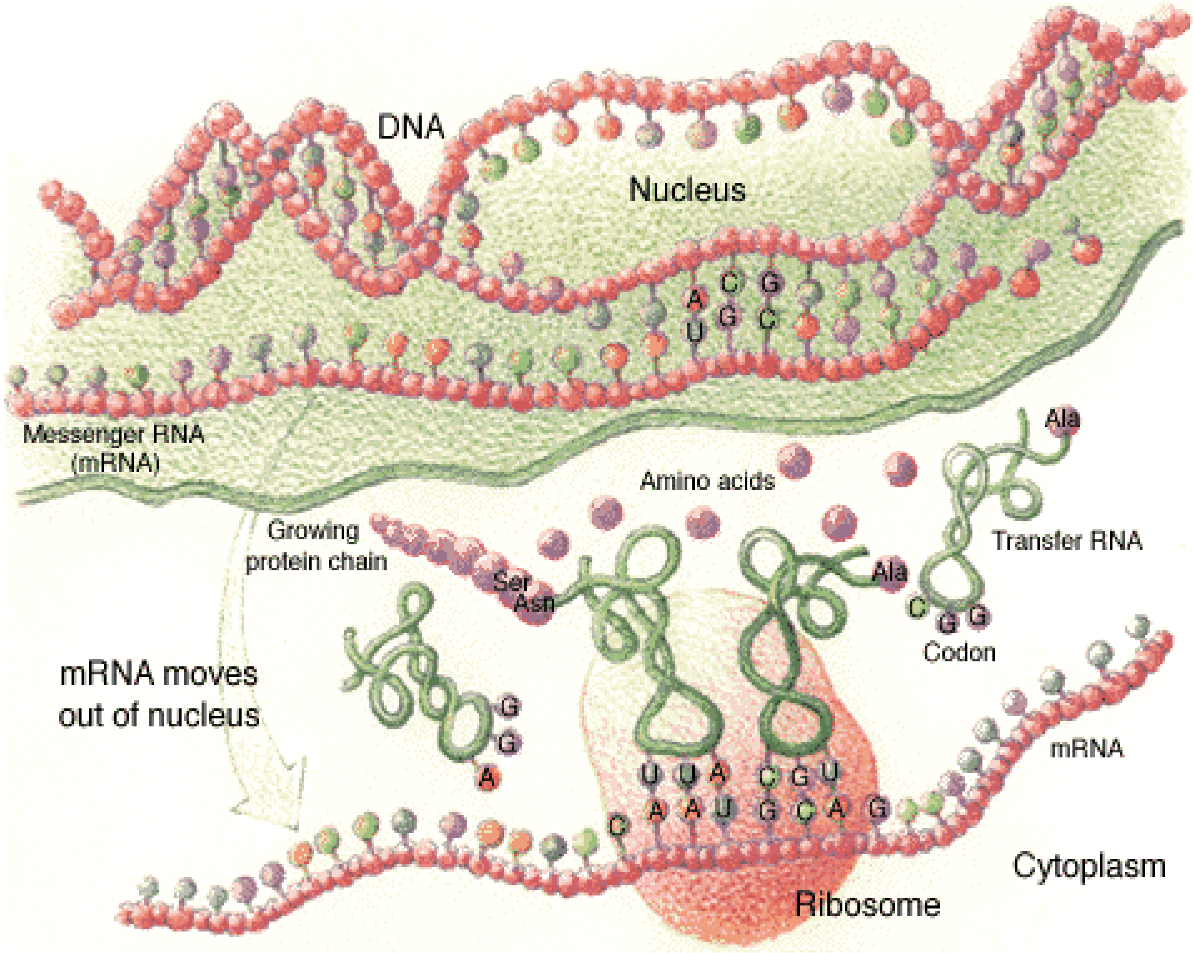
What's the reason for these numbers? Perhaps QC? Patel



Protein synthesis

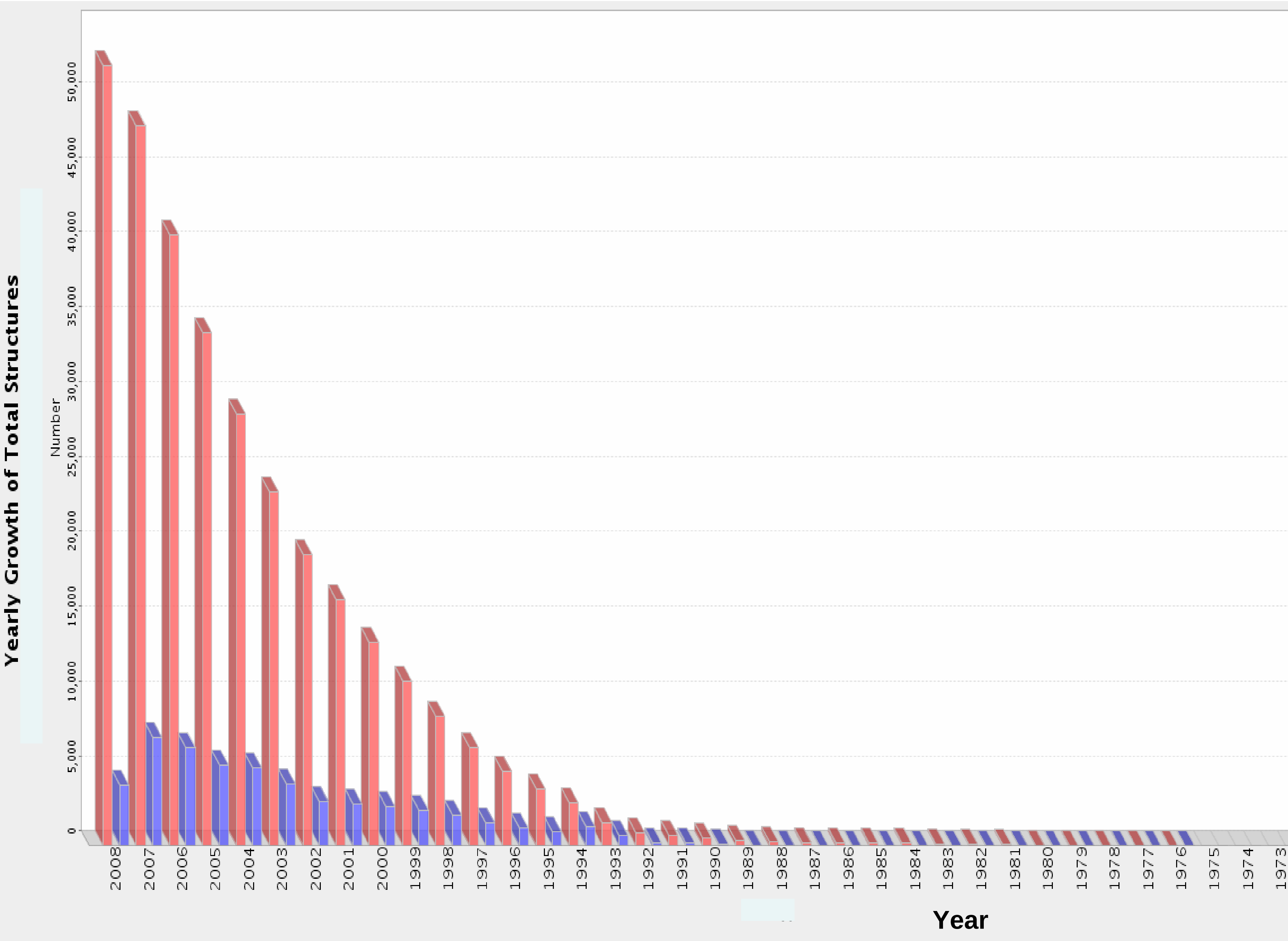
Three bases make a codon

4 bases are A T G C
Adenosine Thymine Guanine Cytosine
Uracyle


DNA
Nucleus
A
U
C
G
G
C
Messenger RNA
(mRNA)
Amino acids
Ala
Transfer RNA
Growing
protein chain
Ser
Asn
Ala
C
G
G
Codon
mRNA moves
out of nucleus
G
G
A
U
U
A
C
G
U
C
A
A
U
G
C
A
G
mRNA
Cytoplasm
Ribosome

PDB Current Holdings Breakdown

		Molecule Type				
		Proteins	Nucleic Acids	Protein/NA Complexes	Other	**Total**
Exp. Method	X-ray	41431	1058	1902	24	**44415**
	NMR	6447	814	138	7	**7406**
	Electron Microscopy	125	11	47	0	**183**
	Other	89	4	4	2	**99**
	Total	**48092**	**1887**	**2091**	**33**	**52103**


Yearly Growth of Total Structures
Number
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
50,000
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
1990
1989
1988
1987
1986
1985
1984
1983
1982
1981
1980
1979
1978
1977
1976
1975
1974
1973
Year

Data

- Data types
  - Amino acid sequences (proteins)
  - Genomic sequences (DNA, RNA,…)
  - 3D-structures of macromolecules
  - Biochemical/Physiological
  - Medical/Epidemiological
  - ……….

There exist about 60 DataBases, each containig trilions of bits

http://expasy.org/

- Metabolic pathways
- siRNA/RNAi
- Peptide antigens
- Protein interactions
- Kinase-Phosphate
- Transcription factors
- Disease Genes
- Protein database
- ………

- Data handling
  - Collecting/Recording
  - Releasing/Validating/Curing
  - Updating/Maintaining

  IN

  - Mining
  - Analyzing/Organizing

  OUT

- Data availability
  - Need easy, standardized, free access to DataBanks
  - Patent laws and regulations

Most important aspect in the production of antibodies or drugs is the design of peptide-antigens. A peptide-antigen is a small segment (15-18 amino acids) of the protein sequence of interest. These peptide-antigens can be used for immunization in order to produce antibodies against the protein or they can be used as a basis for small-molecule/drug targeting.

The Peptide-Antigen database http://www.proteinlounge.com/ contains antigenic peptide targets against all known protein sequences throughout a variety of organisms.

An example

Homo sapiens	24225 Genes with 119846 Peptide Sequences
Mus musculus	39995 Genes with 212168 Peptide Sequences
Rattus norvegicus	19245 Genes with 95044 Peptide Sequences
Bos taurus	1169 Genes with 5809 Peptide Sequences
Danio rerio	1139 Genes with 5651 Peptide Sequences
Drosophila melanogaster	17673 Genes with 87744 Peptide Sequences
Anopheles gambiae	15297 Genes with 75342 Peptide Sequences
Caenorhabditis elegans	26843 Genes with 216976 Peptide Targets
Arabidopsis thaliana	15276 Genes with 75626 Peptide Sequences
Saccharomyces cerevisiae	5655 Genes with 28138 Peptide Sequences

New words for new concepts and needs, like

Proteomics
Genomics
Metabolomics
Reactomics
…

come into play.

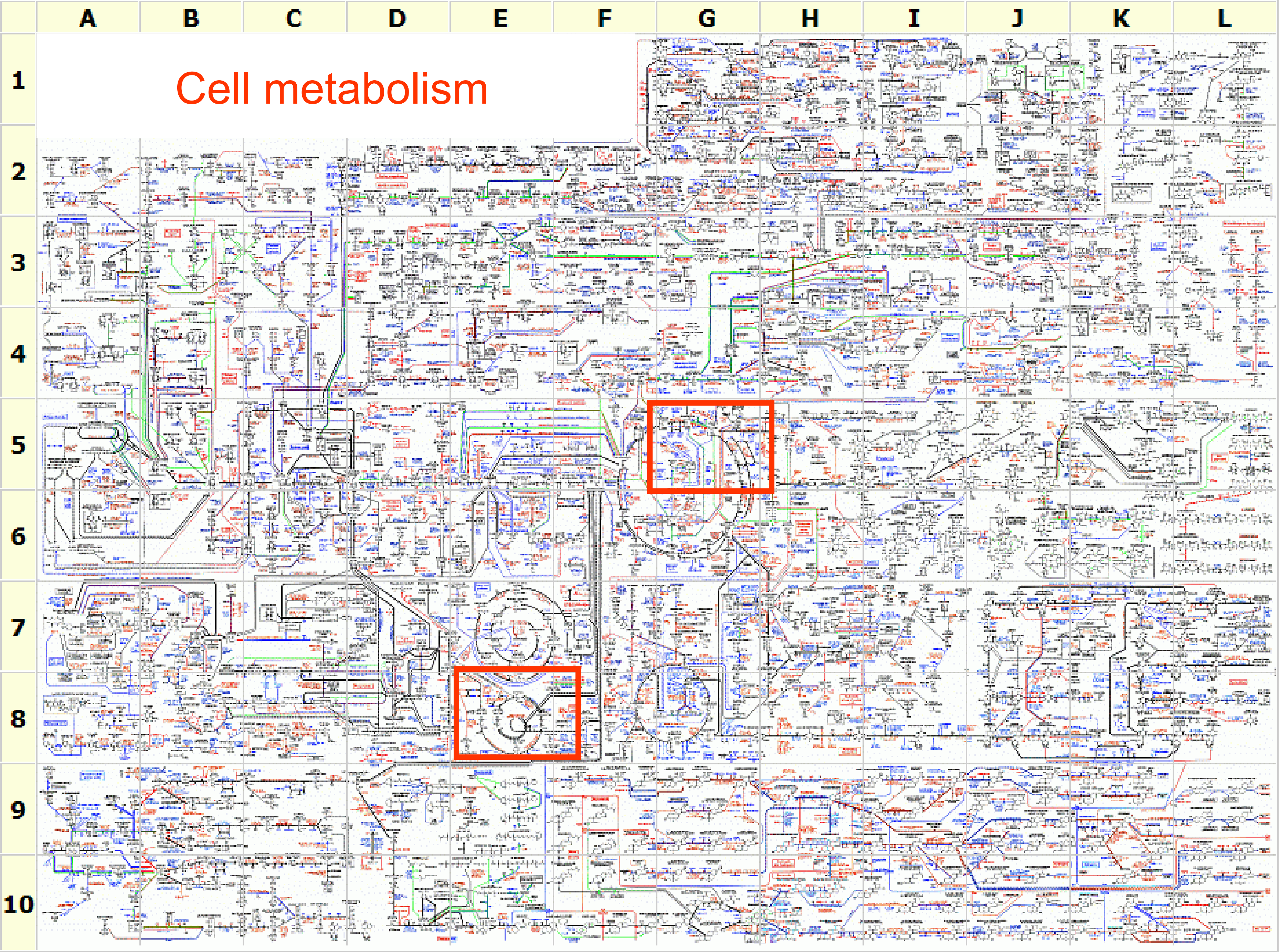
The suffix “omics” is alluding to the fact that are not just the single objects of each class (proteins, genoma, metabolic reactions,…) that matter, but their relations and interconnections

Cell reactome

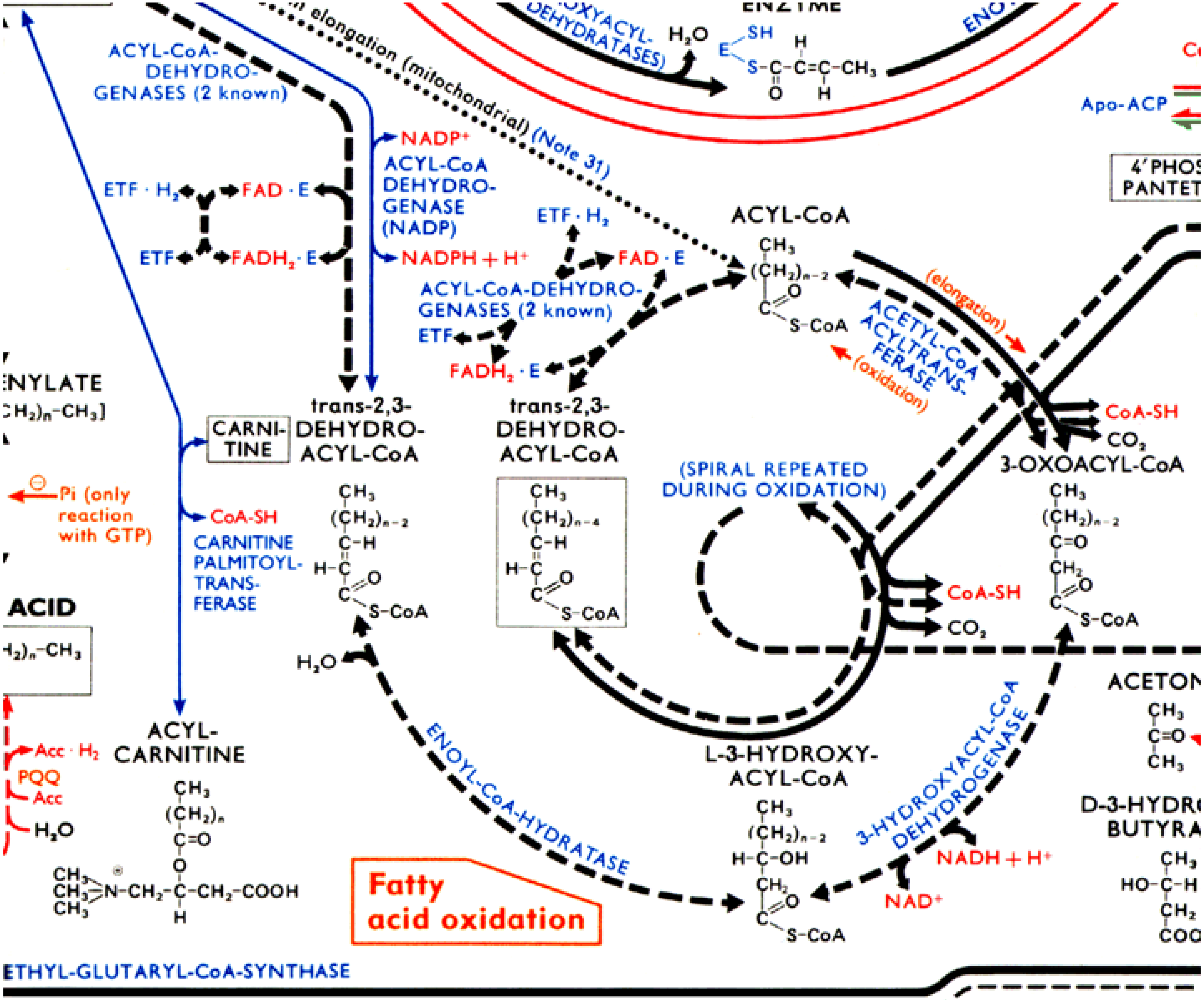
http://www.reactome.org/

Apoptosis	Biological oxidations	Botulinum neurotoxicity	Cell Cycle Checkpoints
Cell Cycle, Mitotic	DNA Repair	DNA Replication	Electron Transport Chain
Gap junction trafficking and regulation	Gene Expression	HIV Infection	Hemostasis
Influenza Infection	Integration of energy metabolism	Lipid and lipoprotein metabolism	Membrane Trafficking
Metabolism of amino acids	Metabolism of carbohydrates	Metabolism of nitric oxide	Metabolism of non-coding RNA
Metabolism of vitamins and cofactors	Nucleotide metabolism	Porphyrin metabolism	Pyruvate metabolism and TCA cycle
Post-translational protein modification	Regulation of beta-cell development	Regulatory RNA pathways	Signaling by BMP
Signaling by EGFR	Signaling by FGFR	Signaling in Immune system	Signaling by Insulin receptor
Signalling by NGF	Signaling by Notch	Signaling by Rho GTPases	Signaling by TGF beta
Signaling by VEGF	Signaling by Wnt	Synaptic Transmission	Telomere Maintenance
Transcription	Translation	mRNA Processing	

Cell metabolism

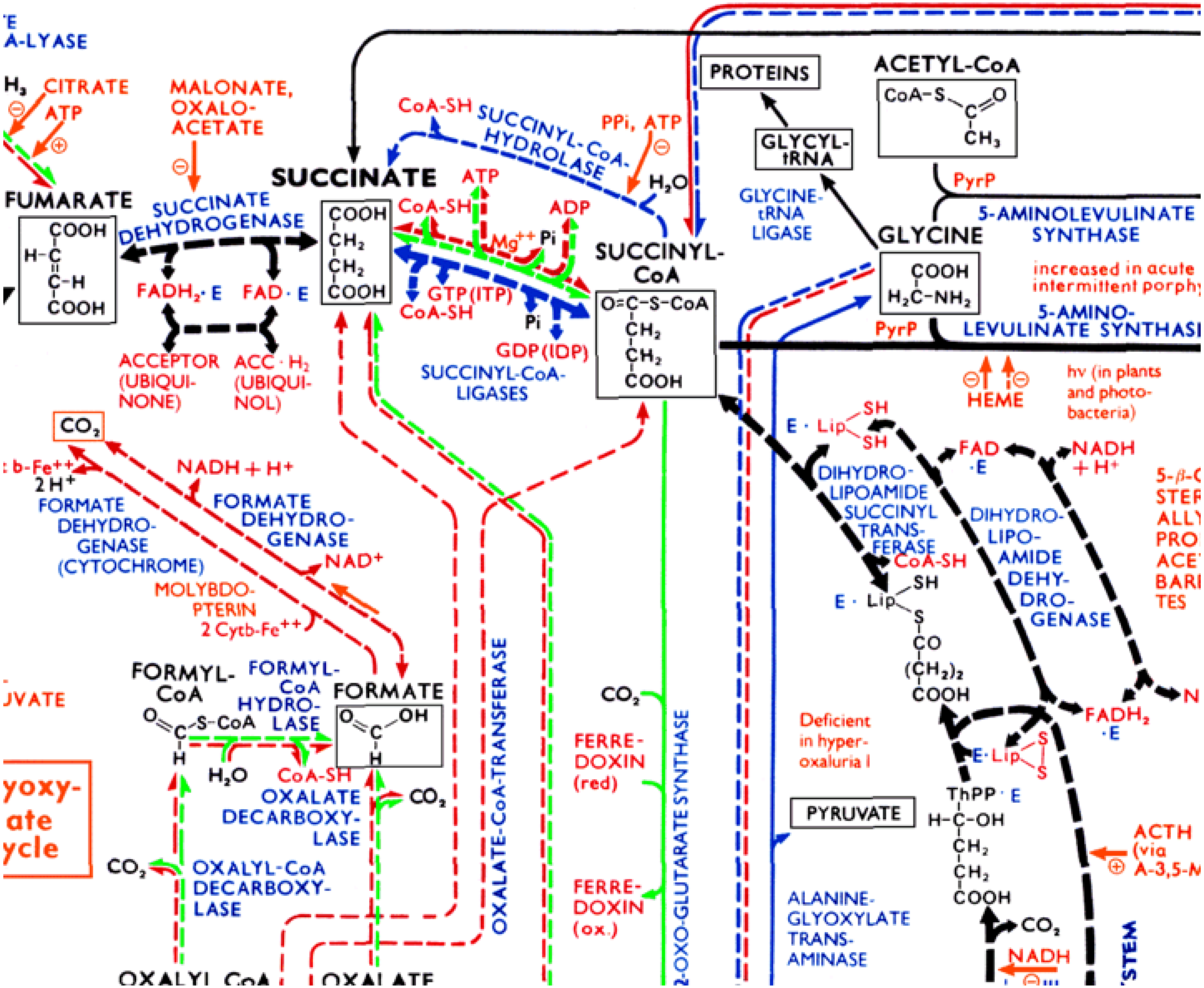


E8



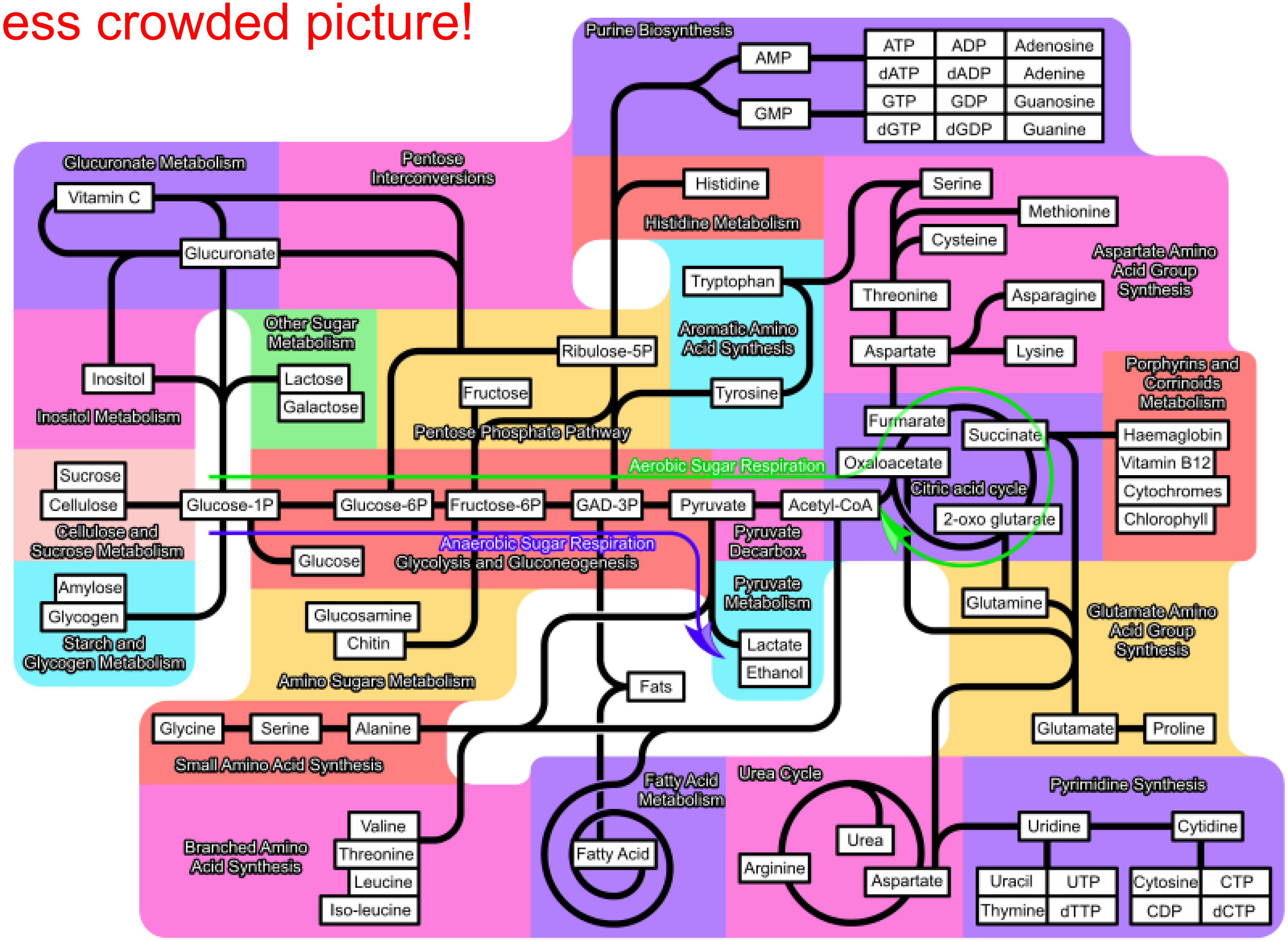
elongation (mitochondrial) (Note 31)
DEHYDRATASES)
H₂O
E
SH
S-C-C=C-CH₃
ENZYME
Apo-ACP
4′PHOS
PANTET
ACYL-CoA-DEHYDROGENASES (2 known)
ETF · H₂
FAD · E
ETF
FADH₂ · E
NADP⁺
ACYL-CoA DEHYDROGENASE (NADP)
NADPH + H⁺
ETF · H₂
FAD · E
ACYL-CoA-DEHYDROGENASES (2 known)
ETF
FADH₂ · E
ACYL-CoA
CH₃
(CH₂)ₙ₋₂
C=O
S-CoA
(elongation)
ACETYL-CoA ACYLTRANSFERASE
(oxidation)
CoA-SH
CO₂
3-OXOACYL-CoA
ENYLATE
H₂)ₙ-CH₃]
CARNITINE
trans-2,3-DEHYDRO-ACYL-CoA
trans-2,3-DEHYDRO-ACYL-CoA
(ₙ₋₄)
C-H
H-C
(SPIRAL REPEATED DURING OXIDATION)
CoA-SH
CO₂
Pi (only reaction with GTP)
CoA-SH
CARNITINE PALMITOYL-TRANSFERASE
ACID
H₂)ₙ-CH₃
H₂O
ENOYL-CoA-HYDRATASE
ACETON
CH₃
C=O
CH₃
L-3-HYDROXY-ACYL-CoA
H-C-OH
CH₂
3-HYDROXYACYL-CoA DEHYDROGENASE
NADH + H⁺
NAD⁺
D-3-HYDR
BUTYRA
HO-C-H
COO
Acc · H₂
PQQ
Acc
H₂O
ACYL-CARNITINE
CH₃
(CH₂)ₙ
C=O
O
CH₃ CH₃ CH₃ N-CH₂-C-CH₂-COOH
H
Fatty acid oxidation
ETHYL-GLUTARYL-CoA-SYNTHASE

G5



A-LYASE
CITRATE
ATP
MALONATE, OXALO-ACETATE
FUMARATE
SUCCINATE DEHYDROGENASE
SUCCINATE
FADH$_2$·E
FAD·E
ACCEPTOR (UBIQUINONE)
ACC·H$_2$ (UBIQUINOL)
CoA-SH
SUCCINYL-CoA-HYDROLASE
PPi, ATP
H_2O
ATP
CoA-SH
ADP
Mg^{++} Pi
GTP(ITP)
CoA-SH
Pi
GDP(IDP)
SUCCINYL-CoA-LIGASES
SUCCINYL-CoA
PROTEINS
GLYCYL-tRNA
GLYCINE-tRNA LIGASE
ACETYL-CoA
PyrP
GLYCINE
5-AMINOLEVULINATE SYNTHASE
increased in acute intermittent porphy
PyrP
5-AMINO-LEVULINATE SYNTHASE
HEME
hv (in plants and photo-bacteria)
CO$_2$
b-Fe^{++}
2 H$^+$
FORMATE DEHYDROGENASE (CYTOCHROME)
NADH + H$^+$
FORMATE DEHYDRO-GENASE
NAD$^+$
MOLYBDO-PTERIN
2 Cytb-Fe^{++}
E·Lip SH SH
DIHYDRO-LIPOAMIDE SUCCINYL TRANS-FERASE
CoA-SH
FAD·E
NADH + H$^+$
DIHYDRO-LIPO-AMIDE DEHY-DRO-GENASE
5-β-C STER ALLY PRO ACE BARI TES
FADH$_2$·E
UVATE
FORMYL-CoA
FORMYL-CoA HYDRO-LASE
FORMATE
H_2O
CoA-SH
OXALATE DECARBOXY-LASE
CO$_2$
yoxy-ate ycle
CO$_2$
OXALYL-CoA DECARBOXY-LASE
OXALATE-CoA-TRANSFERASE
CO$_2$
FERRE-DOXIN (red)
FERRE-DOXIN (ox.)
2-OXO-GLUTARATE SYNTHASE
Deficient in hyper-oxaluria I
PYRUVATE
ALANINE-GLYOXYLATE TRANS-AMINASE
E·Lip S S
ThPP·E
ACTH (via A-3,5-M
NADH
CO$_2$
STEM
OXALYL CoA
OXALATE

A less crowded picture!



METABOLIC NETWORKS

Metabolism of eukaryotic cells

~5000-6000 enzymatic reactions

~3000 metabolites

most simple:

Red blood cells

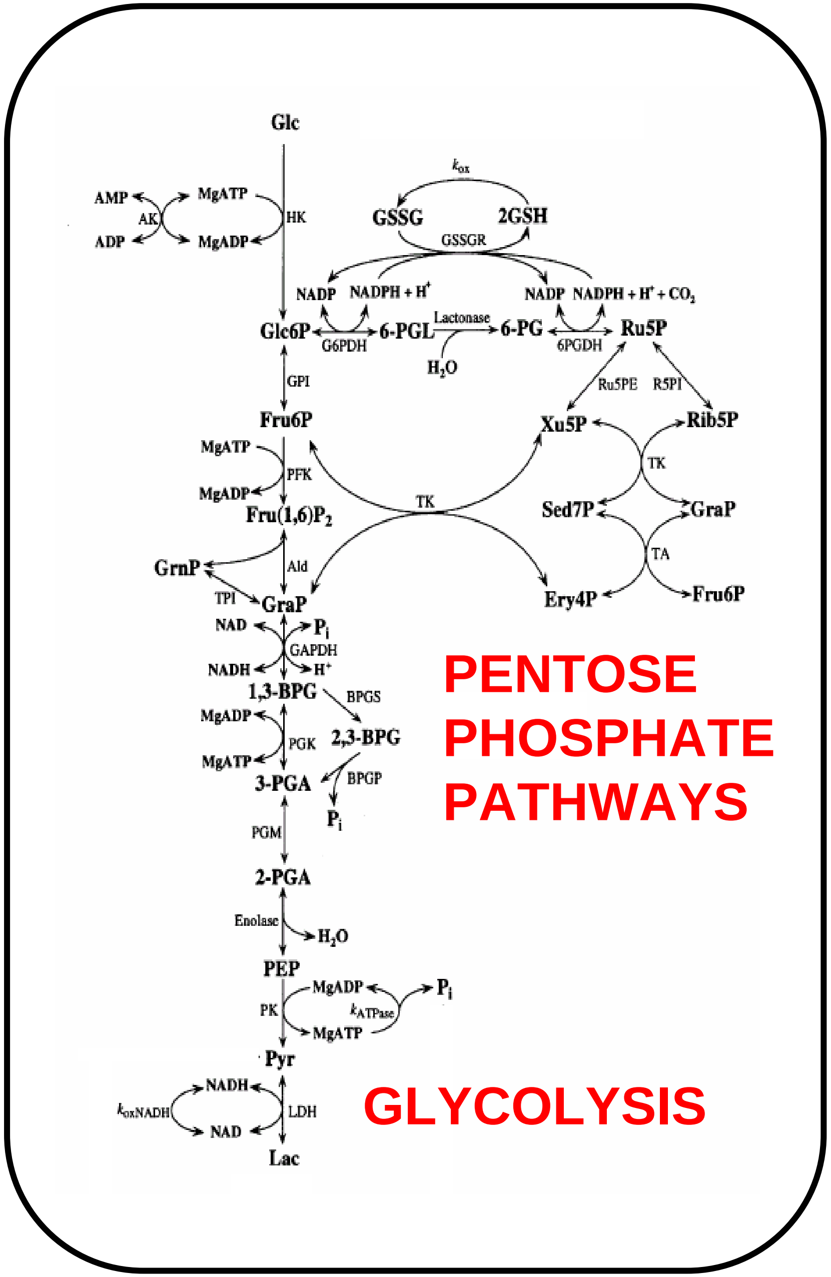
Model system for

Calculation of dynamical properties of whole pathways based on the kinetic properties of single enzymes

metabolite concentration

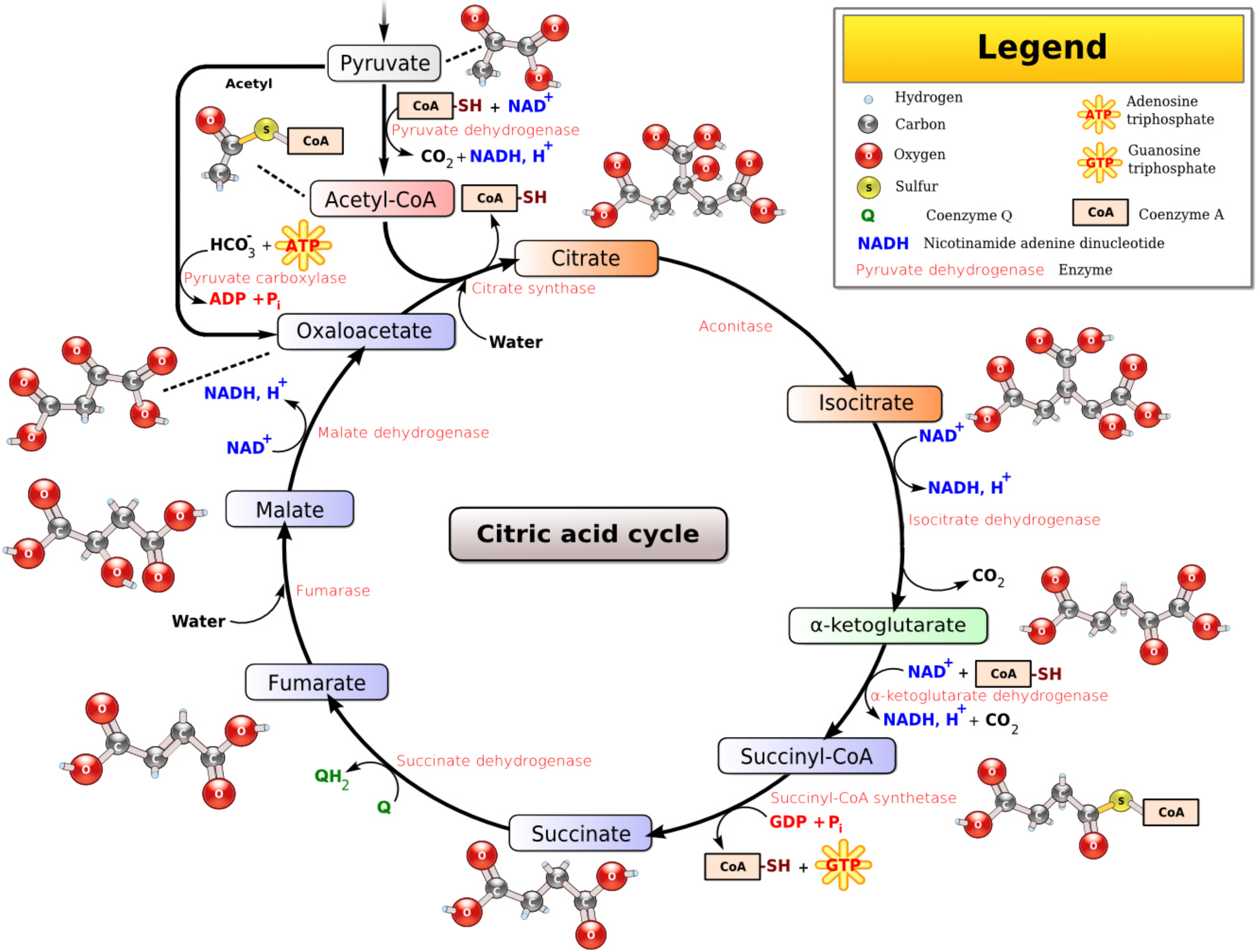
~ 1 μM: $10^8 – 10^9$ molecules/cell

for most substances

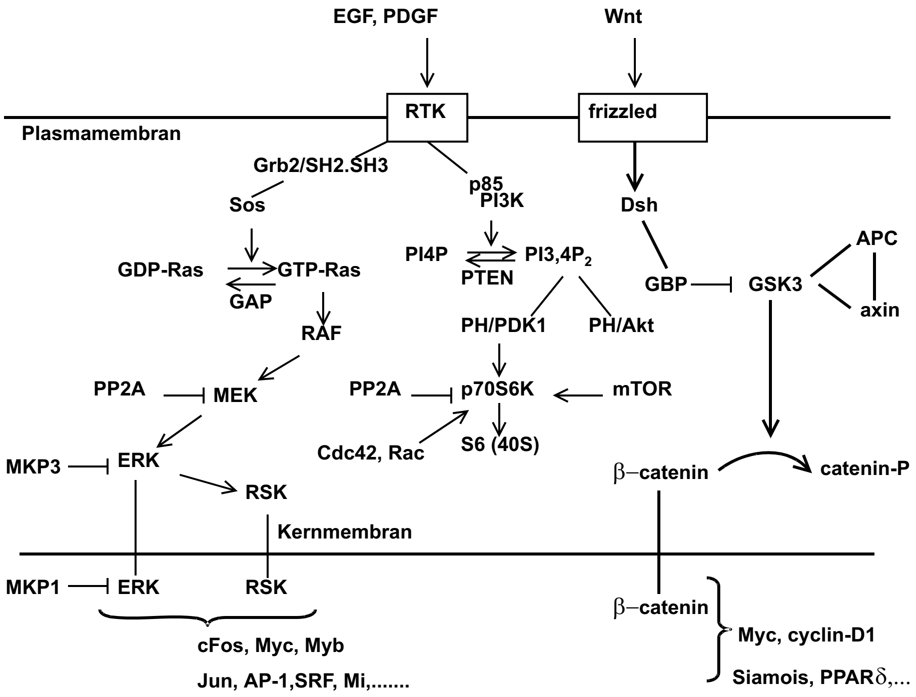


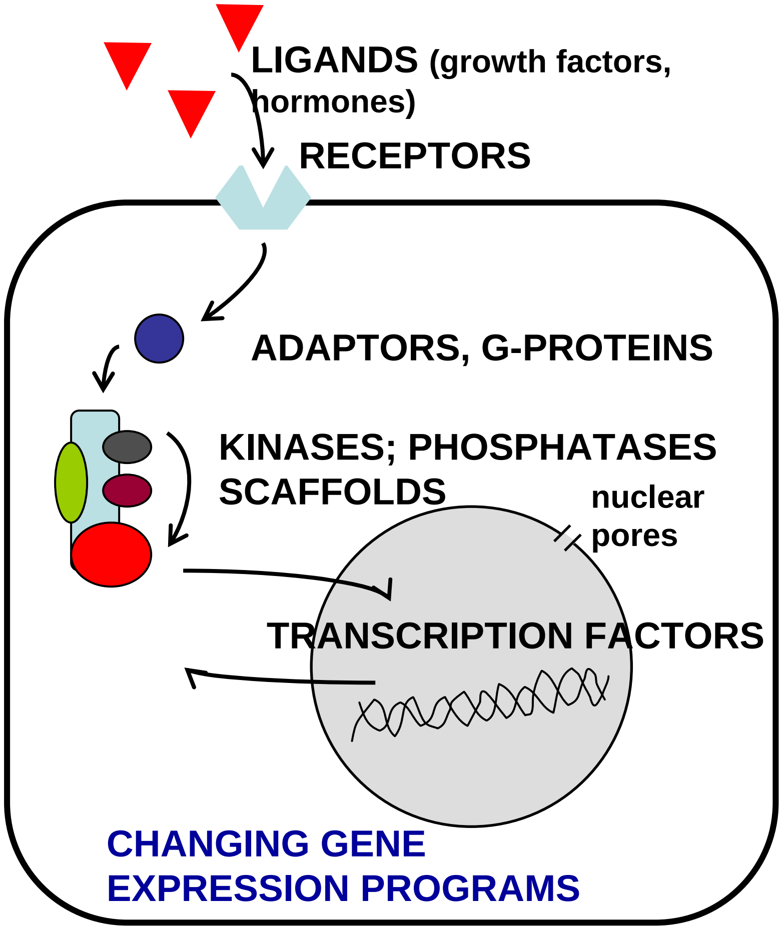
PENTOSE PHOSPHATE PATHWAYS

GLYCOLYSIS


Legend
Hydrogen
Carbon
Oxygen
Sulfur
Q Coenzyme Q
NADH Nicotinamide adenine dinucleotide
Pyruvate dehydrogenase Enzyme
ATP Adenosine triphosphate
GTP Guanosine triphosphate
CoA Coenzyme A
Pyruvate
Acetyl
CoA-SH + NAD^+
Pyruvate dehydrogenase
CO_2 + NADH, H^+
Acetyl-CoA
CoA-SH
HCO_3^- + ATP
Pyruvate carboxylase
ADP + P_i
Citrate
Citrate synthase
Water
Oxaloacetate
Aconitase
Isocitrate
NAD^+
NADH, H^+
Isocitrate dehydrogenase
CO_2
α-ketoglutarate
NAD^+ + CoA-SH
α-ketoglutarate dehydrogenase
NADH, H^+ + CO_2
Succinyl-CoA
Succinyl-CoA synthetase
GDP + P_i
CoA-SH + GTP
Succinate
Succinate dehydrogenase
QH_2
Q
Fumarate
Fumarase
Water
Malate
Malate dehydrogenase
NAD^+
NADH, H^+
Citric acid cycle
C
O
S
CoA

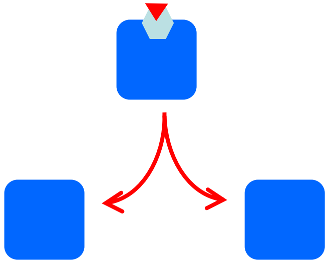
SIGNAL TRANSDUCTION





concentrations of signaling molecules: ~100 nM

10^4 – 10^5 molecules/cell

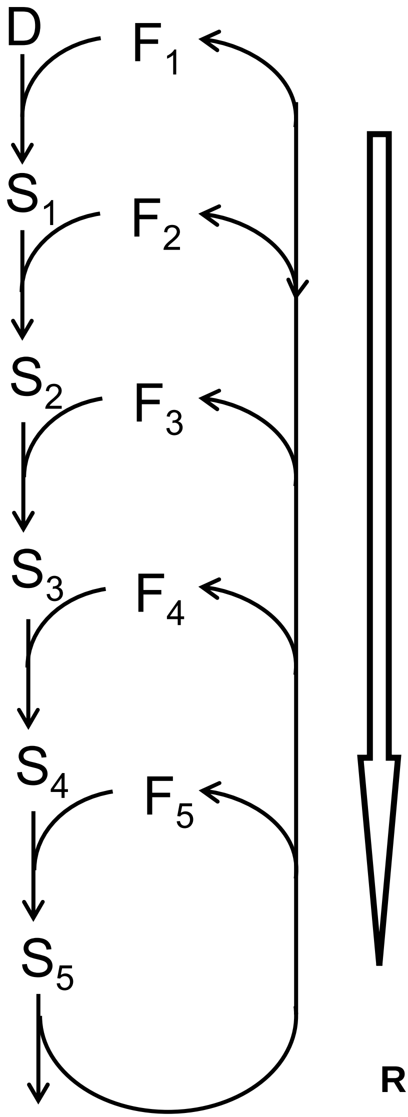


CELL PROLIFERATION

CELL MOVEMENT

DNA-REPAIR

Nucleotide Excision Repair (NER)
Global Genom Repair Pathway

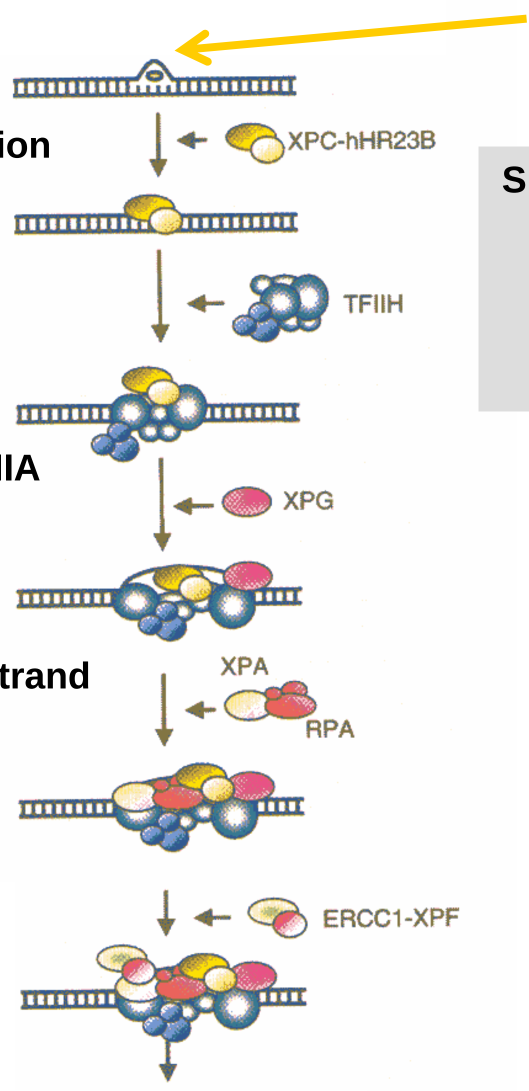


UV-light

damage recognition by XPC

DNA helix unwinding by TFIIA

Cutting of DNA-strand and removing of damaged region



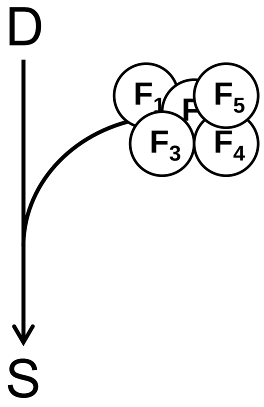
Repair: Synthesis of new DNA strand, ~ 10 nucleotides long

SEQUENTIAL MECHANISM

or

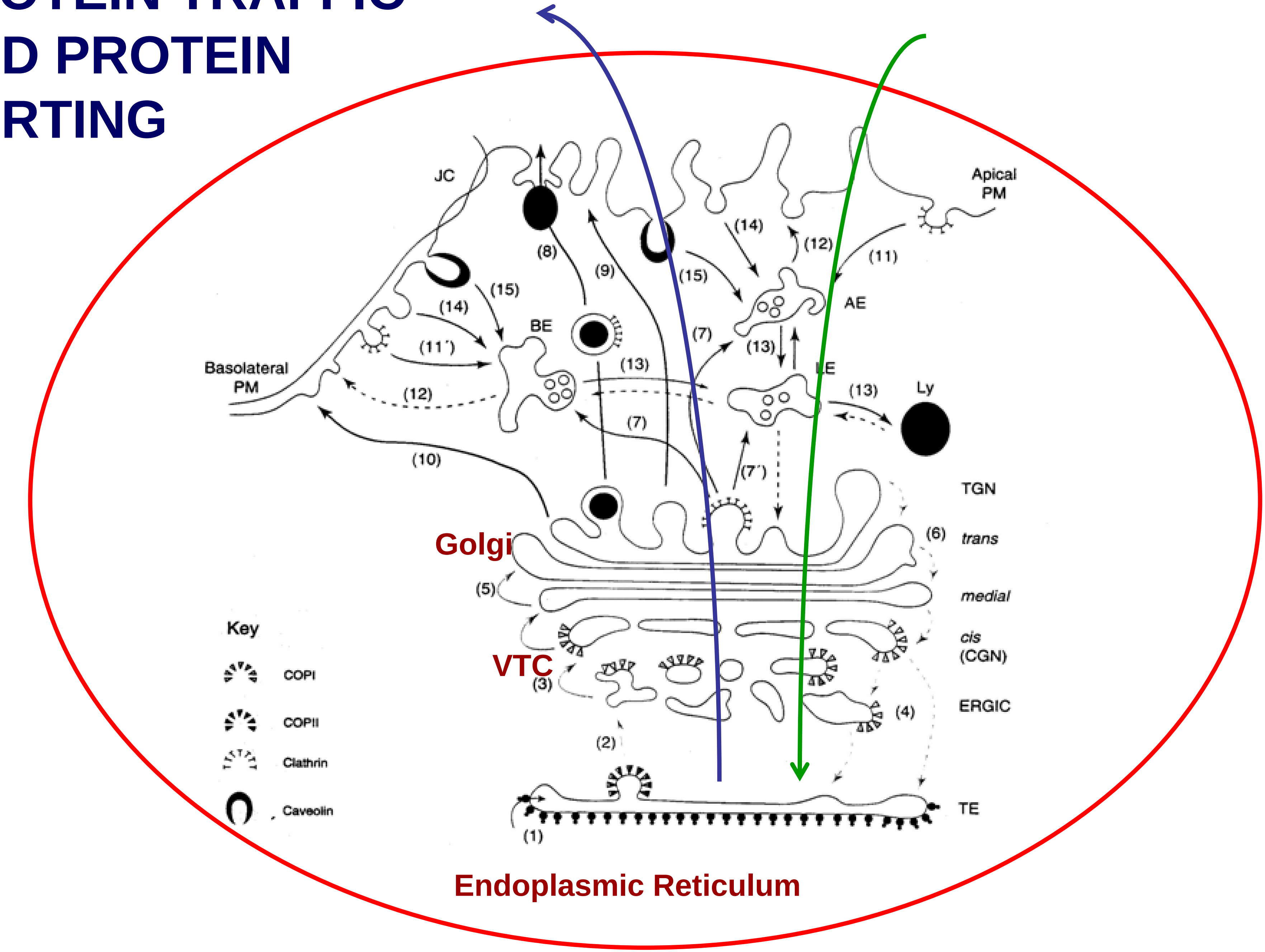
PREASSEMBLED

HOLOCOMPLEX



Mutations in NER-proteins: photosensitivity and sunlight-induced skin cancer

PROTEIN TRAFFIC AND PROTEIN SORTING



Dynamics of Cell Reactions

Network of very many interconnected
sub-networks of related biochemical reactions

Barabasi

- Non-linear diffusive (of the heat type) PDE’s
- Small number of some of the involved molecular species 10^2-10^5/cell
  - large number-fluctuations
  - competition with thermal fluctuations
- Events are discrete with a certain degree of randomness
- Multiplicity of time scales

Gillespie

Realistic (?) Single Cell Simulation

Even for the smallest living organism, Mycoplasma Genitalium

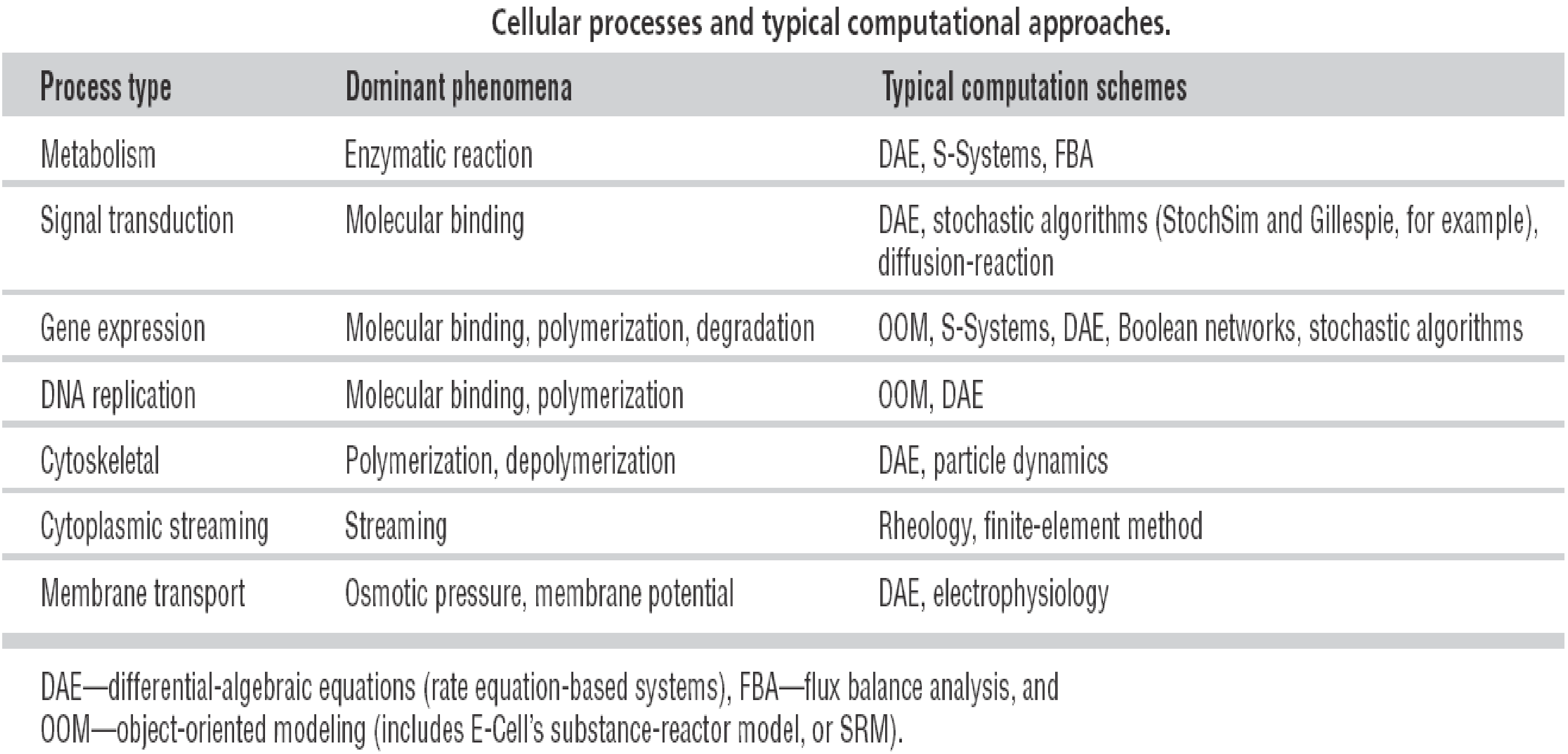
100 genes
500 proteins
100 regulatory elements
10 cellular compartments

E-Cell

http://www.e-cell.org/

Takahashi
Yugi
Hashimoto
Yamada
Pickett
Tomita

The E-Cell Simulation Environment is an object-oriented software suite for modelling, simulation, and analysis of large scale complex systems such as biological cells.

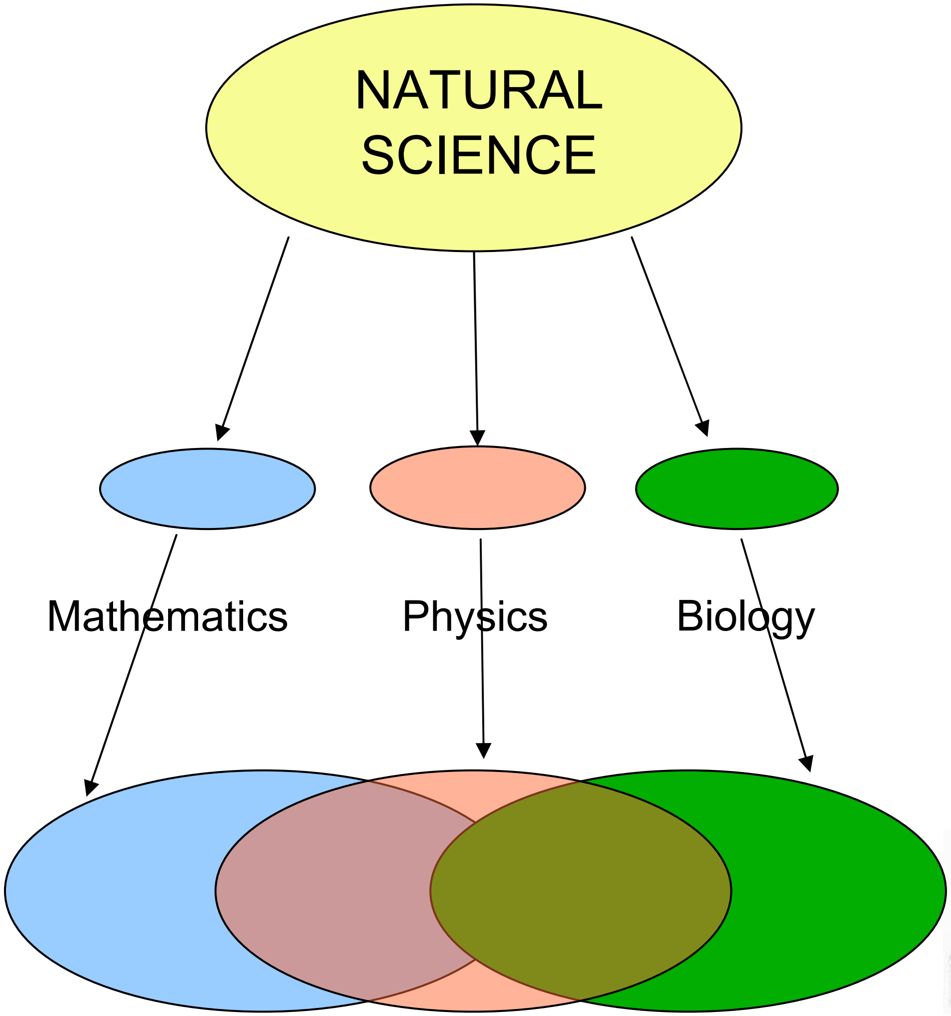


Cellular processes and typical computational approaches.

Process type	Dominant phenomena	Typical computation schemes
Metabolism	Enzymatic reaction	DAE, S-Systems, FBA
Signal transduction	Molecular binding	DAE, stochastic algorithms (StochSim and Gillespie, for example), diffusion-reaction
Gene expression	Molecular binding, polymerization, degradation	OOM, S-Systems, DAE, Boolean networks, stochastic algorithms
DNA replication	Molecular binding, polymerization	OOM, DAE
Cytoskeletal	Polymerization, depolymerization	DAE, particle dynamics
Cytoplasmic streaming	Streaming	Rheology, finite-element method
Membrane transport	Osmotic pressure, membrane potential	DAE, electrophysiology

DAE—differential-algebraic equations (rate equation-based systems), FBA—flux balance analysis, and OOM—object-oriented modeling (includes E-Cell's substance-reactor model, or SRM).

Can we hope to attack such fantastically complicated problems?
Probably yes,
looking back at the development of Natural Science



FROM PROGRESSIVE SEPARATION

TO OVERLAPPING AREAS OF INTEREST

- Similar Mathematical Description and Algorithms

→ Cross-fertilization among nearby Research Fields

Statistical Physics Quantum Field Theory → Structure of Macro-molecules Econo-physics

Stochastic Methods

Meteorology Fluido-dynamics → Metabolic networks Turbulence Chaos

PDE & Stability Analysis

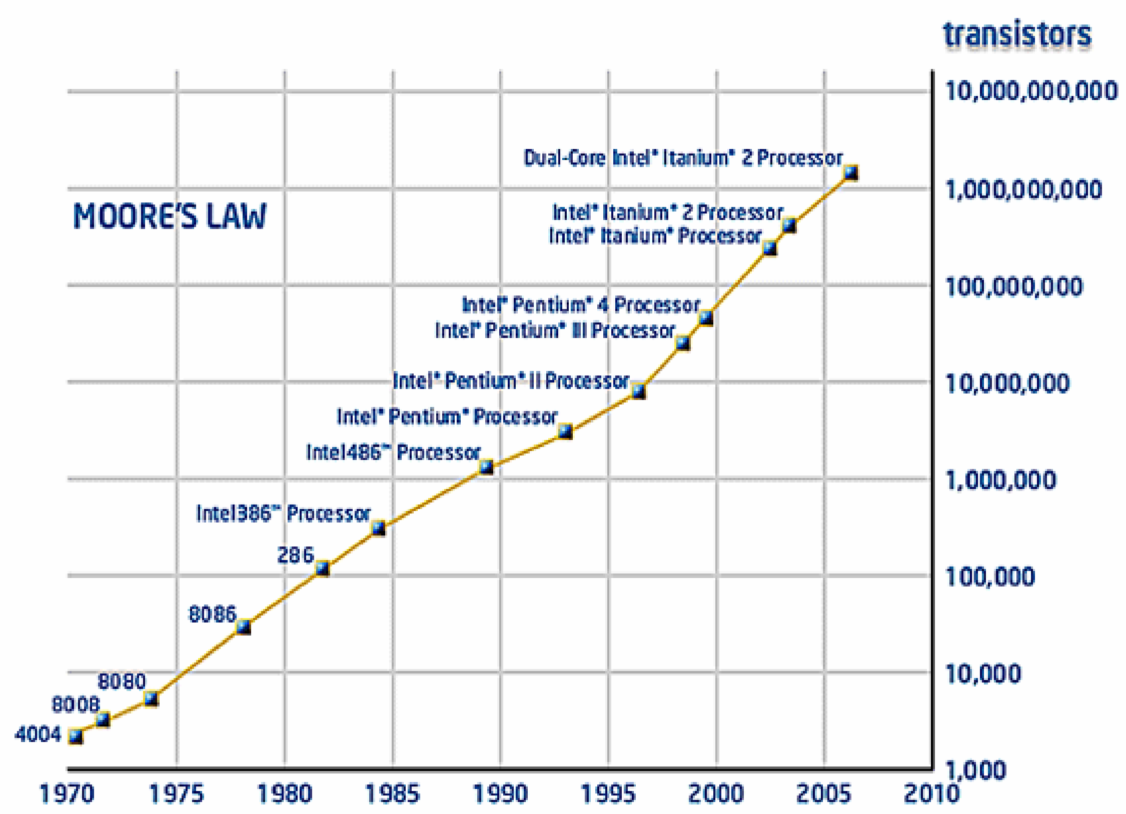
- Dealt with by Numerical Tools

→ Advances in Computer Developments

Numerical Simulations ↔ Dedicated Computers

New Architectures → Parallel Platforms
PC-Clusters
GRID

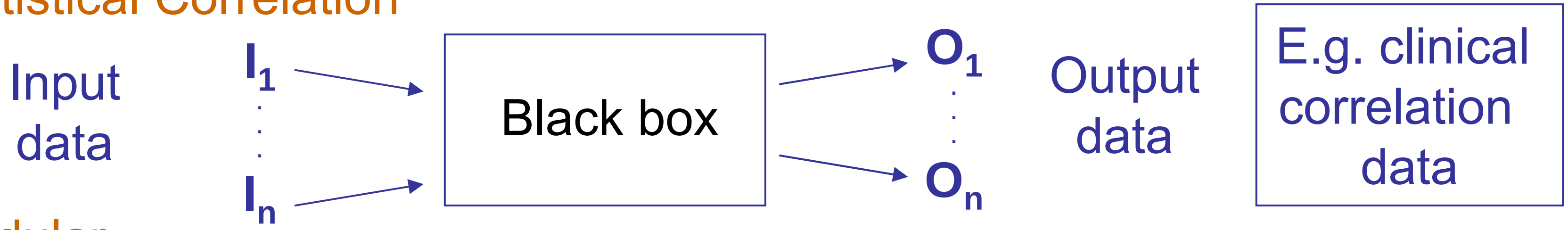
Exponential Increase of → CPU Time
Memory
Storing Capacity



Wide Spectrum of Applications → Lattice QCD
Computational Astrophysics
Weather Forecasting
Genome Project
Computational Biology

Models and modellization Strategies

- Statistical Correlation



Input data

I_1 . . . I_n

Black box

O_1 . . . O_n

Output data

E.g. clinical correlation data

- Modular

Isolating functional modules, well separated

in time: protein folding vs cell duplication
in space: ribosomal protein synthesis vs cell translocation
chemically: metabolic pathways vs DNA transcription
logically: neuronal network vs electrical transmission along the axon

- Comprehensive (holistic)

Full simulations of a living cell

Degrees of freedom

Positions and velocities
Concentrations / reaction constants
Order parameters
Transmembrane potential / ionic current
Physiological / epidemiological data
…………

Mathematical Tools

● Differential equations			
● Ordinary			Metabolic processes
● Partial			Regulatory processes
● Statistical-Mechanics-Inspired Algorithms			
		HMC	Fluids
● Stochastic	Monte Carlo	Simulated Annealing	Proteins/SG
		Multicanonical	Folding/Phase trans.
	Langevin		Cell growth
● Classical Mechanics	Molecular Dynamics	QM/MM	Protein dynamics Cell membrane
● Quantum Mechanics	DFT Car-Parrinello	*bona fide* QM	Local recognition Cell membrane
● Non-equilibrium Statistical Mechanics?			
Einstein, Onsager, Touschek Gallavotti, Jona-Lasinio			Open, almost stationary systems

III. What we would like to know and/or to do

III. What we would like to know and/or to do

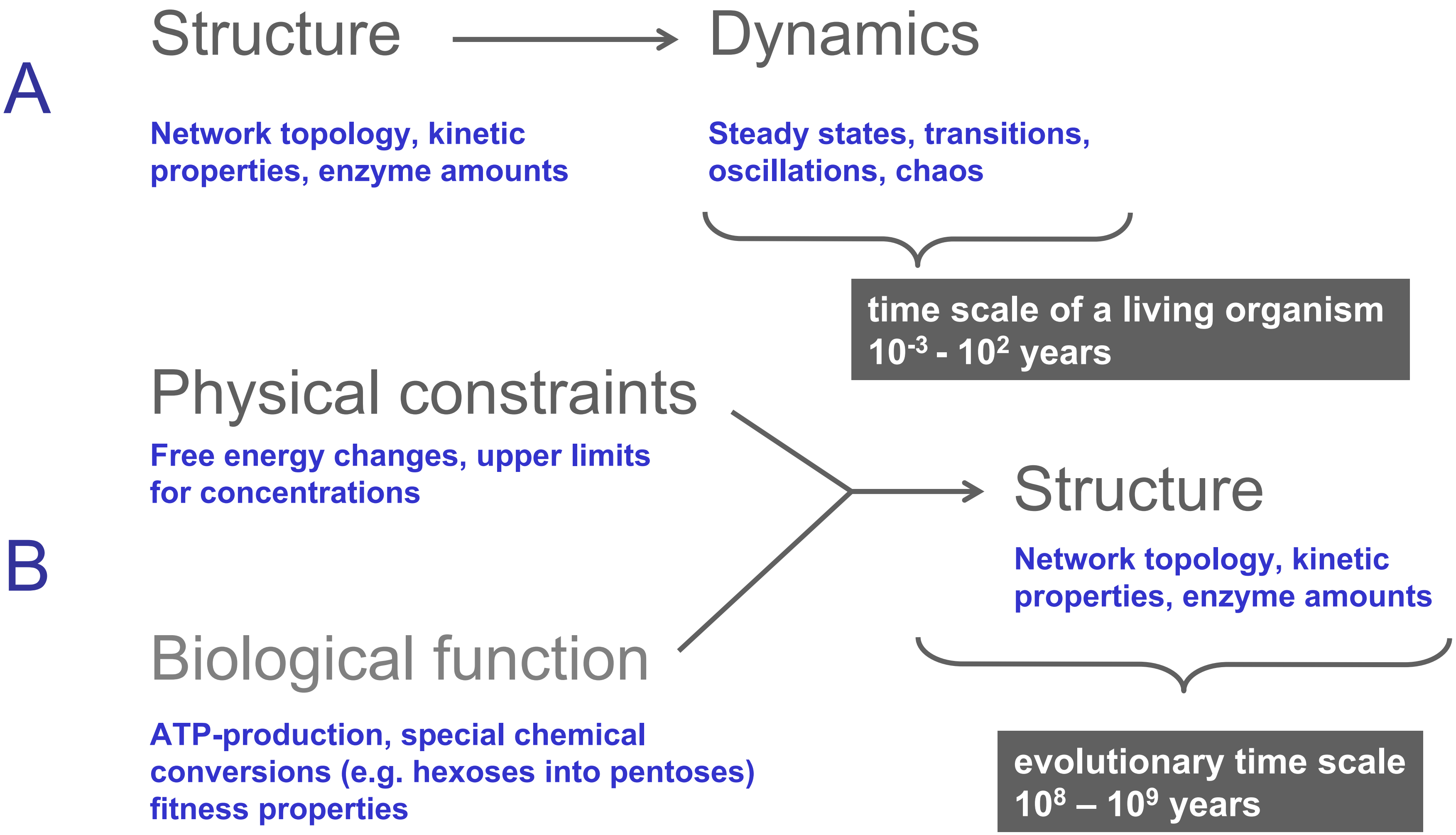
Here is a (partial) list of wishes

- **Protein folding and functioning**
- Protein docking and recognition
- Immunological recognition
- Gene expression and regulation
- **Metabolic networks**
- System biology
- etc.

- Protein/DNA interactions
- **Amyloid aggregation**
- Memory and networking
- miRNA/siRNA
- Signal transduction
- **Nano-bio devices**
- etc.

Not to talk about the ultimate goal,
of curing all possible diseases

What we would like to know about METABOLIC NETWORKS



A

Structure → Dynamics

Network topology, kinetic properties, enzyme amounts

Steady states, transitions, oscillations, chaos

time scale of a living organism
10^{-3} - 10^{2} years

B

Physical constraints

Free energy changes, upper limits for concentrations

Biological function

ATP-production, special chemical conversions (e.g. hexoses into pentoses)
fitness properties

→ Structure

Network topology, kinetic properties, enzyme amounts

evolutionary time scale
10^{8} – 10^{9} years

SIMULATION MODELS

stoichiometric coefficients

metabolite concentrations

reaction rates

$$\frac{dS_i}{dt} = \sum_{j=1}^{r} n_{ij} v_j \qquad \frac{dS}{dt} = N \cdot v$$

$$v = (v_1, \ldots, v_r) \qquad v = v(S, k)$$

Attractors, Chaos (?)

Robustness

Recovery of function

Kinetic parameters

large number 10-1000 of variables

large number 10-1000 of equations

non-linearity

regulatory loops

separation of time scales

natural selection of kinetic parameters

What we would like to know about PROTEINS

- primary structure → folding → function

  linear **3D** **conform. switches**

even tiny atomic displacements matter

  - predict geometry and dynamics of folding and conformational changes

    3D times e.g. heme, rhodopsin

  - predict function

    motif conservation, structural similarity

- evolution/selection → $\#10^7$ among $(\#10^2)^{20}$ possibilities

  folding vs aggregation?

  - understand mis-folding and aggregation

    Mad cow (Prion), Amyloidosis (e.g. β-amyloid in Alzheimer disease)

- recognition/docking

  Ab *vs* Ag, …, transcription factors, promoters, …

  - characterize macromolecules binding
  - clarify molecular mimicry and auto-immune reactions

SIMILATION MODELS

Coarse grained models

- **how general is folding?**
  - Geometrical considerations
  - Lattice models
  - Statistical Mechanics

spin glasses

Atomistic models

- **classical**
  - Minim. of config. energy (no entropy)
  - Canonical/micro-canonical simulations
  - Multi-canonical simulations

"right" ensemble?
"right" thermodynamic variables?

- **QM/MM**
- **QM**
  - Quantum Chemistry
  - DFT
  - Car-Parrinello dynamical simulations

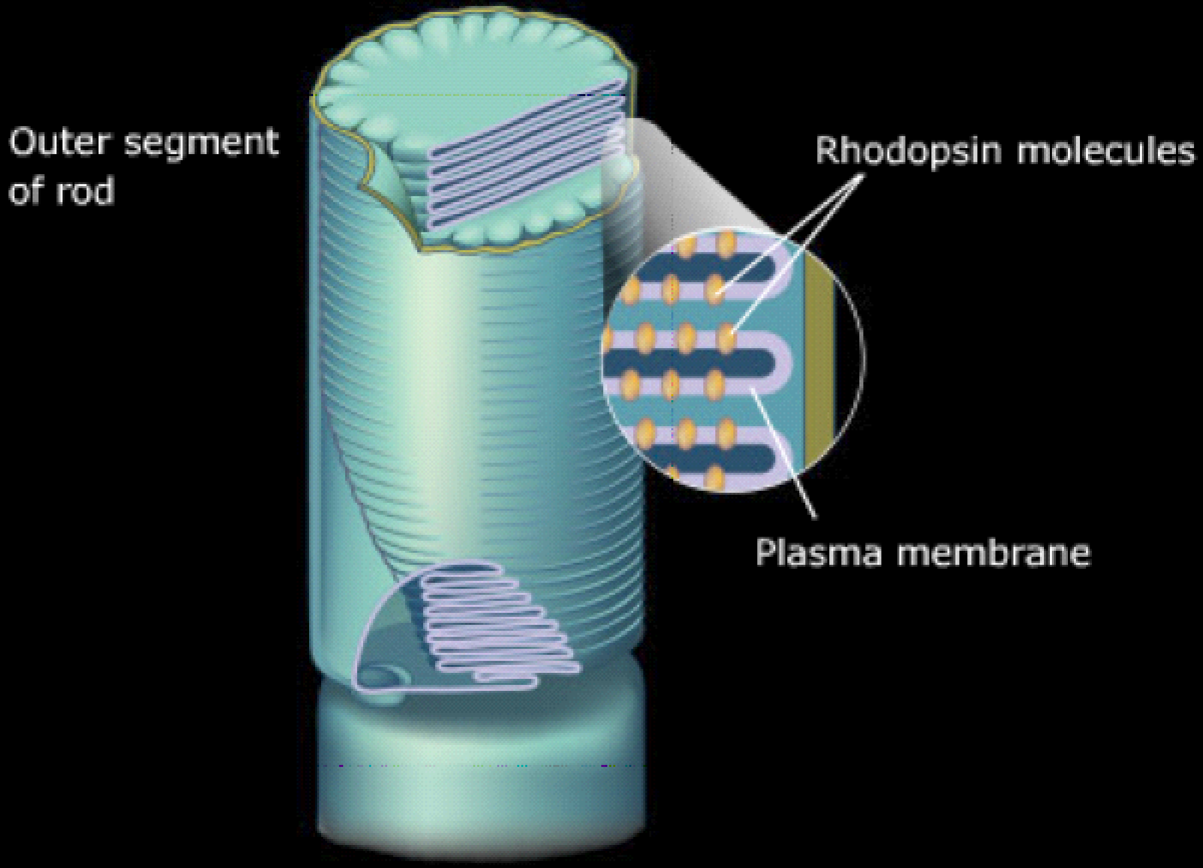
even tiny atomic displacements matter

Two examples, among ∞-ly many

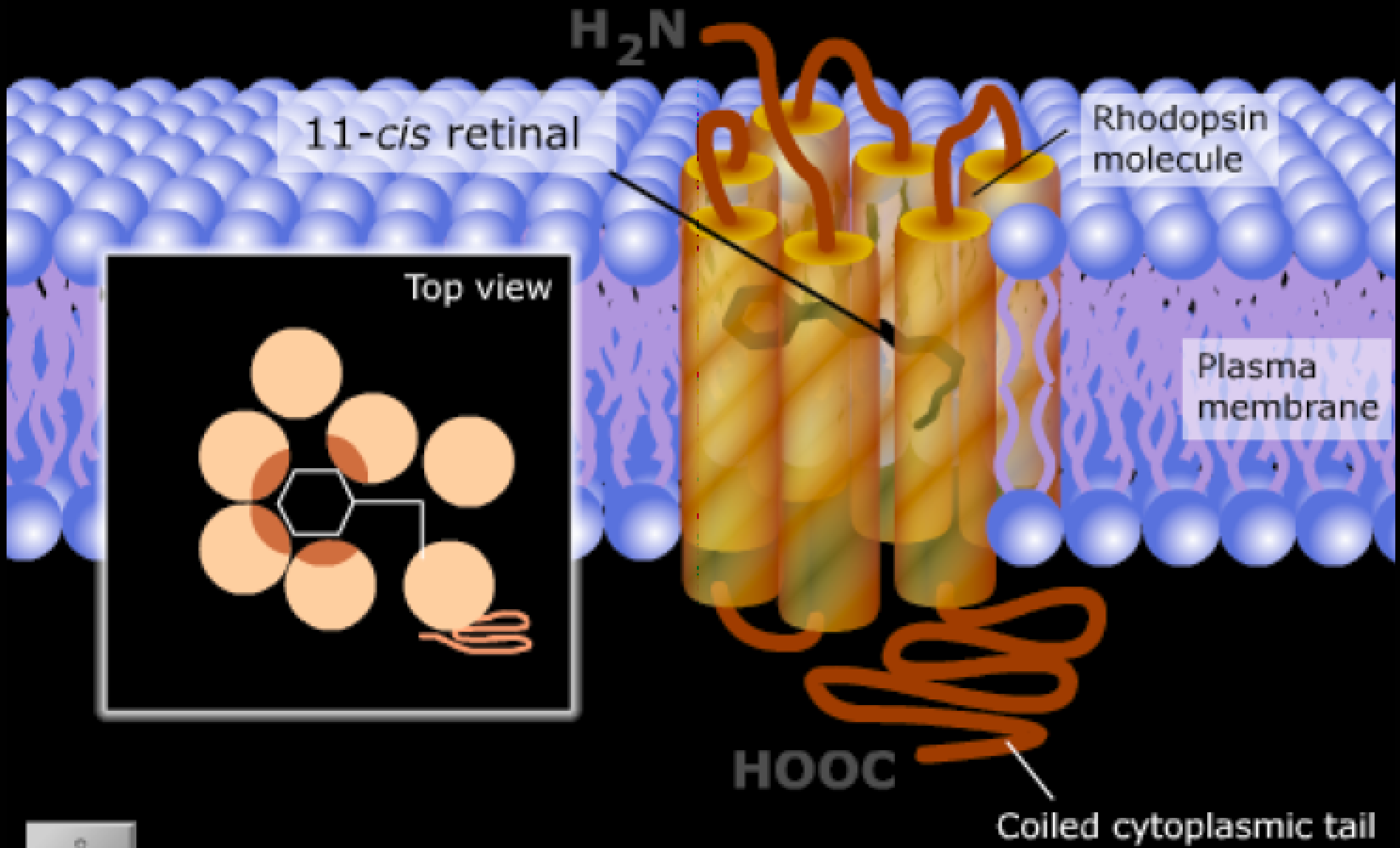
I. Cis →Trans isomerization of 11-cis retinal

II. Hemoglobin “breathing”

Photoisomerization of rhodopsin

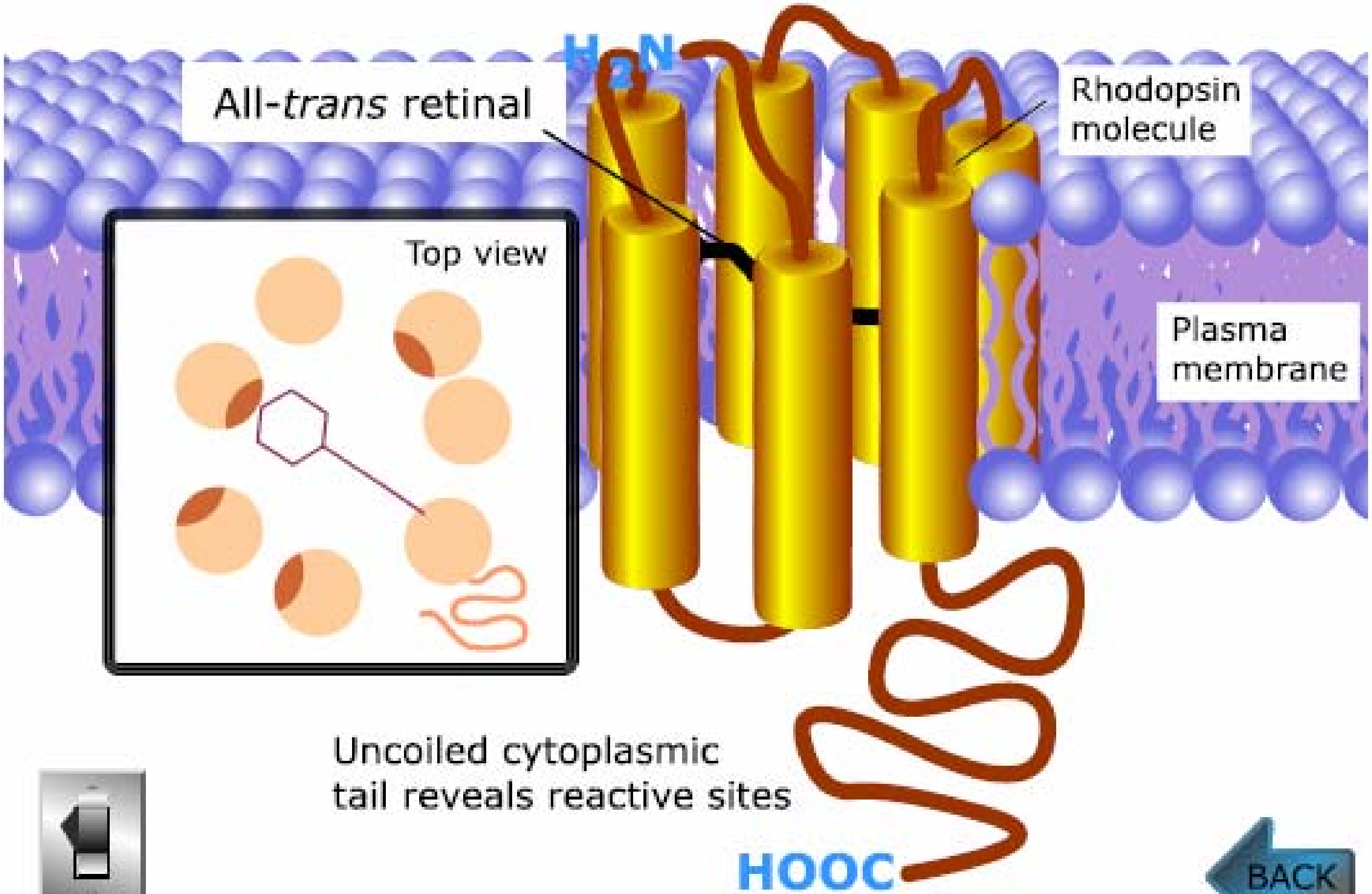


Photoisomerization of rhodopsin



BACK

Photoisomerization of rhodopsin

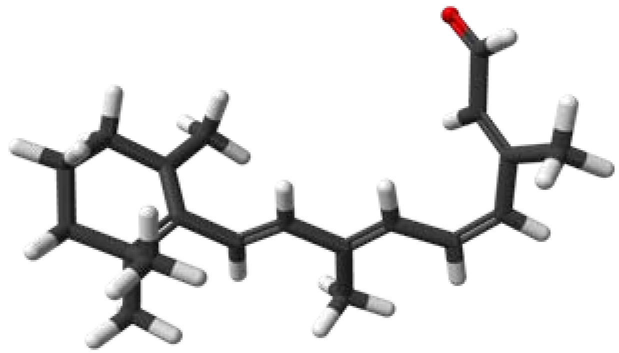


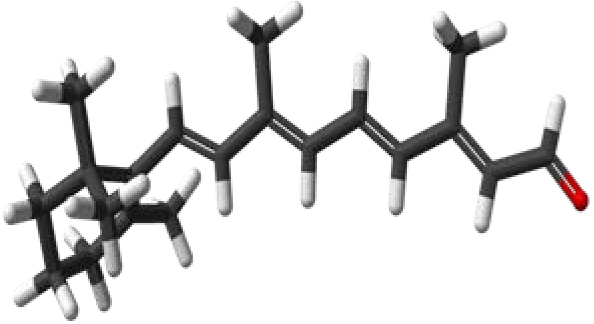
All-*trans* retinal

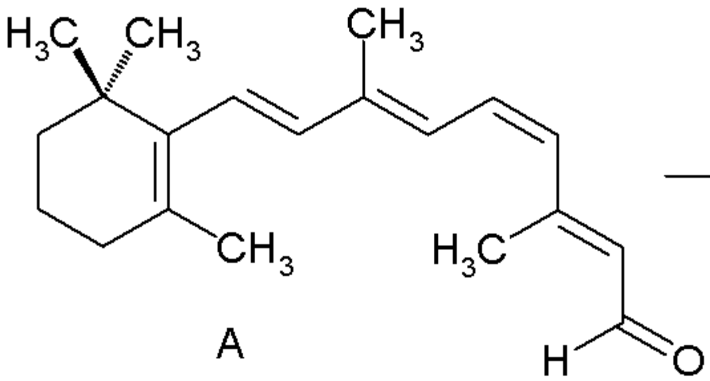
Rhodopsin molecule

Plasma membrane

Uncoiled cytoplasmic tail reveals reactive sites





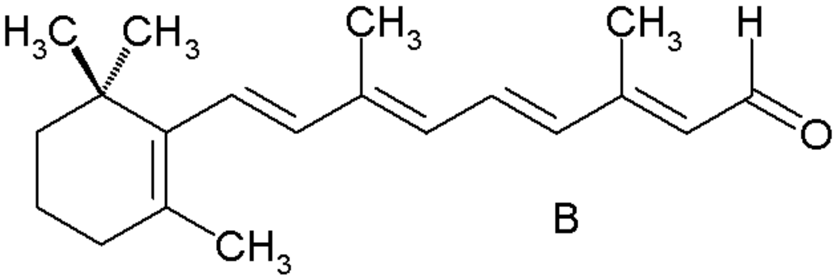


A

γ

light

11-cis retinal

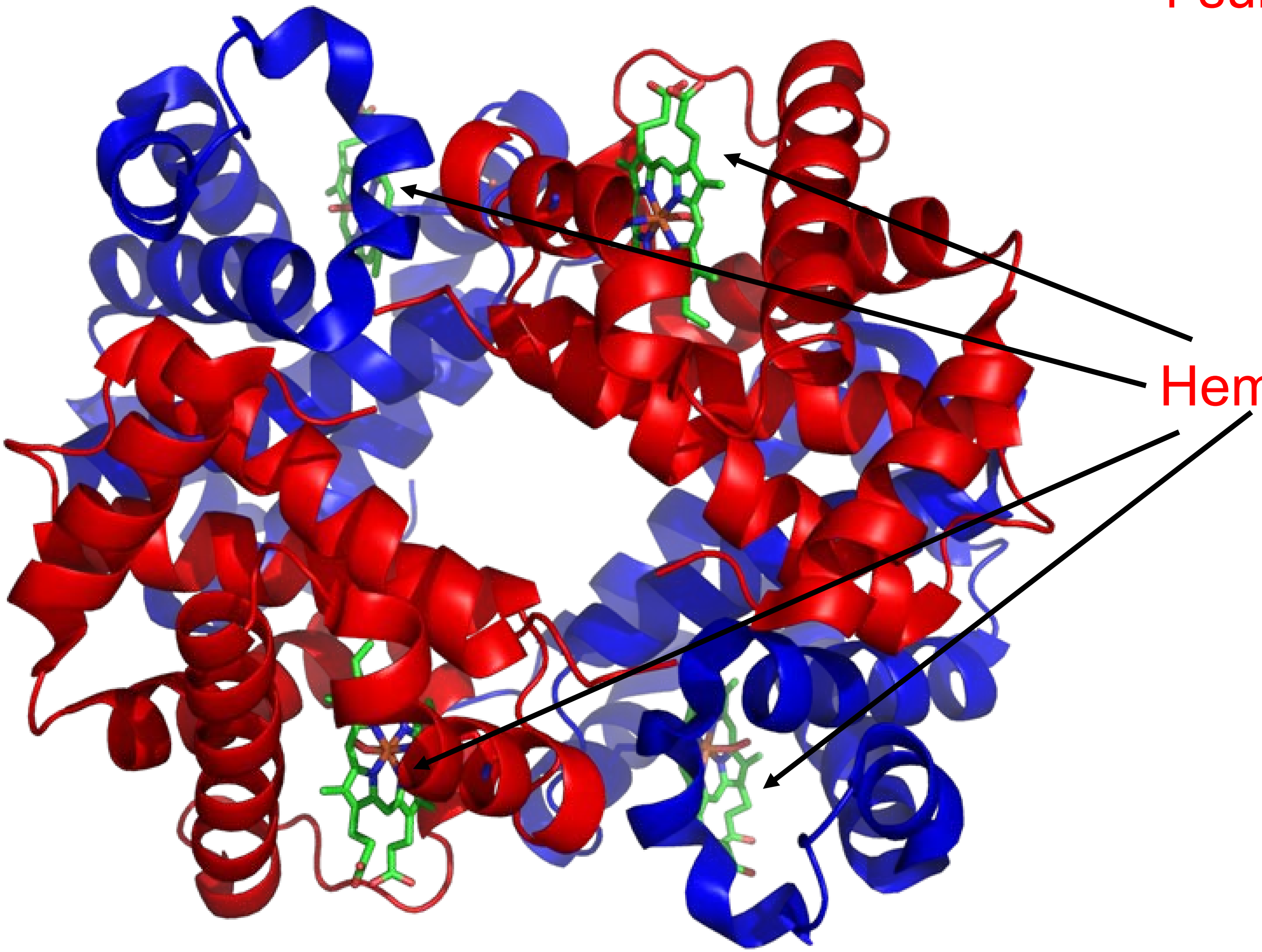


B

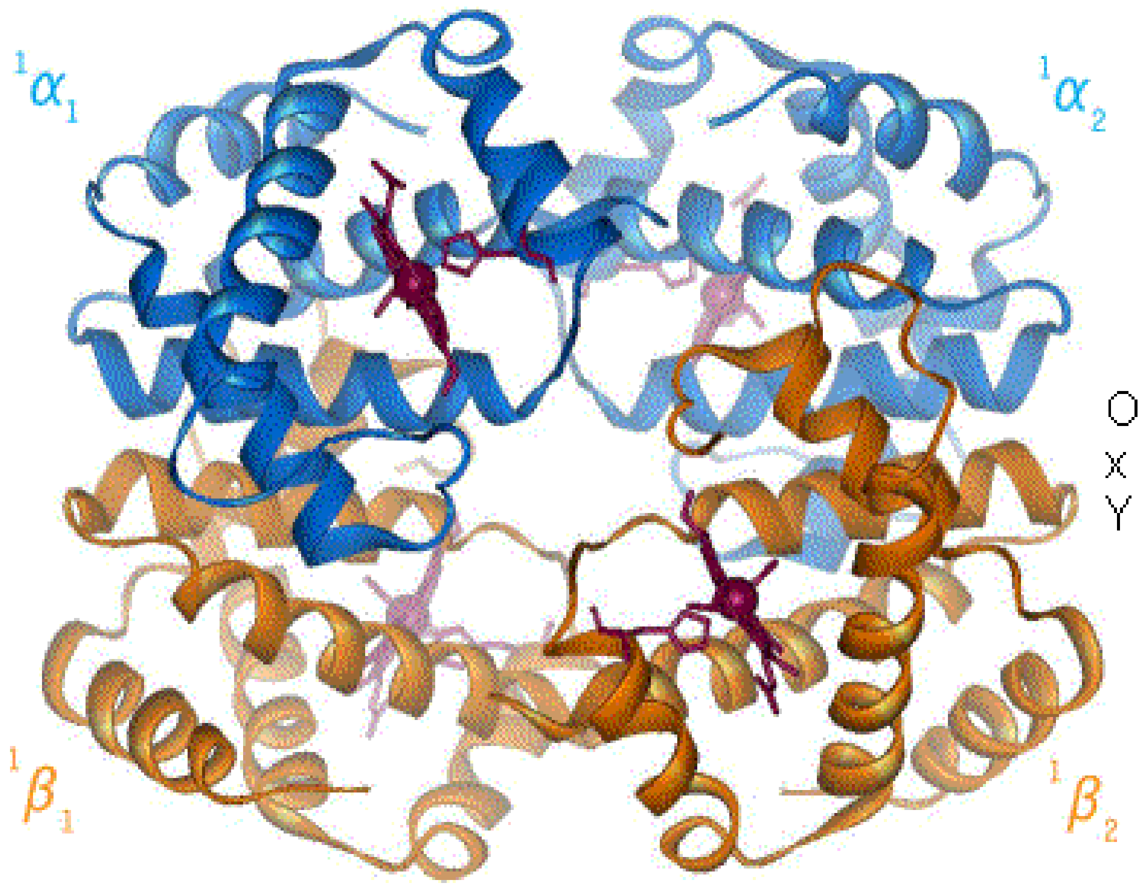
all-trans retinal

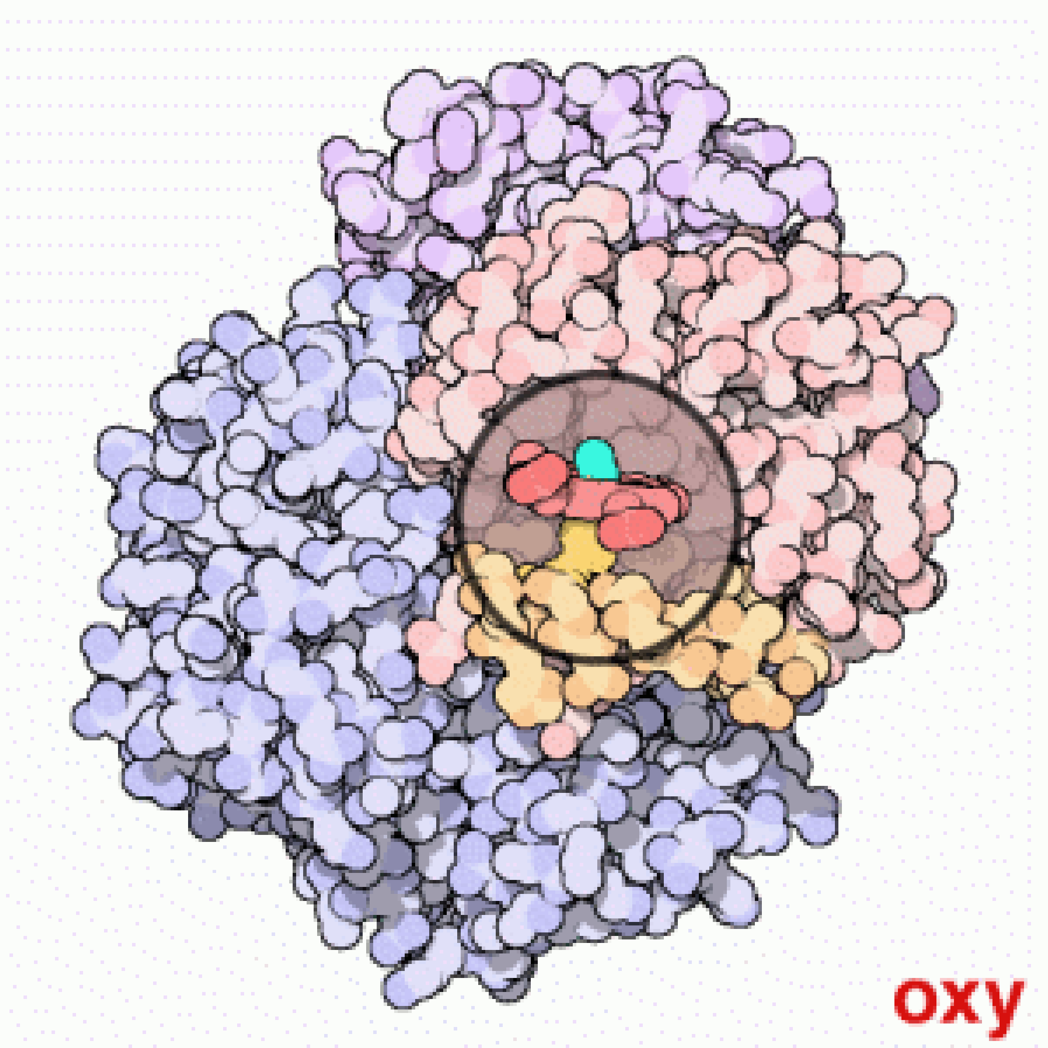
Hemoglobin

4 subunits

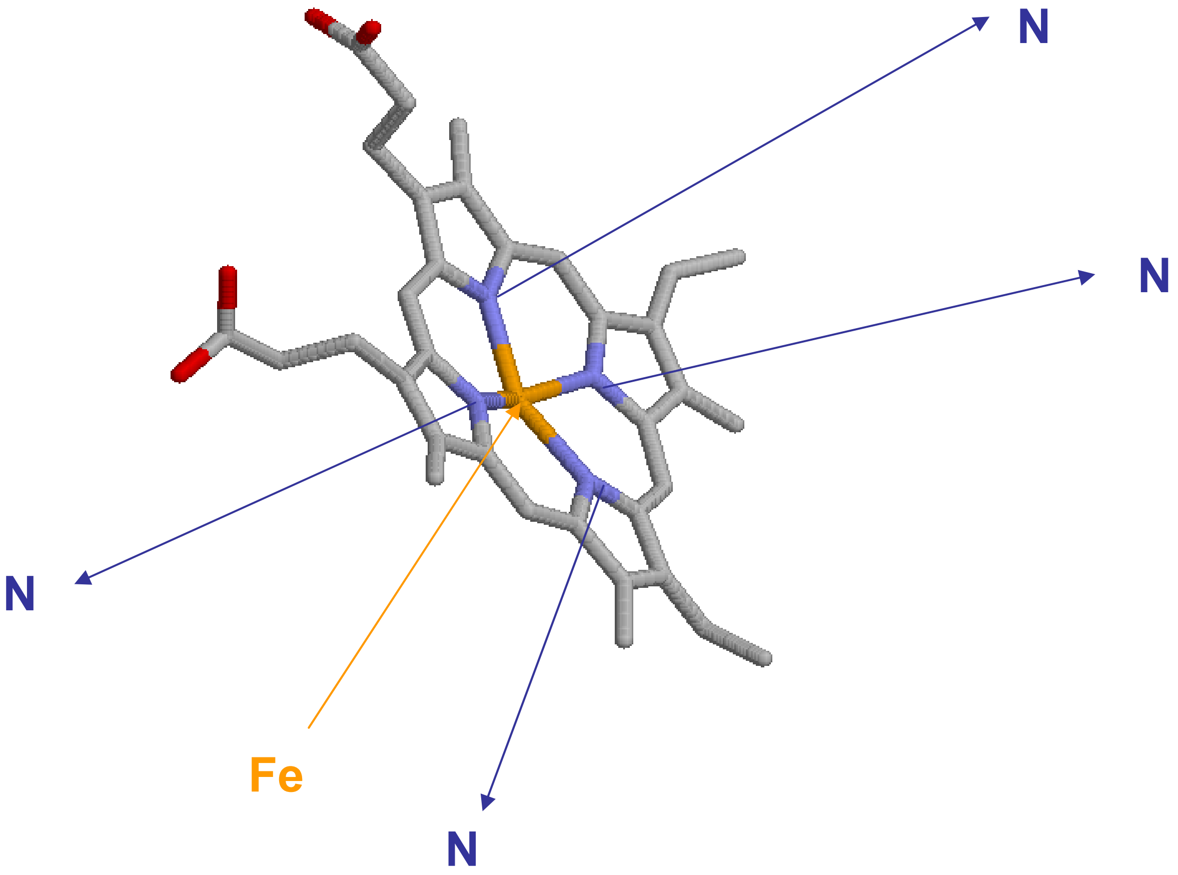


Heme

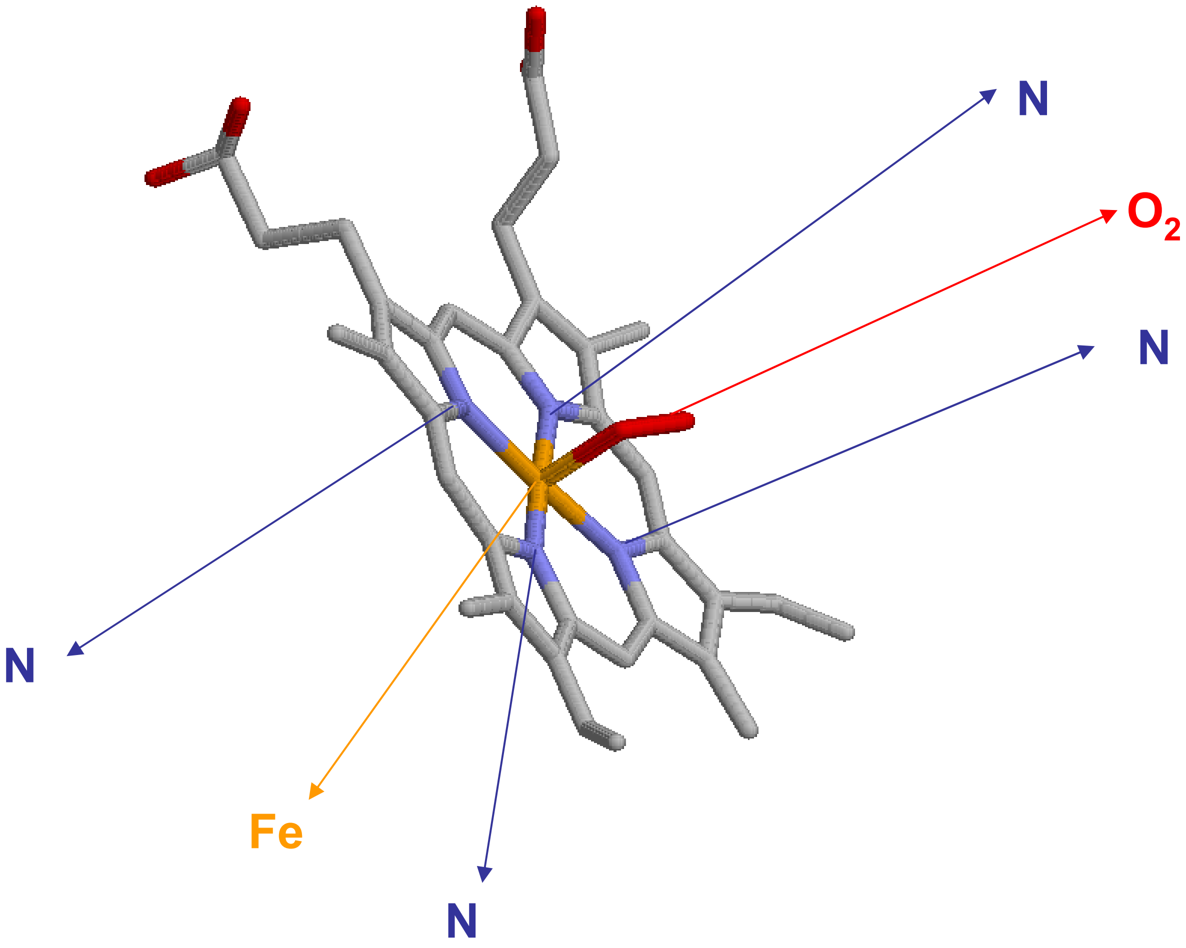

$^1\alpha_1$
$^1\alpha_2$
O
X
Y
$^1\beta_1$
$^1\beta_2$


oxy

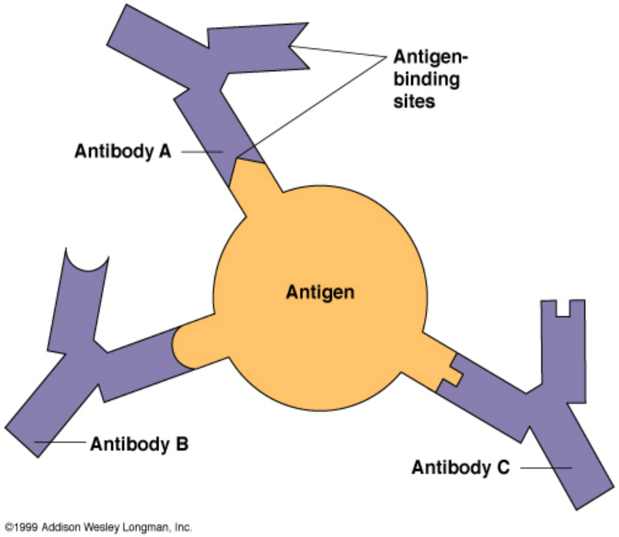
Deoxi-hemoglobin



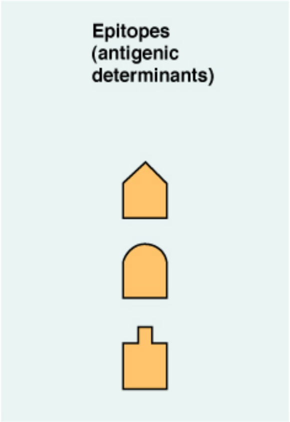
Oxi-hemoglobin



Antigen-functionalized Nanotube for disease diagnosis



©1999 Addison Wesley Longman, Inc.



- Specific Antibodies Ab are produced in response to an external Antigen Ag (like a viral or bacterial protein)

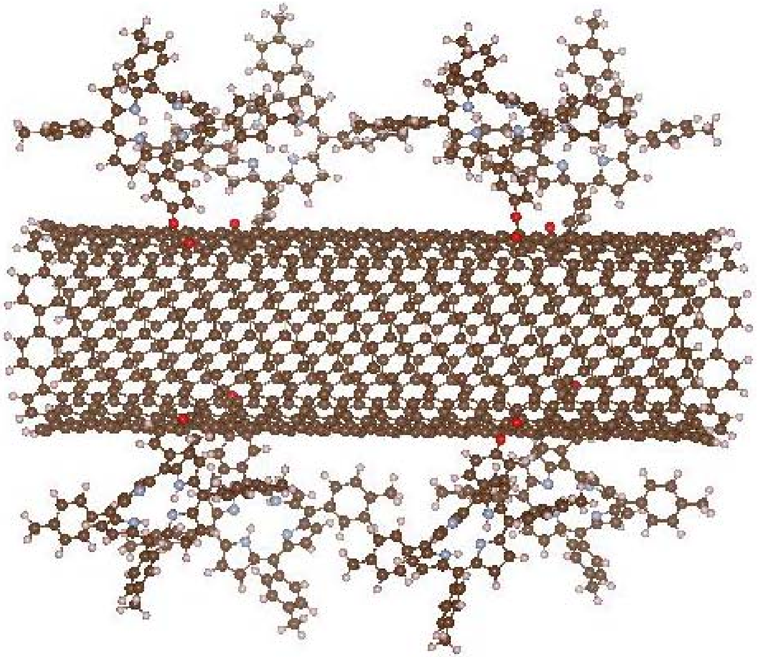
- If you have been infected by Ag_A, you will produced Ab_A, detectable in your blood

- One would like to functionalize a nanotube with the Ag we wish to detected

- Questions

  - can all this be done?
  - will it work (specificity)?
  - can one detect a signal upon Ab_A binding the Ag_A?
  - can simulations be of help?

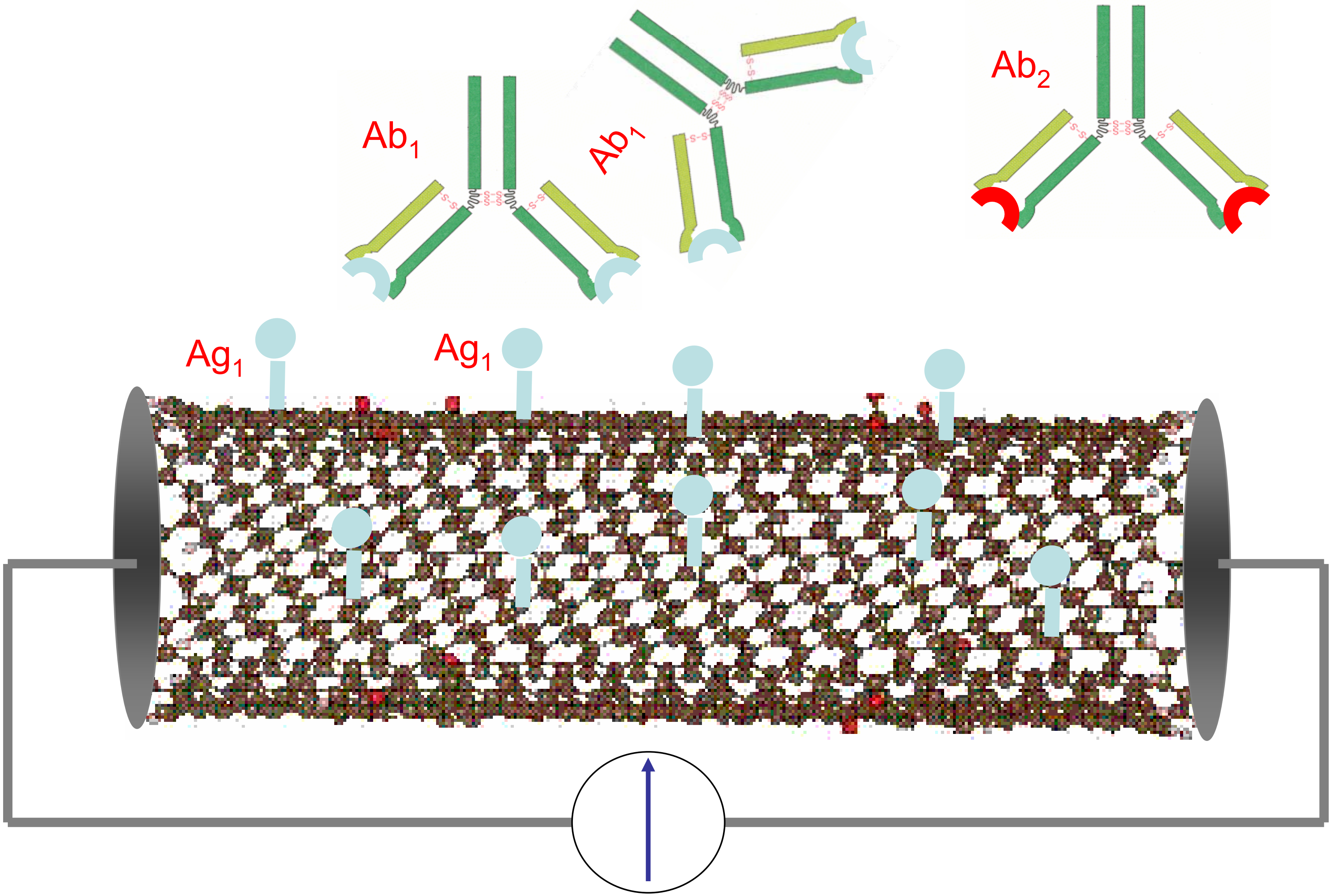
Porphyrin Functionalized Nanotube

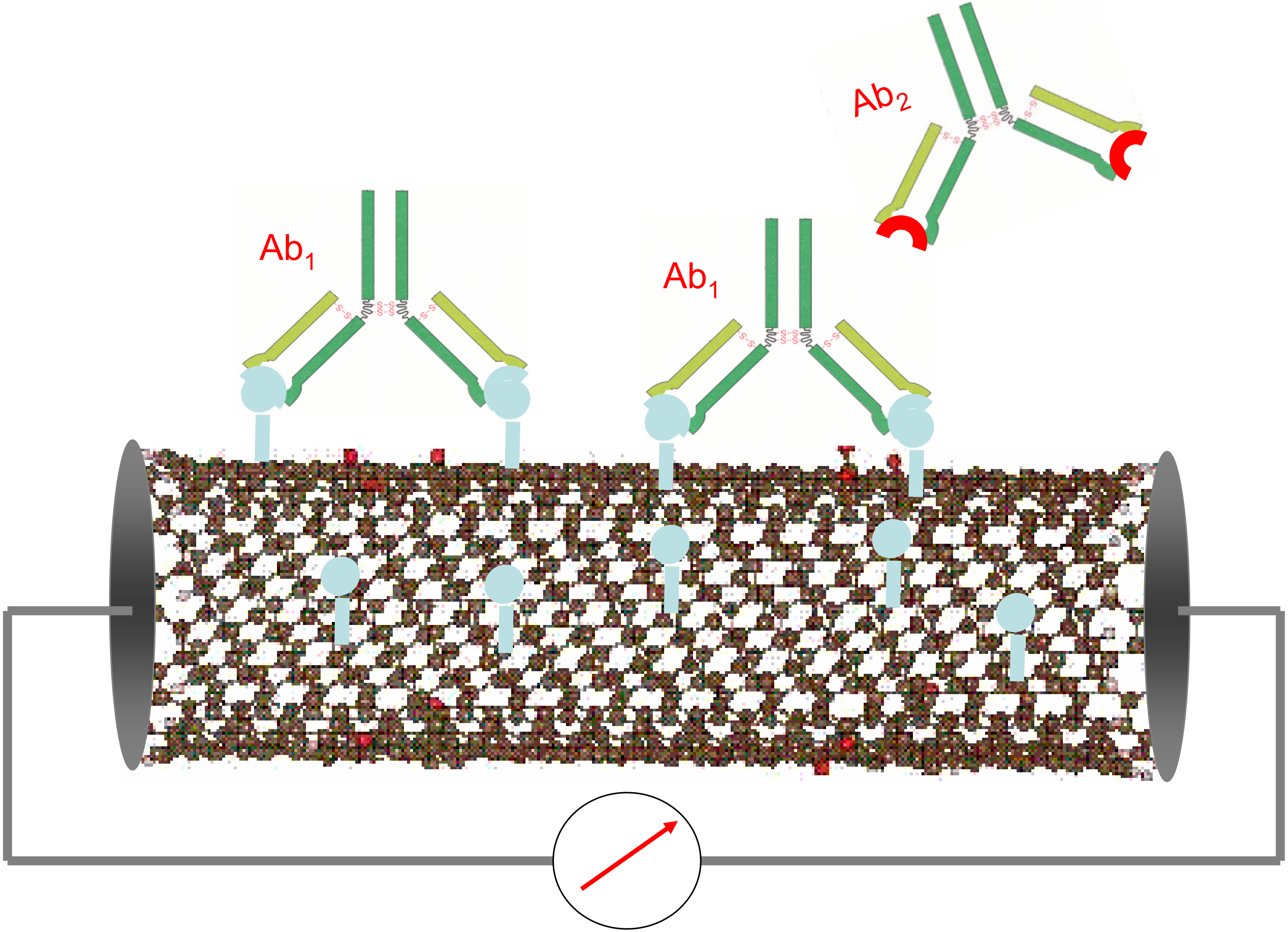


- New materials for **solar energy** applications
- Relatively simple, synthetically feasible (at ORNL-UT) mimics of light-harvesting antenna units
- **Porphyrin** molecules are the light absorbing antenna and the **nanotube** may provide a conducting channel
- Key research questions to address are:
  - How does porphyrin attach to the nanotube?
  - How does the electronic structure change as porphyrin molecules are added to the nanotube (up to 22 % in weight)?
  - How is the conductance affected by surface orientation and composition?
- Problem size **1500** (~ 60 Å) to **5000** atoms (202 Å by 60 Å)

  10 times more electrons

a case for numerical simulations


Ab$_1$
Ab$_1$
Ab$_2$
Ag$_1$
Ag$_1$


Ab$_2$
S–S
S–S
S–S
S–S
Ab$_1$
S–S
S–S
S–S
S–S
Ab$_1$
S–S
S–S
S–S
S–S

IV. What we can actually do and/or are really doing

Two examples

IVa Metabolic networks

IVb Protein folding and aggregation

IV. What we can actually do and/or are really doing

Two examples

IVa Metabolic networks

IVb Protein folding and aggregation

◘ Metabolic networks

- The case of the WNT pathway

  the context and the problem

- Modelization

  data and approximations

- Results

  some numerics

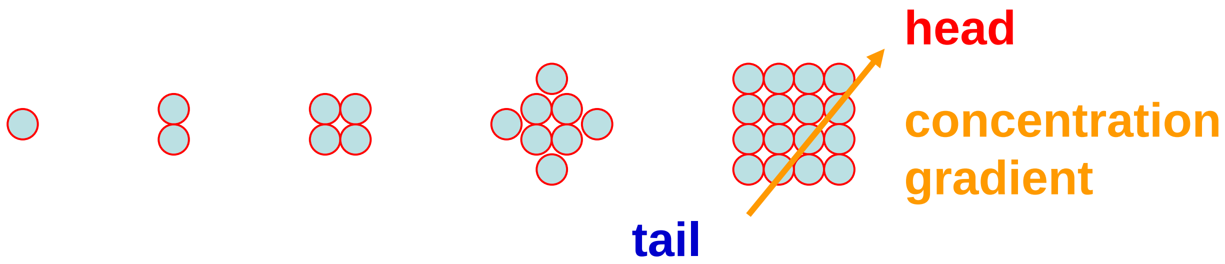
- Outlook

  understanding cancer onset (?)

A paradigmatic case: the WNT pathway

- Morphogenes are proteins that specify the different cell fate in a **concentration** dependent way

- WNT, Hh, BMP, … regulator proteins that (during embryogenesis) provide positional information and organize embryonic patterning



- WNT-signalling mechanism is much studied, because defects in its regulation ultimately lead to cancer

- Normally WNT regulates the level of β-catenine in the cell

1) In the absence of a WNT signal, a multi-component destruction complex, containing GSK3, Axin, ACP,… promotes Phosphorilation of β-catenine, making it ready for degradation by β-TRCP (an E3 Ubiquitin ligase)

2) In the presence of a WNT signal, the activity of the destruction complex is inhibited, and the level of cytoplasmatic β-catenine rises

β-catenine becomes complexed with the transcription factor TCF and activate TCF-target genes (c-myc, cyclinD1, tcf-1,…), which directly influence cell development processes

⇧

mechanism for cell differentiation → embryon **OK**

mechanism for cell differentiation → adult **KO**

⇩

Accumulation of β-catenine in the cell and/or deregulation of the TCF/β-catenine activity can promote carcinogenesis in many tissues

- Mutations in the β-catenine gene CTNNb1 with consequent protein alterations (mostly in the region S29-K49)

- Defects in the WNT pathway, resulting in a deregulation of the cytoplasmatic β-catenine level

● Modeling the canonical WNT pathway

Lee Salic Krueger Heinrich Kirschner
PloS Biology, **1** (2003) 116

MAIN COMPONENTS

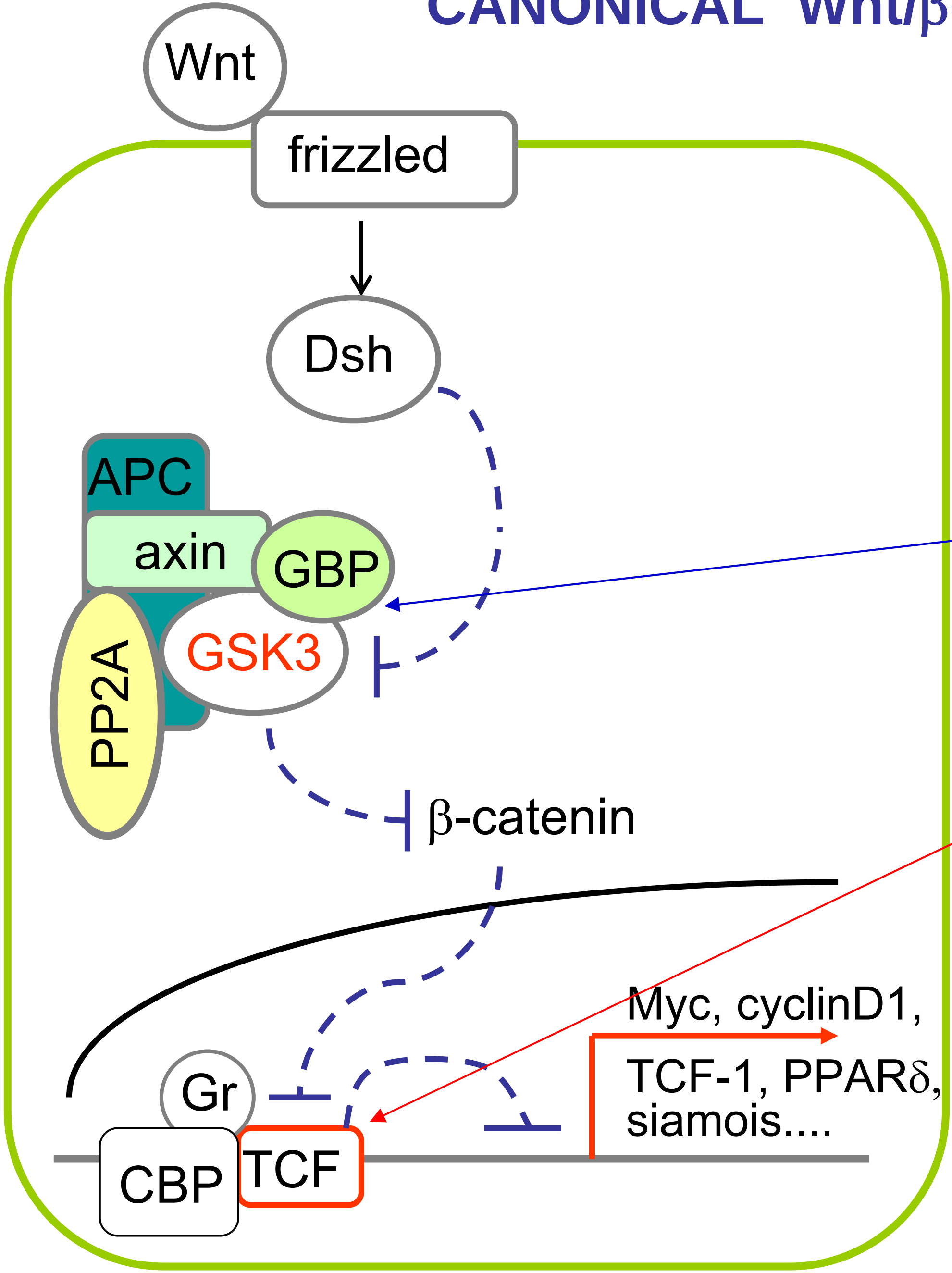
WNT (ligand)
FRIZZLED (receptor)
DISHEVELLED
AXIN (scaffold)
APC (scaffold)
GSK3 (Kinase)
GBP (GSK3 binding protein)
PHOSPHATASE (PP2A)
CASEINE KINASE
β-CATENINE (transcription coactivator)
TCF (transcription factor)

and many, many more …

MUTATIONS IN APC PLAY A PARTICULARLY IMPORTANT ROLE IN COLORECTAL CANCER

APC: ADENOMATOUS POLYPOSIS COLIPROTEIN

CANONICAL Wnt/β-CATENINE PATHWAY



LOGIC OF THE PATHWAY

Either the destruction complex forms

or β-catenin gets complexed with TCF

WITH ACTIVATION OF

WNT-TARGET GENES

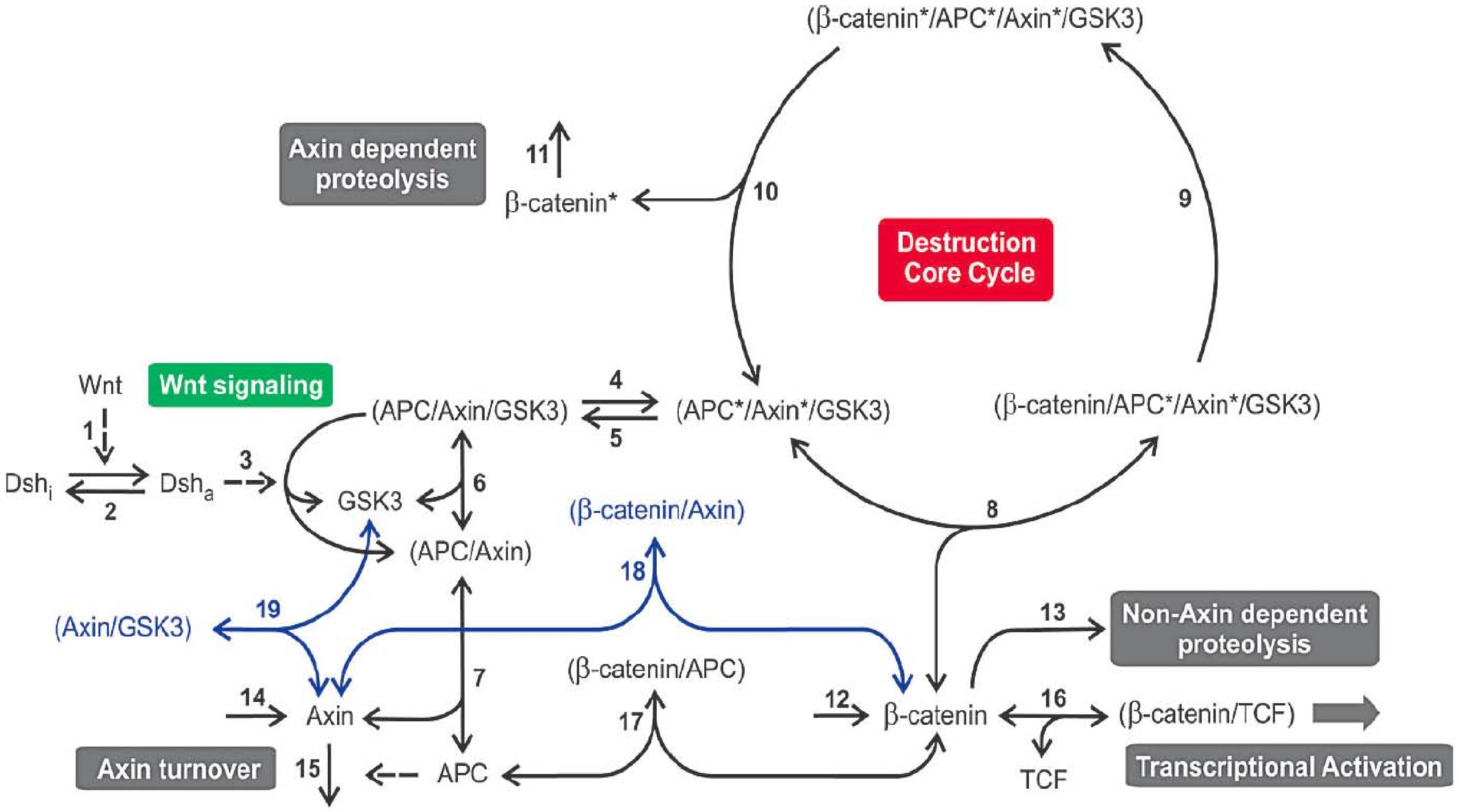
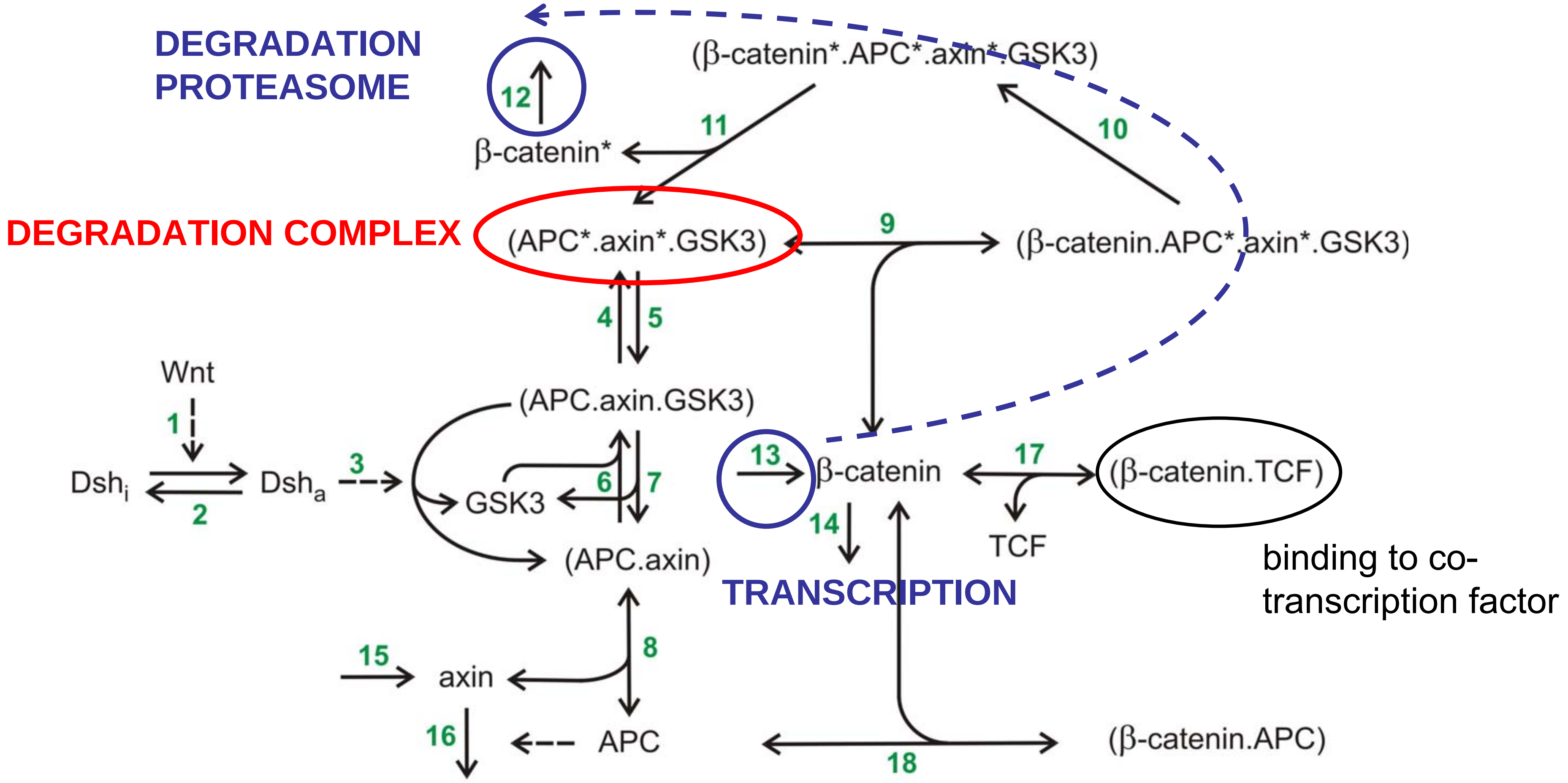


Figure 1. Reaction Scheme for Wnt Signaling

The reaction steps of the Wnt pathway are numbered 1 to 19. Protein complexes are denoted by the names of their components, separated by a slash and enclosed in brackets. Phosphorylated components are marked by an asterisk. Single-headed solid arrows characterize reactions taking place only in the indicated direction. Double-headed arrows denote binding equilibria. Blue arrows mark reactions that have only been taken into account when studying the effect of high axin concentrations. Broken arrows represent activation of Dsh by the Wnt ligand (step 1), Dsh-mediated initiation of the release of GSK3β from the destruction complex (step 3), and APC-mediated degradation of axin (step 15). The broken arrows indicate that the components mediate but do not participate stoichiometrically in the reaction scheme. The irreversible reactions 2, 4, 5, 9–11, and 13 are unimolecular, and reactions 6, 7, 8, 16, and 17 are reversible binding steps. The individual reactions and their role in the Wnt pathway are explained in the text.

DOI: 10.1371/journal.pbio.0000010.g001

Unstimulated reference state
Absence of Wnt



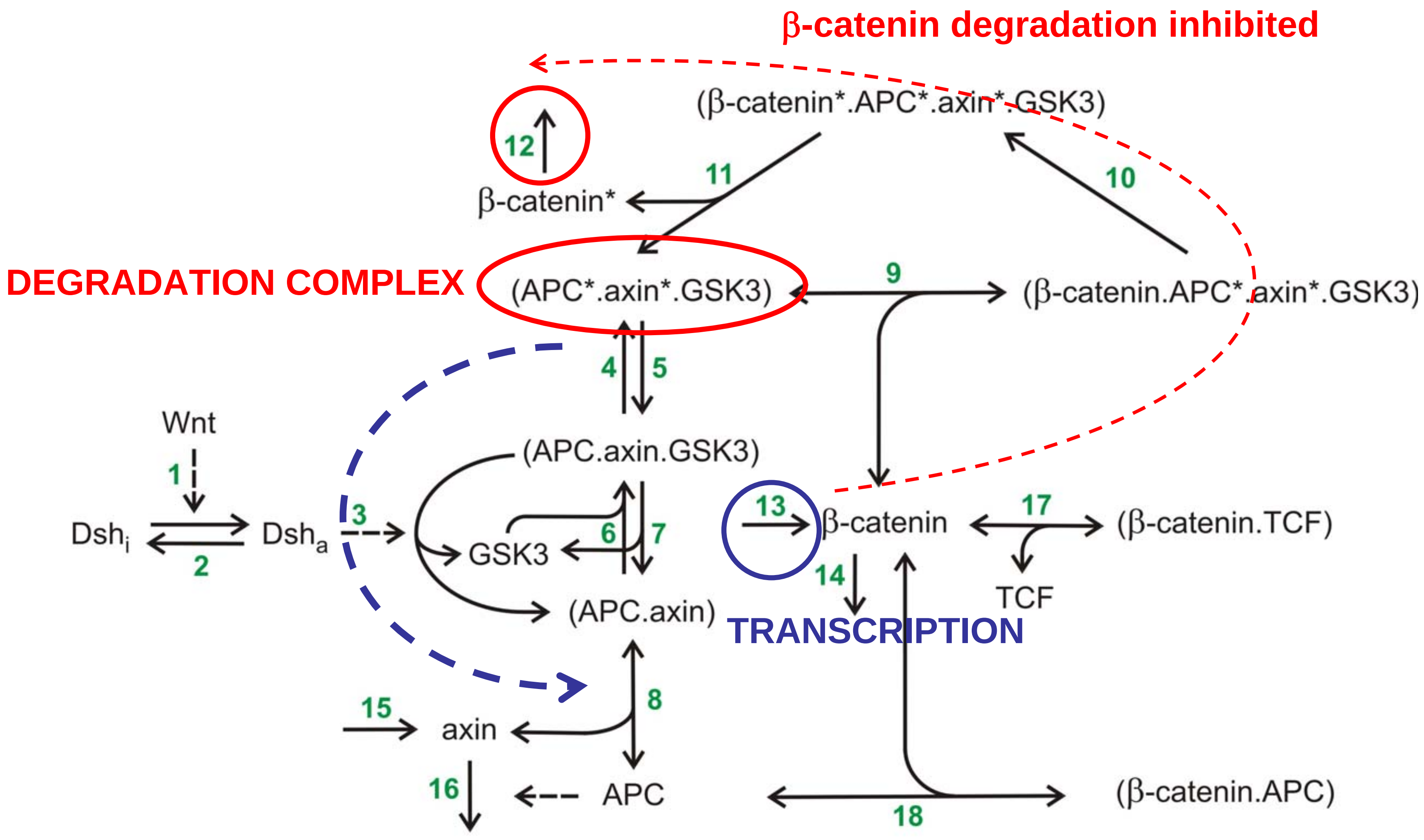
DEGRADATION PROTEASOME

DEGRADATION COMPLEX

TRANSCRIPTION

binding to co-transcription factor

Effect of Wnt-stimulation



β-catenin degradation inhibited

DEGRADATION COMPLEX

TRANSCRIPTION

ACCUMULATION OF β-CATENIN

MAIN INPUT DATA OF THE MODEL

CONCENTRATIONS

total Dsh	**100 nM**
total APC	**100 nM**
total TCF	**15 nM**
total GSK3	**50 nM**
total axin	**0.02 nM**
total β-catenin	**35 nM**
free phosphorylated β-catenin	*1 nM*

DISSOCIATION CONSTANTS

binding of GSK3 to (APC.axin)	*10 nM*
binding of APC to axin	*50 nM*
binding of β-catenin to (APC.axin.GSK)	*120 nM*
binding of β-catenin to TCF	*30 nM*
binding of β-catenin to APC	*1200 nM*

FLUXES

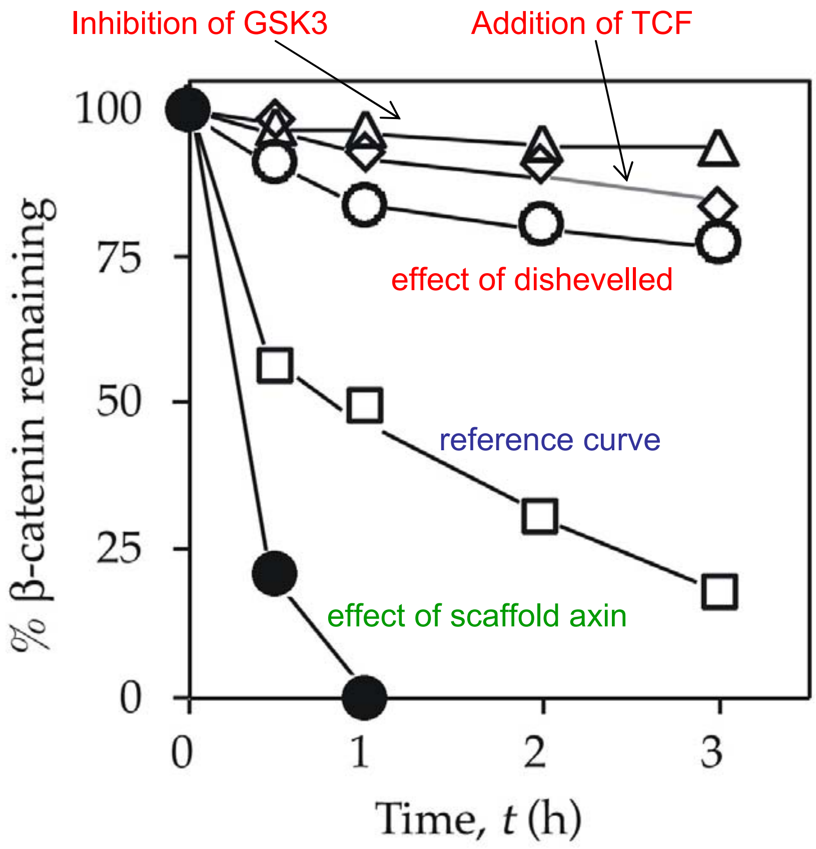
degradation flux of β-catenin via the proteasome	**25 nM/h**
Share of degradation of β-catenin via unphosphorylated form	**1.5 %**

CHARACTERISTIC TIMES

phosphorylation/dephosphorylation of APC and axin	*2.5 min*
GSK3 association/dissociation	*1 min*
Axin degradation	**6 min**

● Results

β-catenin degradation, simulations and comparison with experimental data



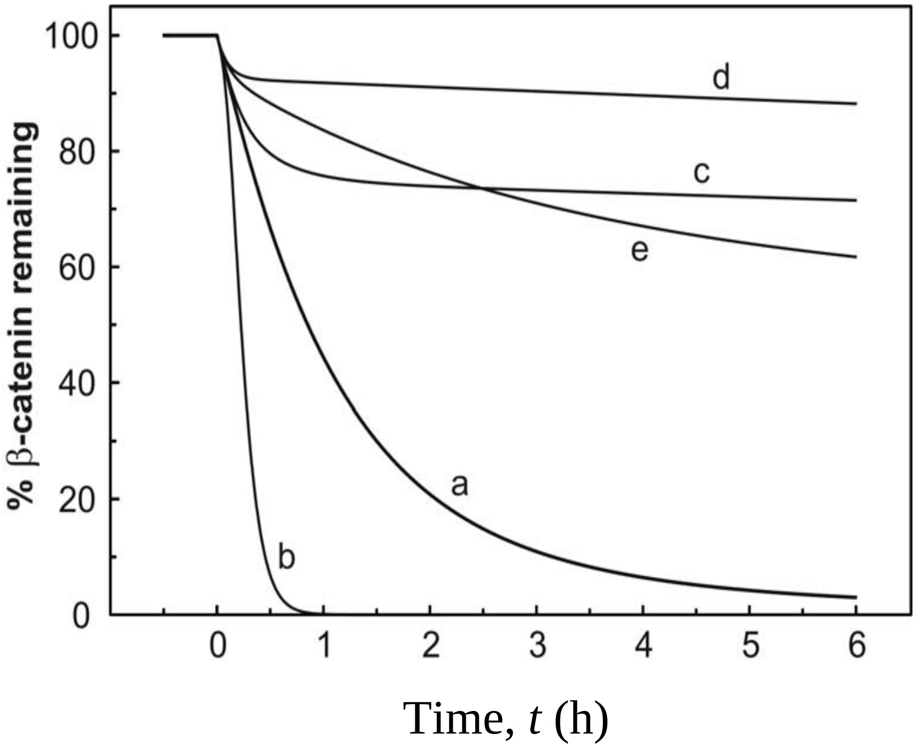
Inhibition of GSK3

Addition of TCF

effect of dishevelled

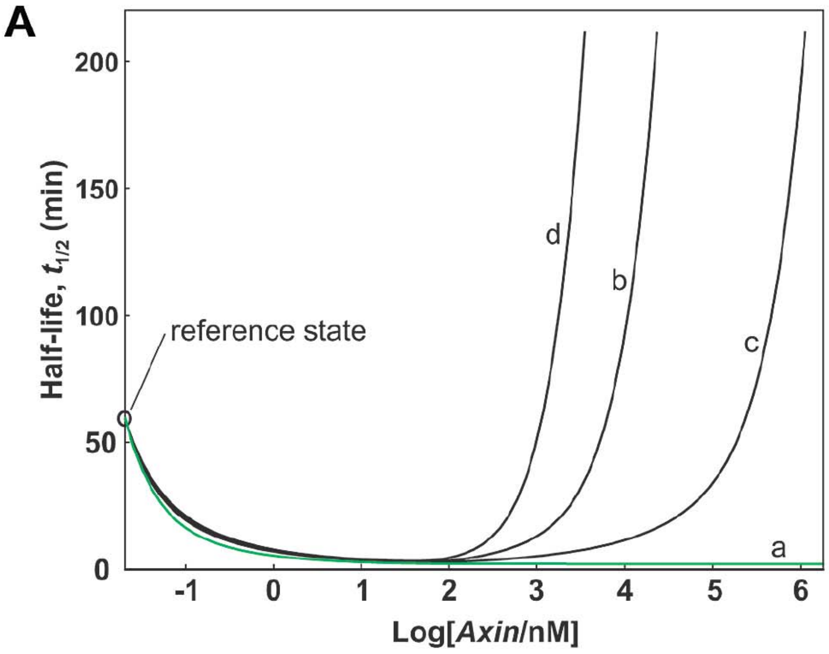
reference curve

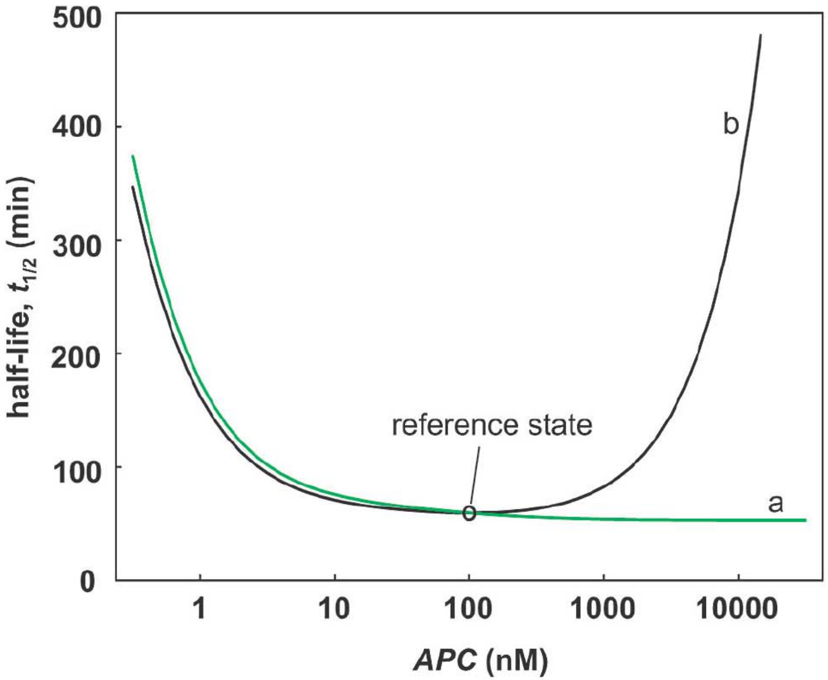
effect of scaffold axin



● Outlook

Tumor suppressor role of Axin and/or APC?





Very complicated to devise a winning strategy (non-linear dynamics)

- Axin degradation is APC dependent
- Axin and APC both involved in the β-catenin destruction complex

HUMBOLDT-UNIVERSITY

Reinhart Heinrich

SIGNAL TRANSDUCTION

Thomas Höfer

Roland Krüger

Holger Nathansen

METABOLIC NETWORKS

Oliver Ebenhöh

Edda Klipp

Stefan Schuster

Jana Wolf

HARVARD MEDICAL SCHOOL, BOSTON

Marc Kirschner

Leon Murphy

Benjamin G. Neel

Tom A. Rapoport

Adrian Salic

Ethan Lee

Stefanie Schalm

BIOCENTRUM AMSTERDAM

Hans Westerhoff

Roel van Driel

Martijn Moné

UNIVERSITY BORDEAUX II

Jean-Pierre Mazat

Christine Reder

Summary

what can be/was done about metabolic networks

- Bio-chemical data suggest the set of relevant
  - compounds/reaction constants
  - chemical reactions
  - network topology
- Construct the set of (non-linear) diff. eqs (time and space) for concentrations
- Solution
  - identify relevant initial states
  - evolve the equations
  - stability studies around different points in concentration space
- Devise experiments and compare
- Identify the key features of the system
  - compounds/reaction constants
  - chemical reactions
  - network topology
- hence what is needed to correct what goes wrong

(like accumulation of β-catenin in adult cell, as it would
promote unwanted expression of the silenced TRCβ gene)

▣ Protein folding and aggregation

- Generalities
- Universality *vs* natural selection

  the case of random hetero-polymers
- Folding *vs* aggregation

  the case of the Prion protein (PrP)
  and the role of Cu
- XAS (NMR, EPR) experiments

  data analysis and EXAFS theory
- QM calculations

  DFT and Car-Parrinello dynamics

• Generalities

Protein is a complex (and complicated) system

- Many degrees of freedom

  protein: ~ 300 a.a.'s x 10 atoms = ~ 3000 atoms
  solvent: ~ 1000 atoms

  → 3 to 4 times more "active" electrons

- Large range of folding times

  from *μsec's* to *sec's*

  the Levinthal's paradox {
  too fast for an exhaustive search
  too slow for a straight descent to absolute minimum

- Interaction is not short-range

  choice of a phenomenologically acceptable potential in MD
  a Q.M. treatment (DFT, Car-Parrinello) is often needed

- Free-energy landscape looks very corrugated

  many hierarchically organized local minima, separated by high barriers

- System is not living at thermodynamic equilibrium

  flux of energy and matter

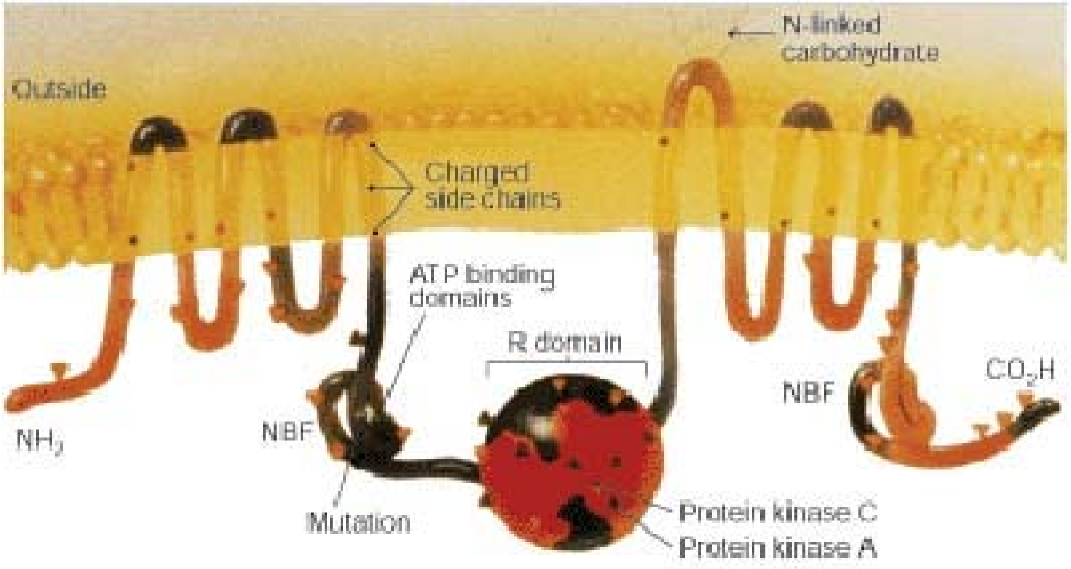
- Even single mutations matter

  though not always

a dramatic case

The CFTR gene is found at the q31.2 locus of chromosome 7, is 230 000 base pairs long, and creates a protein that is 1,480 amino acids long. The most common mutation, ΔF508 is a deletion (Δ) of three nucleotides that results in a loss of the amino acid phenylalanine F at the 508th position on the protein. This mutation accounts for two-thirds of CF cases worldwide and 90 percent of cases in the United States, however, there are over 1,400 other mutations that can produce CF.

There are several mechanisms by which these mutations cause problems with the CFTR protein. ΔF508, for instance, creates a protein that does not fold normally and is degraded by the cell. Several mutations, which are common in the Ashkenazi Jewish population, result in proteins that are too short because production is ended prematurely. Less common mutations produce proteins that do not use energy normally, do not allow chloride to cross the membrane appropriately, or are degraded at a faster rate than normal. Mutations may also lead to fewer copies of the CFTR protein being produced.



Even a single mutation (deletion) can be fatal

Cystic Fibrosis

The protein cannot be crystallized. No full resolution of the critical a.a. 508 region → simulations?

We expect numerical approaches to be difficult

- Which atoms are going to be bound?

  structure of the potential is not *a priori* known (QM)

- Force computation time grows like NxN

  two-body potential

- The system is very heterogeneous

  the problem is not “embarassingly” parallel

- Dynamics time step is of the order of a *femptosec*

  the system can be followed for very short times

- The system gets easily trapped in metastable states

  the exploration of the system phase-space is far from ergodic

- Energy may not be a good label of the states of the system

  states with largely different 3D-structures can have similar energies
  states with only slightly different 3D-structures can have very different energies

Countless number of approaches

- Geometrical approaches
- Simulated annealing
- Molecular Dynamics
- Monte Carlo simulations
- Simulated tempering and variations thereof
- Multi-canonical simulations
- Effective free-energy profile evaluation
- ...

Different levels of description

- Systems with discretized degrees of freedom
- String of beads
- Detailed atomistic description
  - with effective interaction potentials
  - with *ab initio* potentials
- ...

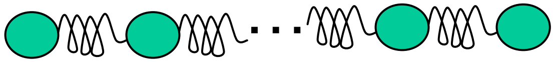
● Universality vs natural selection

Iori Marinari
Parisi Struglia

Self-interacting random hetero-polymers

♦ The complexity of the system is encoded in a certain amount of randomicity of the Hamiltonian

- $H = \sum_{i=1}^{N} \sum_{i>j} E_{ij}$, $N \geq 30$


beads

- $E_{ij} = k\delta_{i,j+1} r_{ij}^2 + \dfrac{B}{r_{ij}^{12}} + \dfrac{\eta_{ij} - A}{r_{ij}^6}$, $r_{ij}^2 = |\vec{x}_i - \vec{x}_j|^2$

binding repulsive $(B = 2)$ it depends on the sign of $\varepsilon - A$

- η_{ij} uncorrelated random gaussian variables

$$< \eta_{ij} > = 0 \qquad < \eta_{ij}^2 > = \varepsilon$$

♦ The system is brought to equilibrium at $\beta=1/k_B T$ under the Boltzmann probability distribution $\propto$ exp [-β*H*]

♦ During the evolution the shape of the chain is continuously monitored and various interesting features are revealed

Coil (open)	$\delta >> \rho, \lambda$
Unshaped globule	$\delta \geq \lambda$
Frozen well-shaped structures	$\delta < \rho, \lambda$
↪ ~ folding?	

configurations

- $\delta^2_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^{N} |\vec{x}_i^{(\alpha)} - \vec{x}_i^{(\beta)}|^2 \rightarrow$ "distance" between $\{\vec{x}_i^{(\alpha)}\}$ and $\{\vec{x}_i^{(\beta)}\}$

- $\rho = \frac{1}{N_{conf}}\sum_{\alpha} \frac{1}{N}\sum_{i=1}^{N} |\vec{x}_i^{(\alpha)} - < \vec{x}_i^{(\alpha)} > | \rightarrow$ average giration radius

- $\lambda = \frac{1}{N_{conf}}\sum_{\alpha} \frac{1}{N-1}\sum_{i=1}^{N-1} |\vec{x}_i^{(\alpha)} - \vec{x}_{i+1}^{(\alpha)} | \rightarrow$ average link length

I. $\varepsilon = 0$, no randomicity → homo-polymer

- phase transition at $A \approx 2$

coil (open) → **un**-shaped globule (closed)

$P(\delta^2)$ peaked at large δ^2 → small δ^2

II. $\varepsilon \neq 0$, some random interaction → hetero-polymer

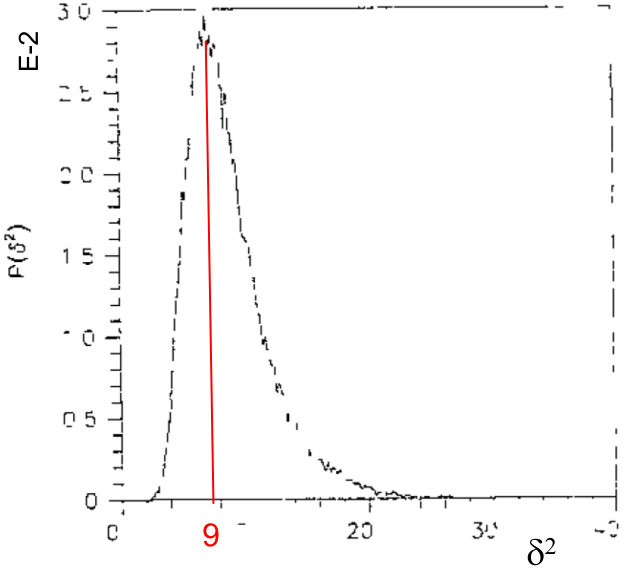
- new phase beyond a critical $\varepsilon_c > A$

well-shaped globule (~ glassy phase in SG ?)

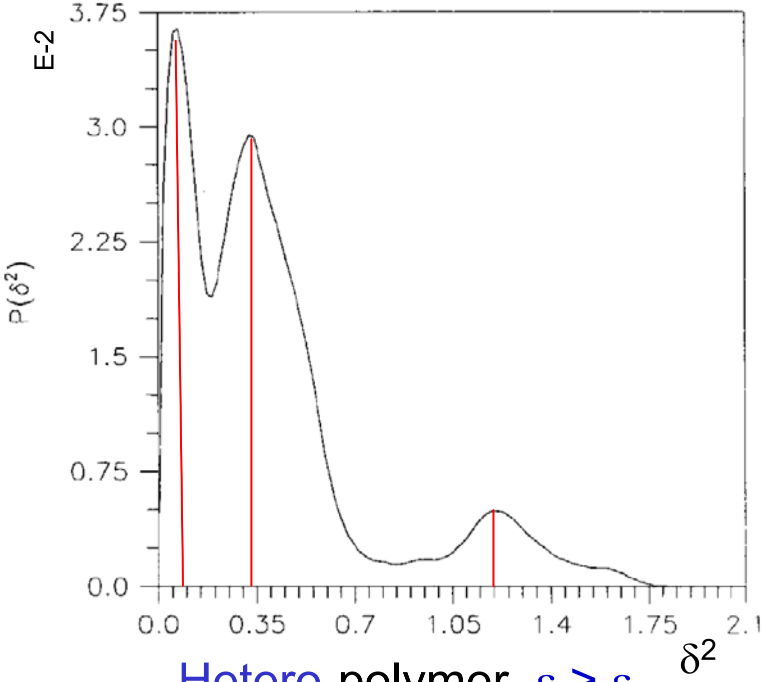
$P(\delta^2)$ is endowed with a lot of structure

Main result →
Sufficiently random hetero-polymers generically fold

Speculation → Perhaps (all the) other a.a. sequences do not fold.
Do they rather aggregate?



Homo-polymer, ε = 0
open phase A = 1.6

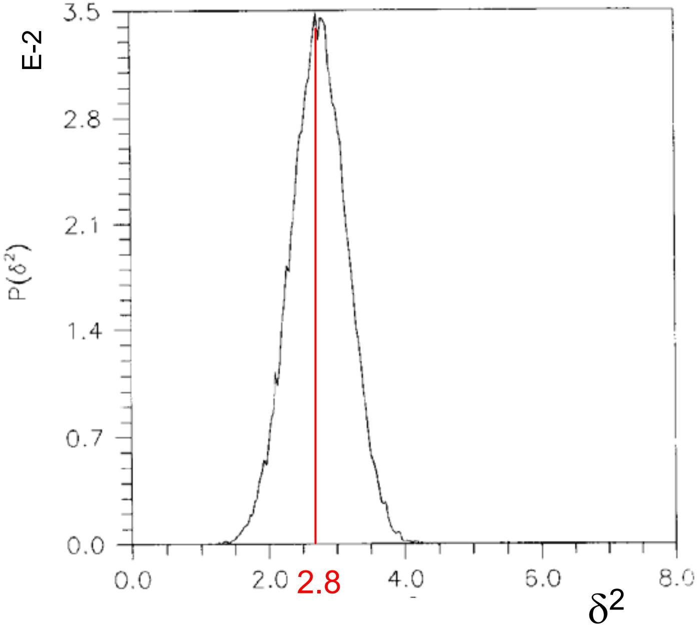


Hetero-polymer, $\varepsilon > \varepsilon_c$
“folded” phase

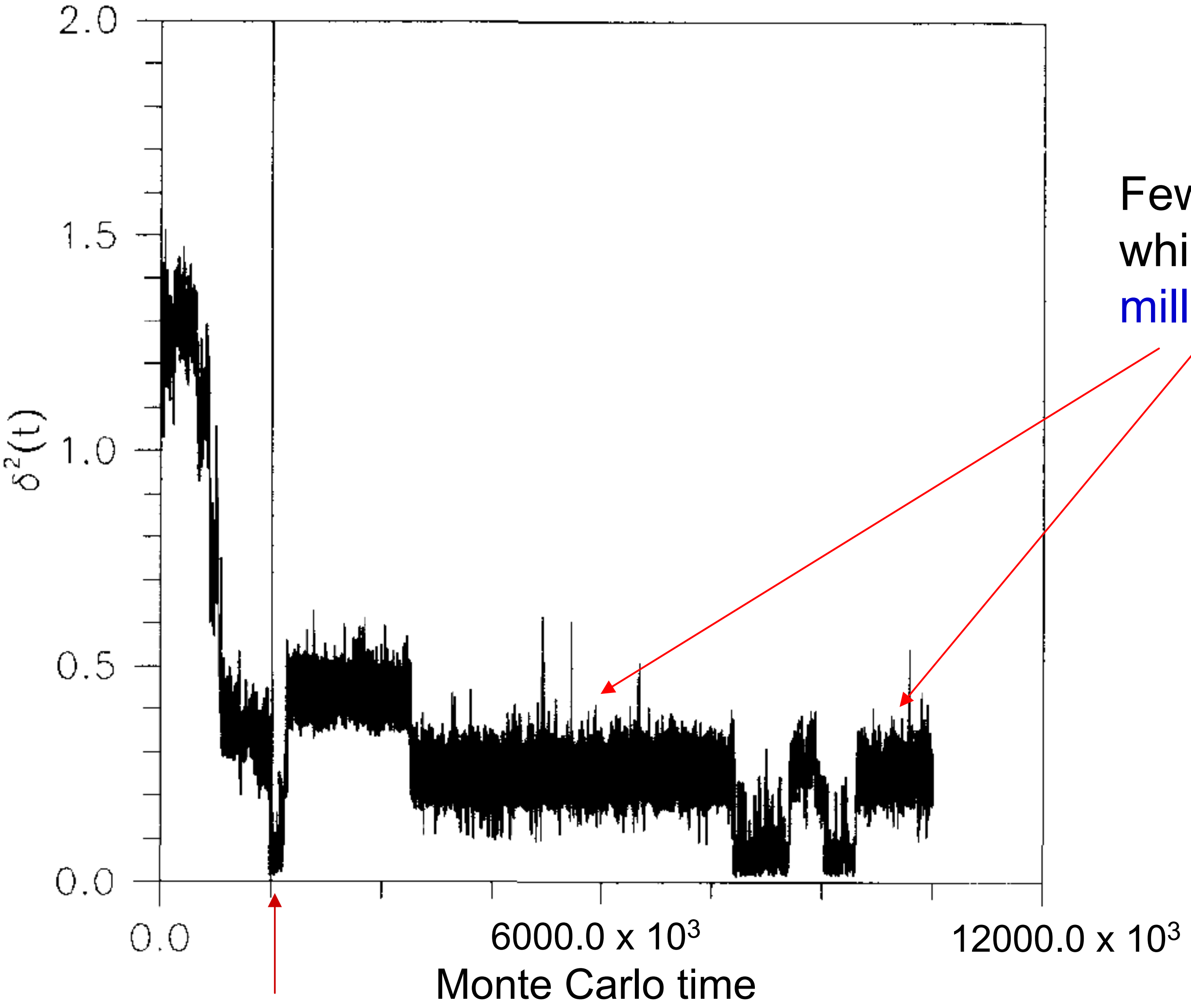
Peaks → macroscopically different folded states

Not δ-functions → many (only) microscopically different states

Macrostates are very long living (see next figures)

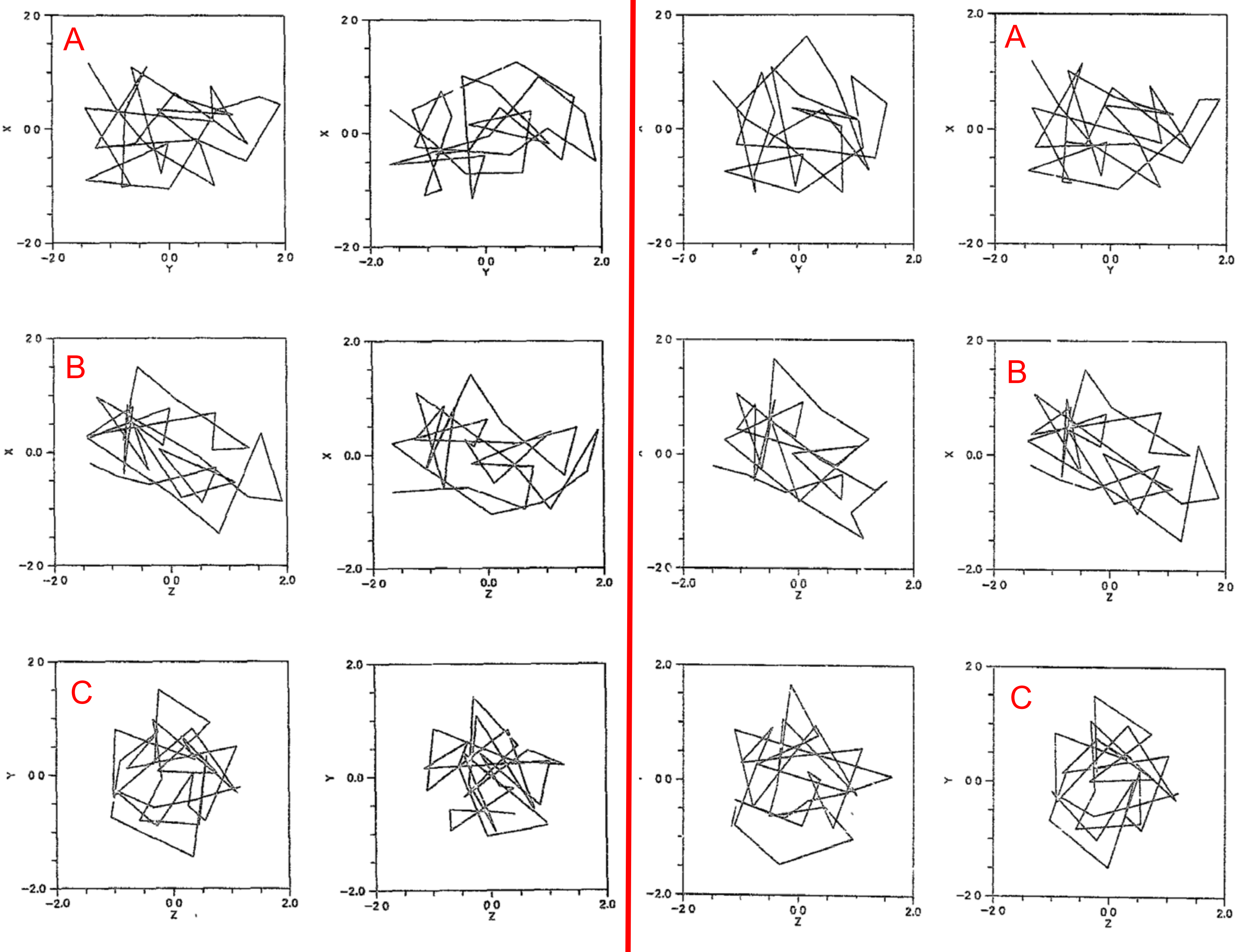


Homo-polymer, ε = 0
globular phase A = 3.8



Few metastable states which reappear after millions of MC steps

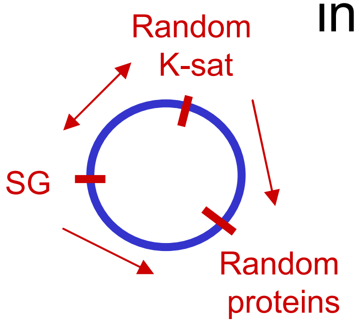
reference configuration


A
A
B
B
C
C
X
Y
Z
2.0
0.0
-2.0

Comments

- In the ”folded” phase the situation displays analogies with what one finds in the glassy phase of SG
  - Many long living, hierarchically organized states at sufficiently large randomicity (frustration)
  - Very long (actually not well defined) correlation times (stretched exponentials: $\propto \exp[-(t/\tau)^{\alpha}]$, $\alpha<1$, “aging”)
  - Complexity of protein folding is reflected in the NP-completeness of SG

- Can one make the SG analogy more stringent and useful?
  - Perhaps yes, taking inspiration from results in K-sat problem theory



- Random K-sat problems can be mapped to SG
- Alg’s borrowed from SG can help solving Random K-sat problems in polynomial time with probability ~1
- Can a random protein be folded in polynomial time?

- Should we instead move to a more reductionist point of view?

K-sat problems and SG

www.satlib.org

- **K-sat problem:** M constraints among N boolean variables, $p_1, p_2, \ldots, p_N$
- **Constraint:** clause among K variables (or their negation, $\neg$)

e.g. $(p_1 \vee \neg p_2) \wedge (p_2 \vee p_3) \wedge (\neg p_1 \vee \neg p_3) \rightarrow$
$[p_1 = t, p_2 = t, p_3 = f]$ or $[p_1 = f, p_2 = f, p_3 = t]$

Conjuntive Normal Form (CNF)

- $K \geq 3 \Rightarrow$ NP-complete problem

K-sat problems	Spin systems
- p_i = true/false	- spin $\Rightarrow \sigma_i$ +1/-1
- clause among a set of p_i	- interaction among a set of spins
- negated / non-negated variables	- coupling $J = -1$ / $+1$
- clauses satisfied / violated	- energy = 0 / 1
- # of violated clauses	- total energy H
- 2^N possible ansatz's	- $s = 1, 2, \ldots, 2^N$ possible states

$$P(\sigma,\beta) \propto \exp[-\beta H(\sigma)]$$

- minimal # of violated clauses	- minimum of H → SM at $\beta \rightarrow \infty$ $(T = 0)$

Random K-sat problem: building the a-th clause, C_a ($a = 1, 2, \ldots, M$)

1) $p_{i1}, p_{i2}, \ldots, p_{iK}$ ($K \geq 3$) are picked up with uniform probability among the N variables $p_1, p_2, \ldots, p_N$

2) variables $p_{i1}, p_{i2}, \ldots, p_{iK}$ are randomly negated

Spin Glass: building the a-th interaction term, E_a ($a = 1, 2, \ldots, M$)

1) $\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{iK}$ ($K \geq 3$) are picked up with uniform probability among the N variables $\sigma_1, \sigma_2, \ldots, \sigma_N$

2) coupling is $J_a = J_{i1}\, J_{i2} \ldots J_{iK}$ with $J_{ir} = -1$ or $J_{ir} = +1$, according to whether p_{ir} was randomly negated or not.

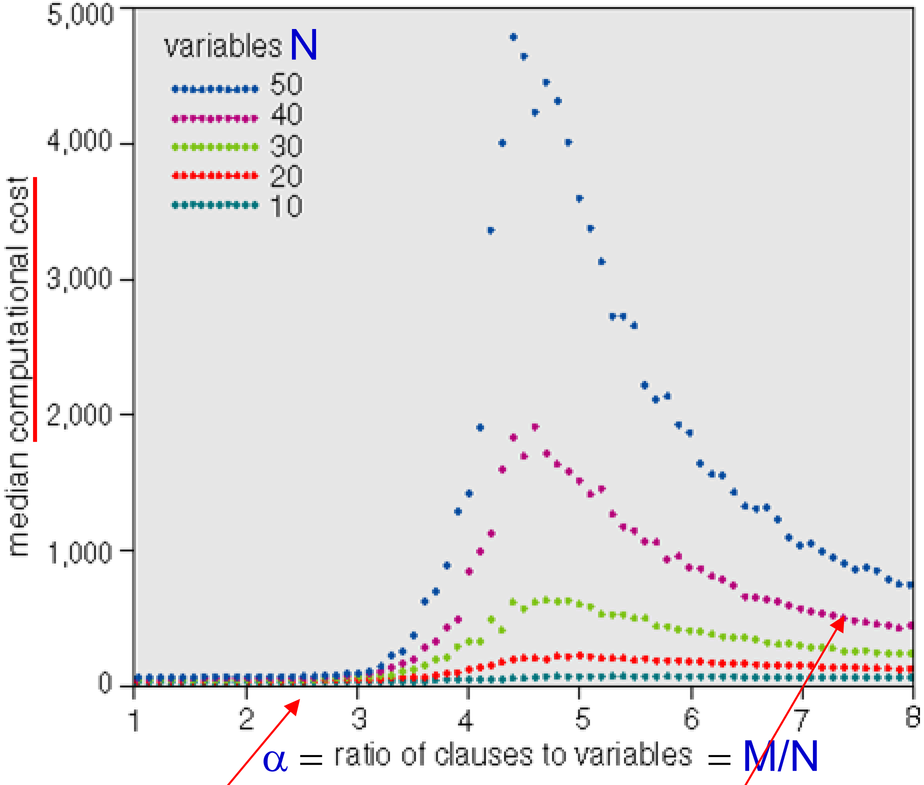
Mézard Monasson
Parisi Zecchina

Some interesting result

1) Emergence of a phase transition as $N \to \infty$, at a critical value of $\alpha = M/N$

2) Methods developed in SG theory can be used to solve hard K-sat problems (cavity method, decimation alg., …)

3) The average random (not the worst) case can be solved in polynomial time with probability ~ 1

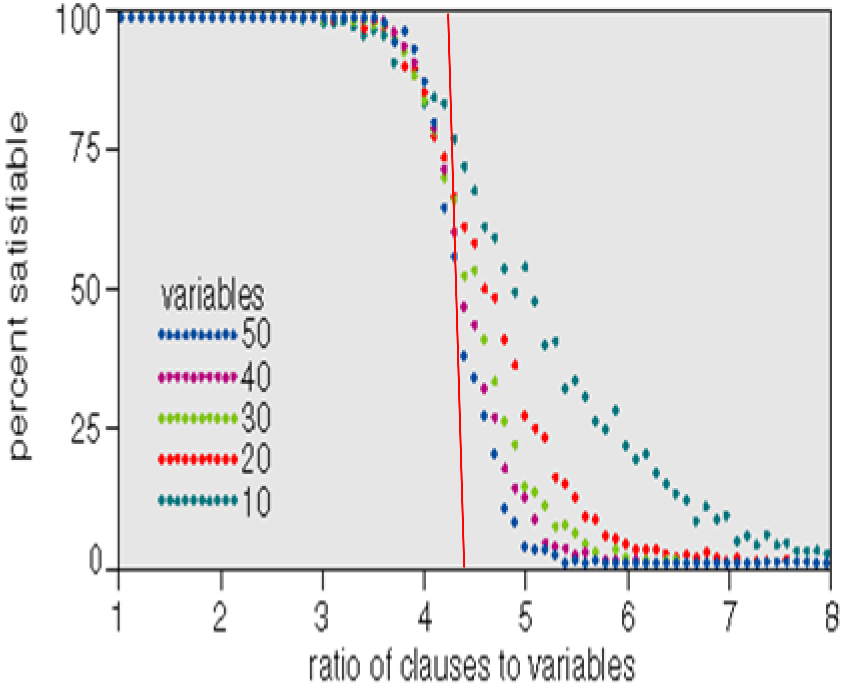


many variables
few clauses
under-constrained
⇓
SATisfiable

not so many variables
compared to # of clauses
over-constrained
⇓
UNSATisfiable

Mitchell Levesque Selman

Hardest problems
around $\alpha_c \approx 4.3$,
where SAT propositions
tend to become UNSAT



Phase transition →
the jump becomes sharper
as N gets larger

Protein folding and aggregation

- Generalities
- Universality *vs* natural selection

  the case of random hetero-polymers
- Folding *vs* aggregation

  the case of the Prion protein (PrP)

  the role of Cu
- XAS (NMR, EPR) experiments

  data analysis and EXAFS theory
- Q.M. simulations

  DFT and Car-Parrinello dynamics

● **Folding vs aggregation**

A test case: Prion Protein - PrP

(A bit of phenomenology)

- PrP is a cell membrane glycoprotein (highly expressed in the central nervous system of many mammals), whose physiological role is unclear
- It is, however, known to selectively bind copper, Cu
- Mature PrP has a flexible, disordered, N-terminal (23-120) and a globular C-terminal (121-231)
- **Misfolding** of PrP is held responsible for brain plaque formation and the development of Transmissible Spongiform Encelopathies (TSE)

● The N-terminal domain contains four (in humans) copies (repeats) of the octa-peptide PHGGGWGQ, each capable of binding Cu

● Experiments, more specifically, indicate that the Cu binding site is located within the HGGG tetra-peptide

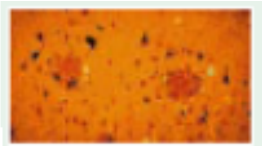
● Cu seems to play a crucial role

● Cries for (Car-Parrinello) *ab initio* simulations

Quantum Chemistry
in the BO approx

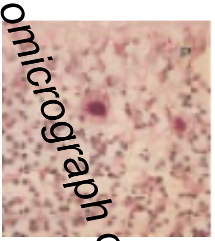
K. Wilson

- **Alzheimer's disease**

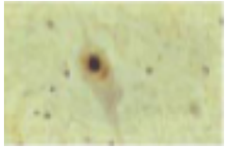


- **Transmissible Spongiform Encephalopaties (TSEs)**

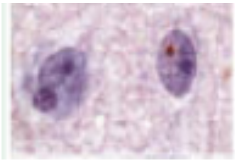
in humans: Creutzfeldt-Jakob Disease
sporadic
familial
iatrogenic
variant
in sheeps: Scrapie
in cattle: Bovine Spongiform Encephalopathy



- **Parkinson's disease; Dementia with Lewy bodies**



- **Amyotrophic Lateral Sclerosis**



- **Huntington's Disease**

post mortem photomicrograph of an histological section of the brain tissue

in vivo diagnosis by Positron Emission Tomography, PET

DISEASE	AGGREGATING PROTEINS
Alzheimer’s disease	Amyloid β-peptide
Transmissible Spongiform Encephalopathies	Full-length prion protein or fragments
Hereditary cerebral haemorrhage with amyloidosis	Amyloid β-peptide or Cystatin C
Parkinson’s disease; dementia with Lewy bodies	α-Synuclein
Frontotemporal dementia with parkinsonism	Tau
Type II diabetes	Amylin
Medullary carcinoma of the thyroid	Procalcitonin
Atrial amyloidoses	Atrial natriuretic factor
Amyotrophic lateral sclerosis	Superoxide dismutase
Huntington’s disease	Long glutamine stretches within proteins
Primary systemic amyloidosis	Intact immunoglobulin light chains or fragments
Secondary systemic amyloidosis	Fragments of serum amyloid A protein
Familial amyloidotic polyneuropathy 2	Fragments of apolipoprotein A1
Senile systemic amyloidosis	Wild-type transthyretin and fragments
Familial amyloidotic polyneuropathy 1	Mutant transthyretin and fragments
Familian Mediterranean fever	Fragments of serum amyloid A protein
Haemodialysis-related amyloidosis	β_2-Microglobulin
Finnish hereditary amyloidosis	Fragments of mutant gelsolin
Lysozyme amyloidosis	Full-length mutant lysozyme
Insulin-related amyloid	Full-length insulin
Fibrinogen α-chain amyloidosis	Fibrinogen α-chain variants

How do we go about such a complicated problem?

1) Hints from physiological/biological/biochemical data → **PrP accumulated data**

2) Make a working hypothesis and/or a model for misfolding or aggregation → **The role of Cu**

3) Test it against appropriately designed experiments → **EXAFS experiments**

4) Phenomenological interpretation of EXAFS data → **EXAFS theory**

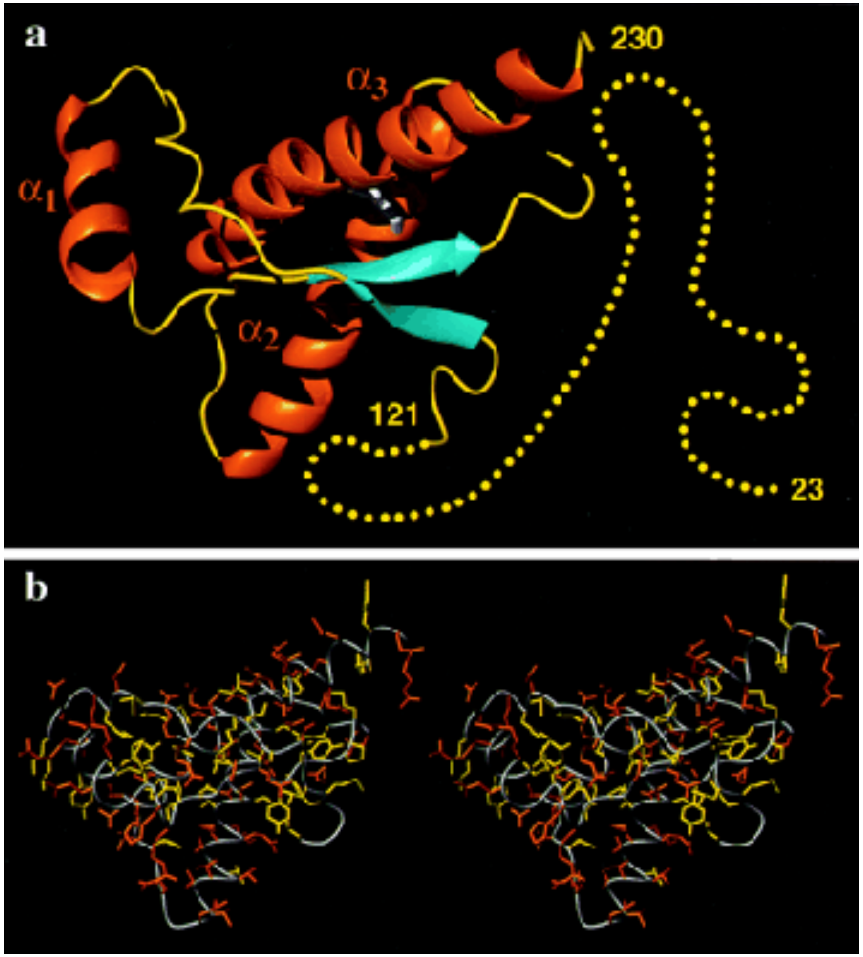
5) Go to an atomic description to check 4) and interpret the model → ***Ab initio* calculations**

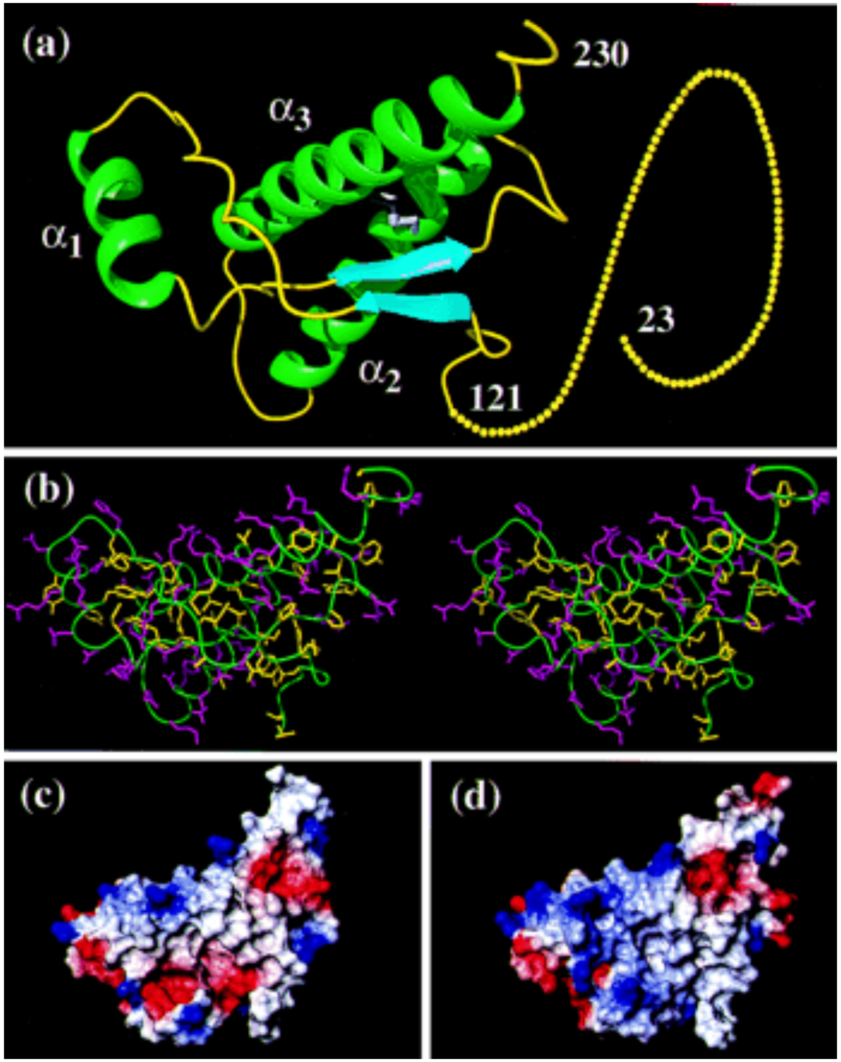
At this point, if you think you have understood something

6) Devise (?) an anti-aggregation strategy → **Test it *in vivo***

7) Most probably → **need to go back to**
2)

We start with some data

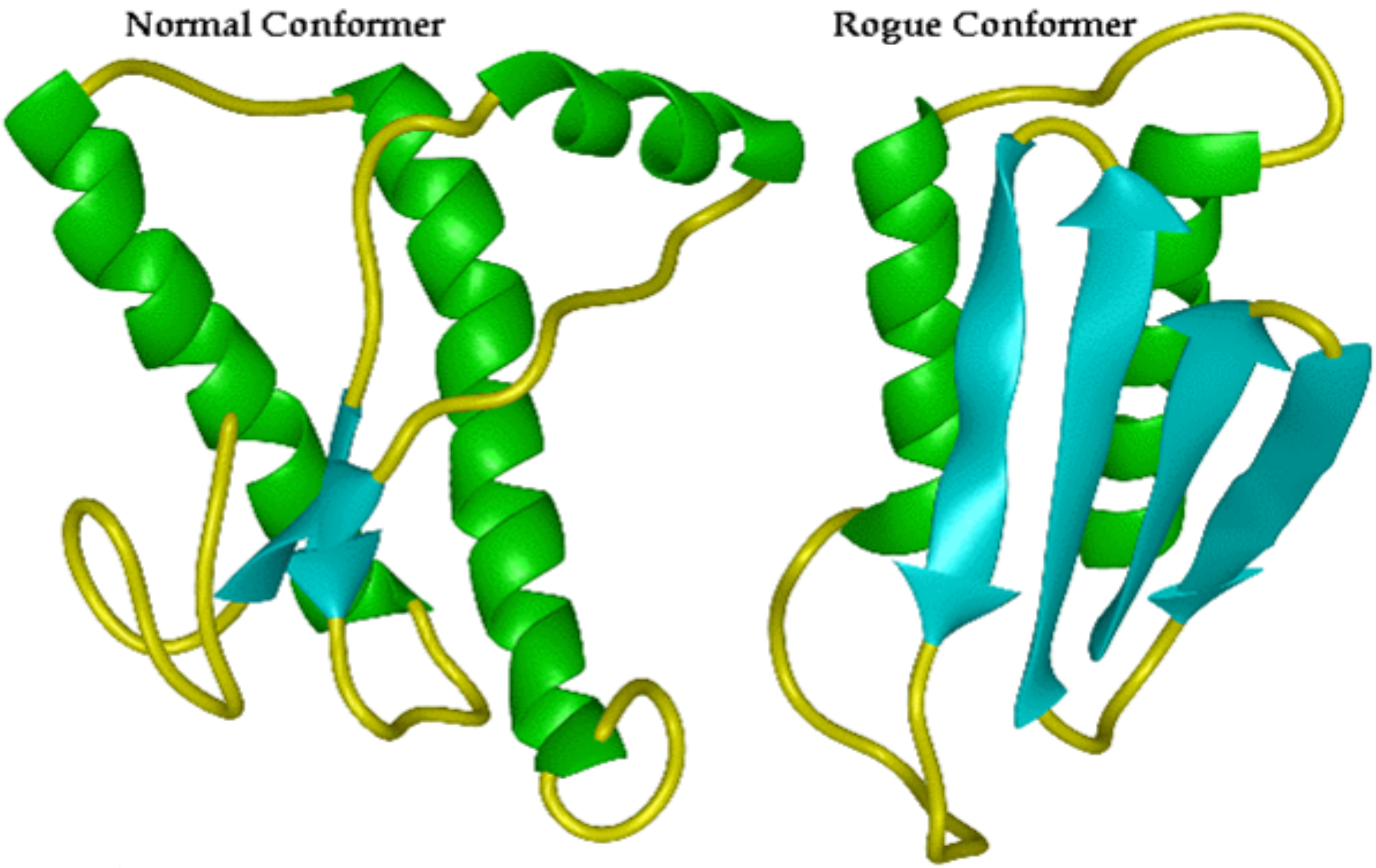




HuPrP (human)
α-helices = orange
β-strands = cyan
non-regular secondary structure = yellow,
flexible disordered "tail" (23-121) = yellow dots

BoPrP (bovine)
α-helices = green
β-strands = cyan,
non-regular secondary structure = yellow
flexible disordered "tail" (23-121) = yellow dots

C-terminal part

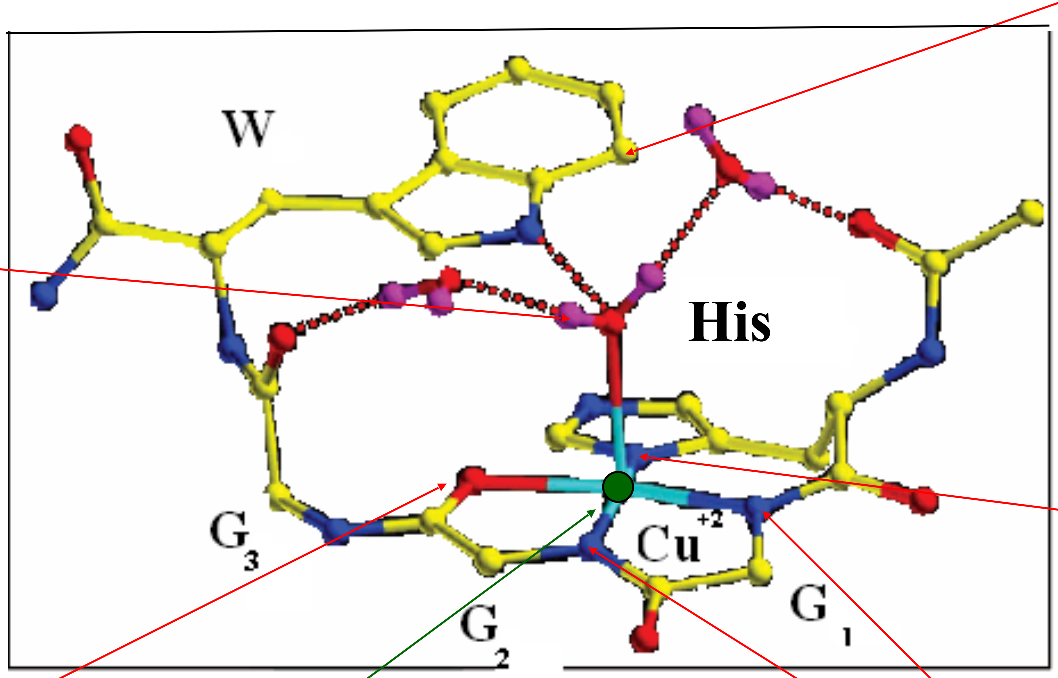


X-ray crystallography

Speculative

X-crystallography of the HGGGW-Cu^{+2} complex

Burns, et al. *Biochemistry* **41**:3991 (2002)



Trp side-chain parallel to the equatorial plane, possibly keeping water in site

$O(H_2O)$

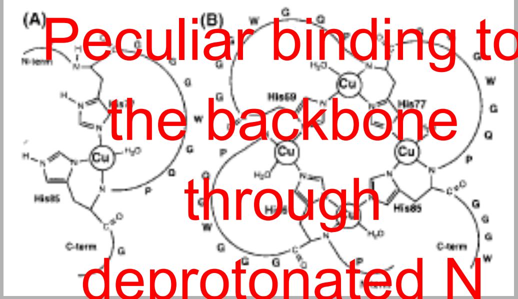
N_δ(His)

$O(G_2)$

Cu^{2+}

Peculiar binding to the backbone through deprotonated N from G_1 and G_2

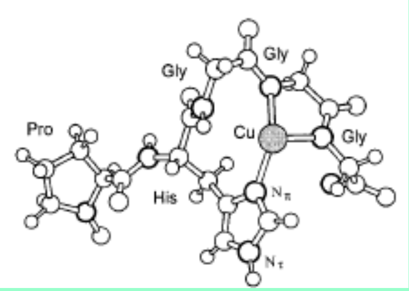
ESR, CD, and visible absorption spectra



Viles et. al. (1999) PNAS. 96: 2042

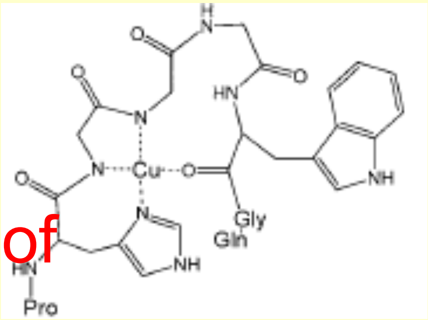
Peculiar binding to the backbone through deprotonated N from Gly$_2$ and Gly$_3$

Raman



Miura et al. Biochemistry (1999) 38:11560

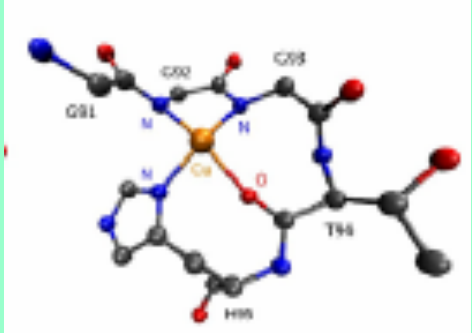
EPR



Aronoff-Spencer et al. Biochemistry (2000) 39: 13760.

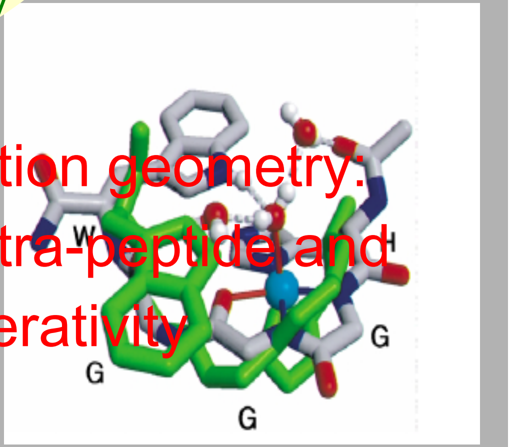
pH dependence of Cu binding

Ab initio computation on EPR data



Cox et. al. (2006) Bioph. J.

NMR



Cu coordination geometry: di-peptide, tetra-peptide and cooperativity

Zahn (2003) J. Mol. Biol. 334: 477

● EXAFS experiments

Synchrotron Radiation (SR)

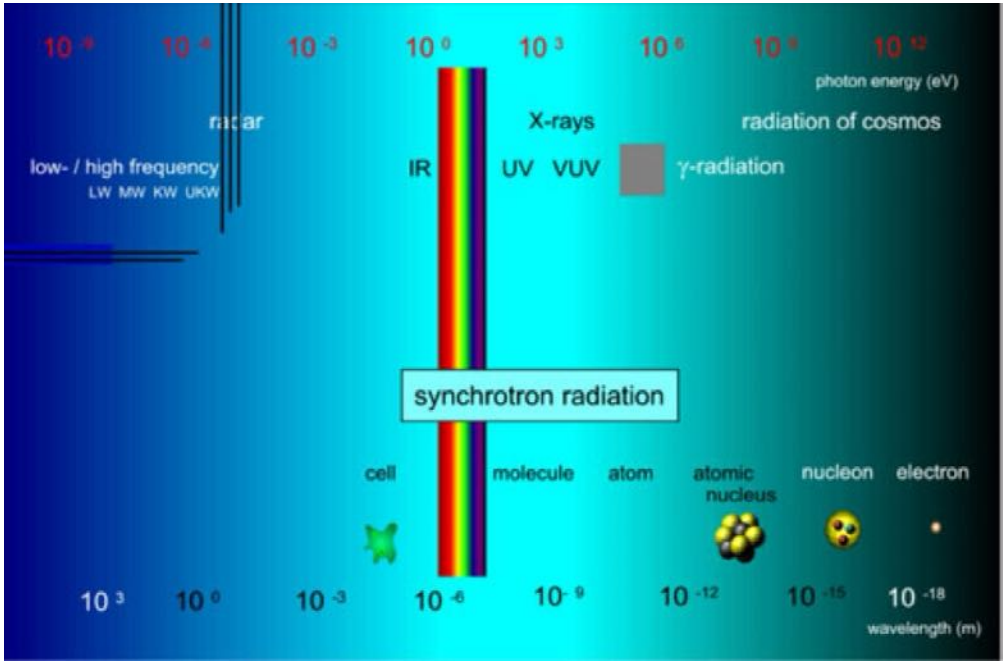
charged (electrons)
accelerated
relativistic ($E = \gamma\, mc^2$)

particles

radial acceleration
(e.g. deflection by a magnet)
Lorentz force
$\boldsymbol{F} = e\, \boldsymbol{v} \times \boldsymbol{B}$

SR is always emitted in the
forward direction
and is observed in a narrow cone
tangentially to the orbit

The higher the electron kinetic energy
the narrower the emission cone

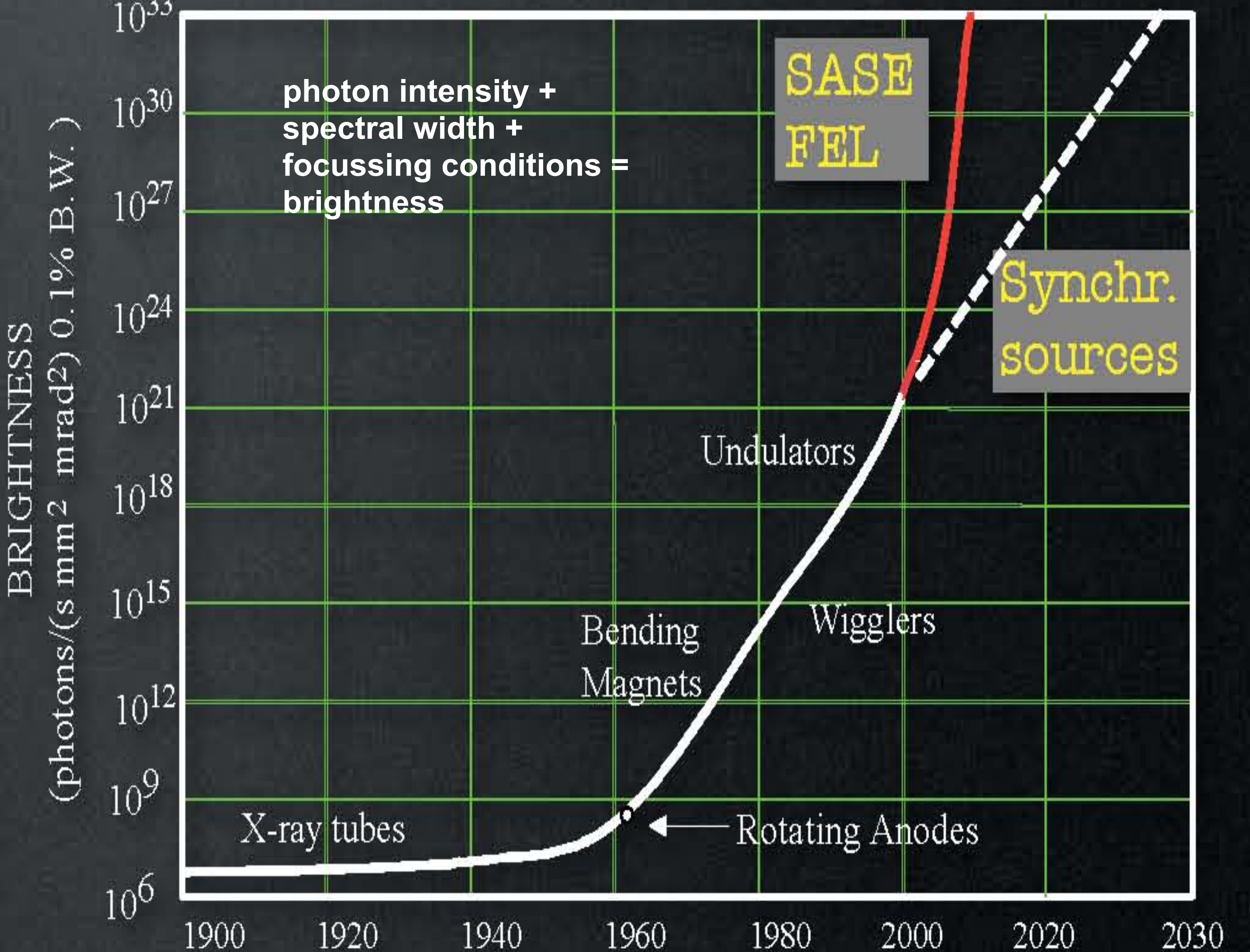


**SR spans the electromagnetic spectrum
from infrared (IR) to X-ray radiation**

Synchrotron Light Sources of the World

April 2005

1. Advanced Light Source (ALS), Berkeley, California
2. Advanced Photon Source (APS), Argonne, Illinois
3. ALBA Synchrotron Light Facilty (formerly Laboratorio de Luz Sincrotrón), Vallés, Spain
4. ANKA Synchrotron Strahlungsquelle, Karlsruhe, Germany
5. Australian Synchrotron, Melbourne, Victoria
6. Beijing Synchrotron Radiation Facility (BSRF), Beijing
7. Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY), Berlin
8. Canadian Light Source (CLS), Saskatoon, Saskatchewan
9. Center for Advanced Microstructures and Devices (CAMD), Baton Rouge, Louisiana
10. Center for Advanced Technology (INDUS-1 and INDUS-2), Indore, India
11. Cornell High Energy Synchrotron Source (CHESS), Ithaca, New York
12. diamond, Rutherford Appleton Laboratory, Didcot, England
13. Dortmund Electron Test Accelerator (DELTA), Dortmund, Germany
14. Electron Stretcher Accelerator (ELSA), Bonn, Germany
15. Elettra Synchrotron Light Source, Trieste, Italy
16. European Synchrotron Radiation Facility (ESRF), Grenoble, France
17. Hamburger Synchrotronstrahlungslabor (HASYLAB) at DESY, Hamburg, Germany
18. Institute for Storage Ring Facilities (ISA, ASTRID), Aarhus, Denmark
19. Laboratoire pour l'Utilisation du Rayonnement Electromagnétique (LURE), Orsay, France
20. Laboratório Nacional de Luz Síncrotron (LNLS) Sao Paolo, Brazil
21. MAX-lab, Lund, Sweden
22. National Synchrotron Light Source (NSLS), Brookhaven, New York
23. National Synchrotron Radiation Laboratory (NSRL), Hefei, China
24. National Synchrotron Radiation Research Center (NSRRC), Hsinchu, Taiwan, R.O.C
25. National Synchrotron Research Center (NSRC), Nakhon Ratchasima, Thailand
26. Photonics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST)
27. Photon Factory (PF) at KEK, Tsukuba, Japan
28. Pohang Accelerator Laboratory, Pohang, Korea
29. Shanghai Synchrotron Radiation Facility, (SSRF),
30. Siberian Synchrotron Radiation Centre (SSRC), Novosibirsk, Russia
31. Singapore Synchrotron Light Source (SSLS), Singapore
32. SOLEIL Synchrotron, Saint-Aubin, France
33. Stanford Synchrotron Radiation Laboratory (SSRL), Menlo Park, California
34. Super Photon Ring - 8 GeV (SPring8), Nishi-Harima, Japan
35. Swiss Light Source (SLS), Villigen, Switzerland
36. Synchrotron Radiation Center (SRC), Madison, Wisconsin
37. Synchrotron Radiation Source (SRS), Daresbury, U.K.
38. Synchrotron Ultraviolet Radiation Facilty (SURF III) at NIST, Gaithersburg, Maryland
39. UVSOR Facility, Okazaki, Japan
40. VSX Light Source, Kashiwa, Japan

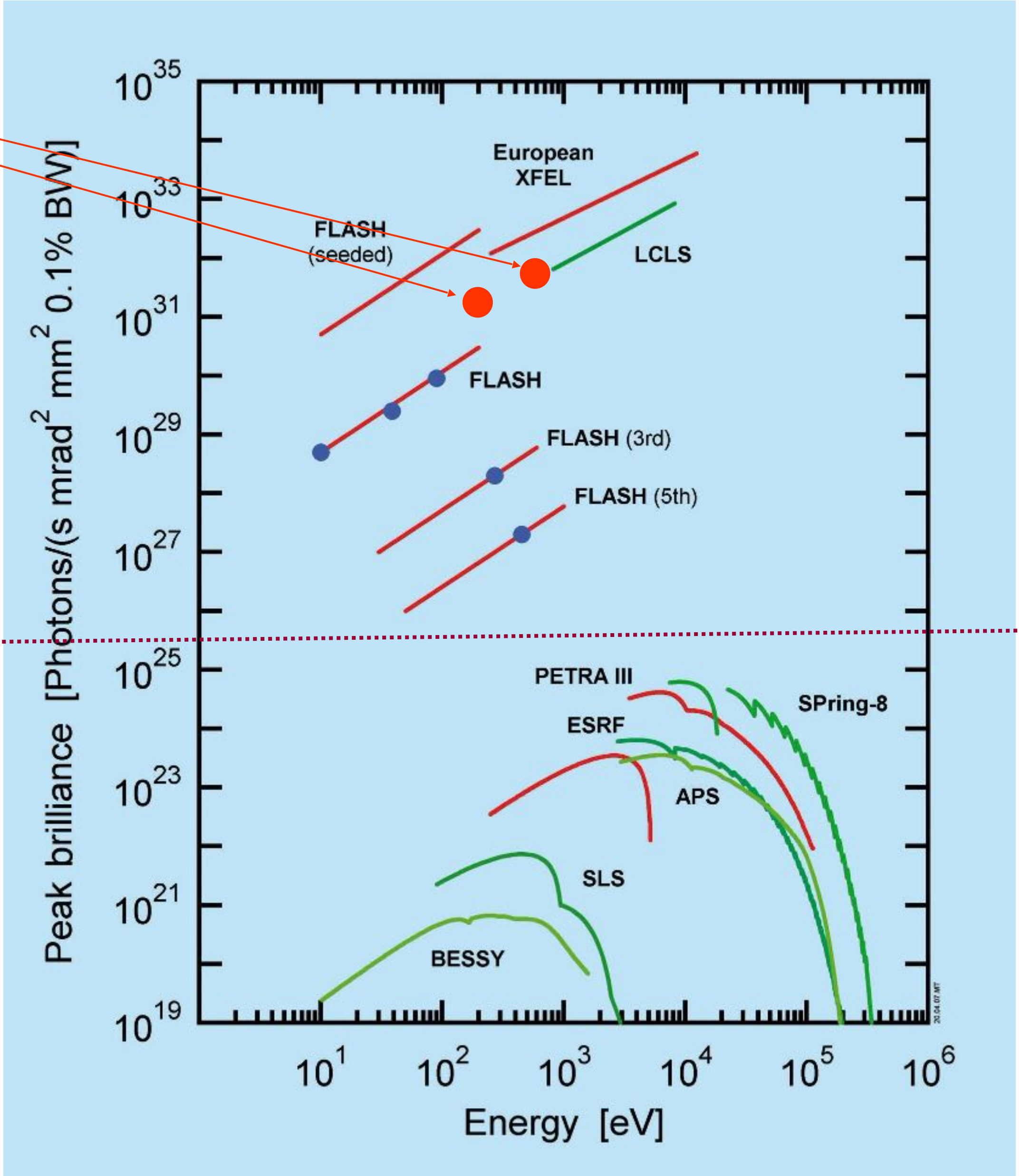


photon intensity +
spectral width +
focussing conditions =
brightness

Sparx

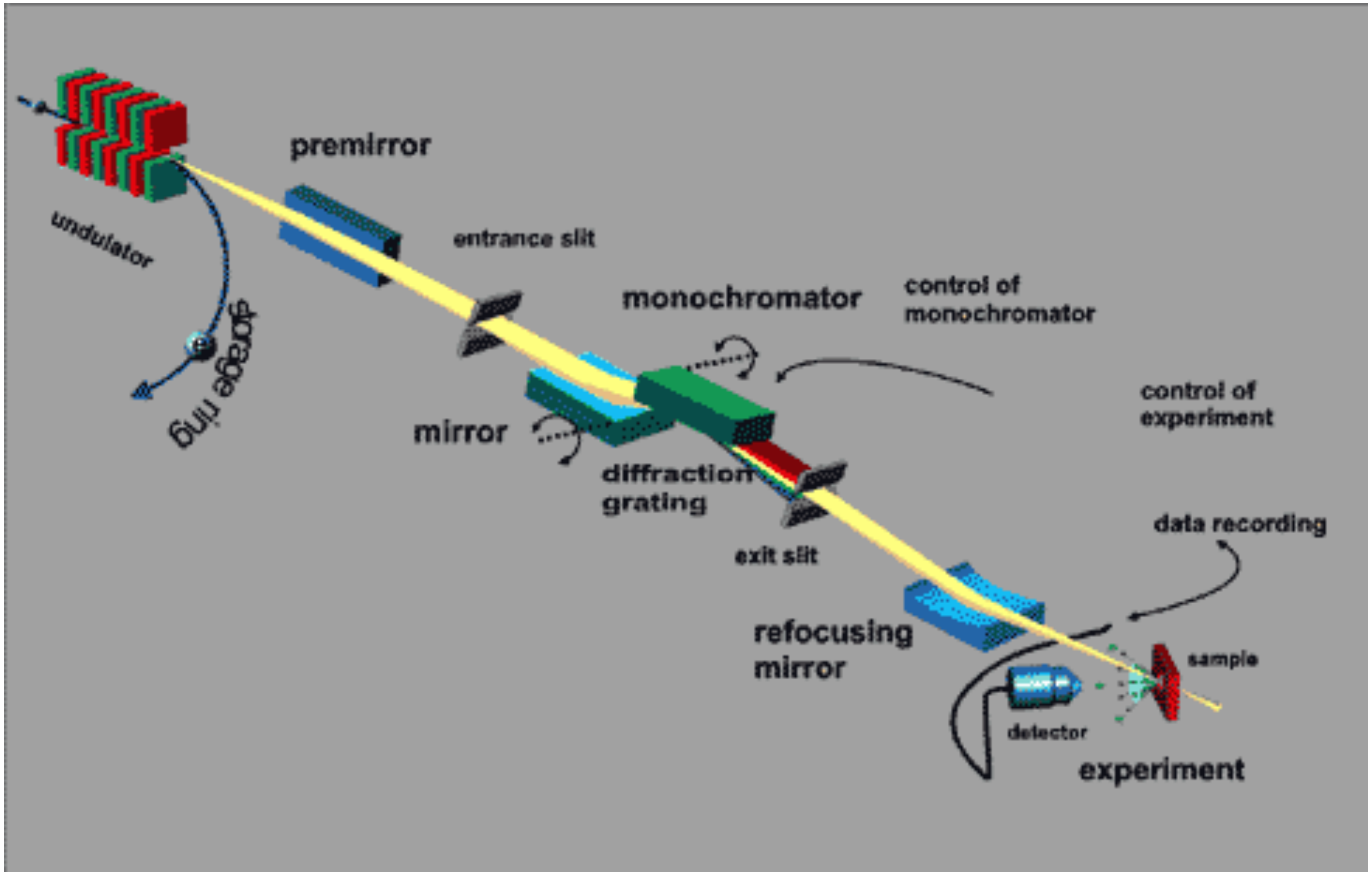
FEL Region

Synchrotron Radiation



Experimental setting

- radiation is directed by optical elements to the monochromator
- monochromator selects the desired wavelength of the spectrum
- the radiation is directed to the sample



Hard X-ray photons ⇒ $\lambda \sim$ inter-atomic distances in crystals

Radiation absorption ⇒ photo-electric effect $I = I_0 \exp[-\mu(E_k)d]$

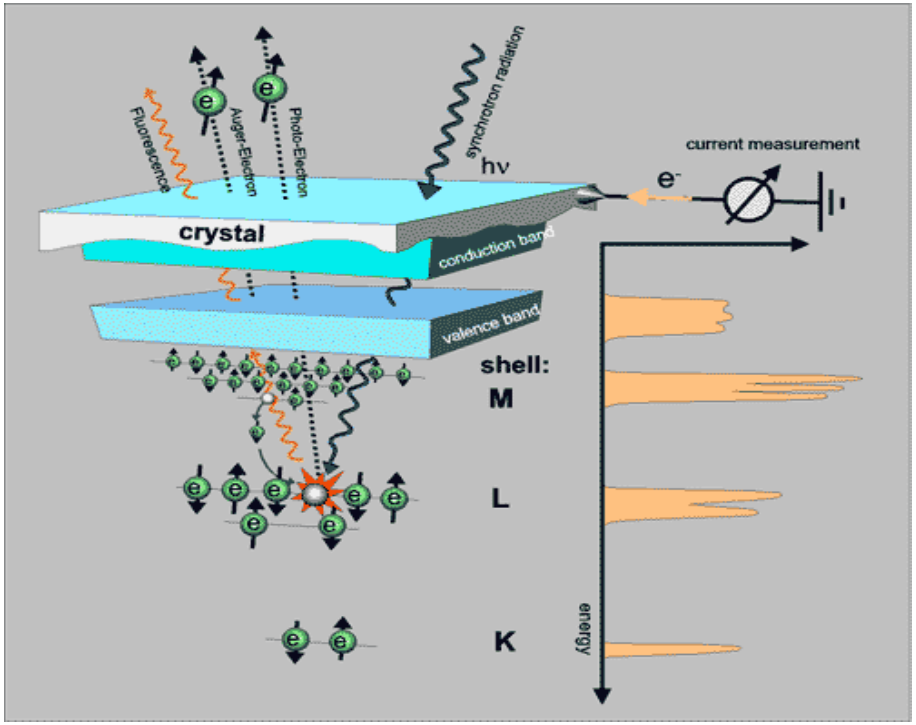
$$E_k = h\nu - E_0$$

E_k = kinetic energy of the emitted photo-electron

$h\nu$ = energy of the photon

E_0 = electron binding energy *

*characteristic of the specific material and bound state of the electron



extracted photo-electron leaves behind a positively charged state

⇓

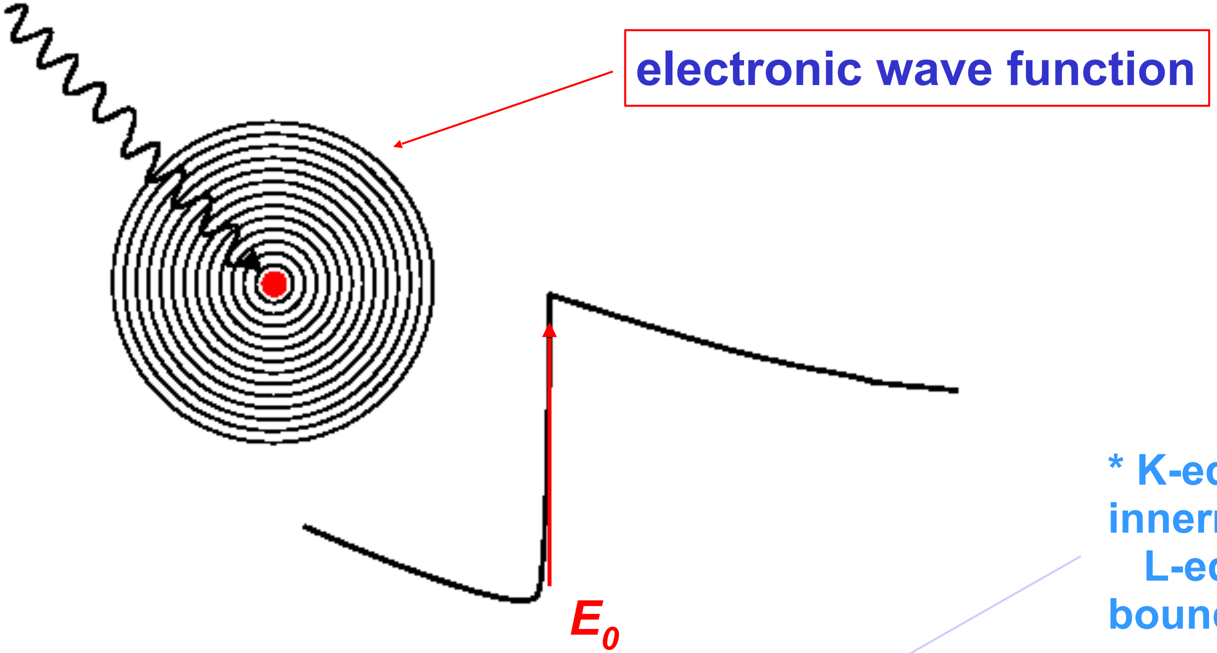
holes in the inner shell are filled by electrons from outer shells

⇓

emission of photons (fluorescence)

or further electrons from outer shells (Auger electrons)

XAS spectrum from an isolated atom (e.g. mono-atomic gas)



electronic wave function

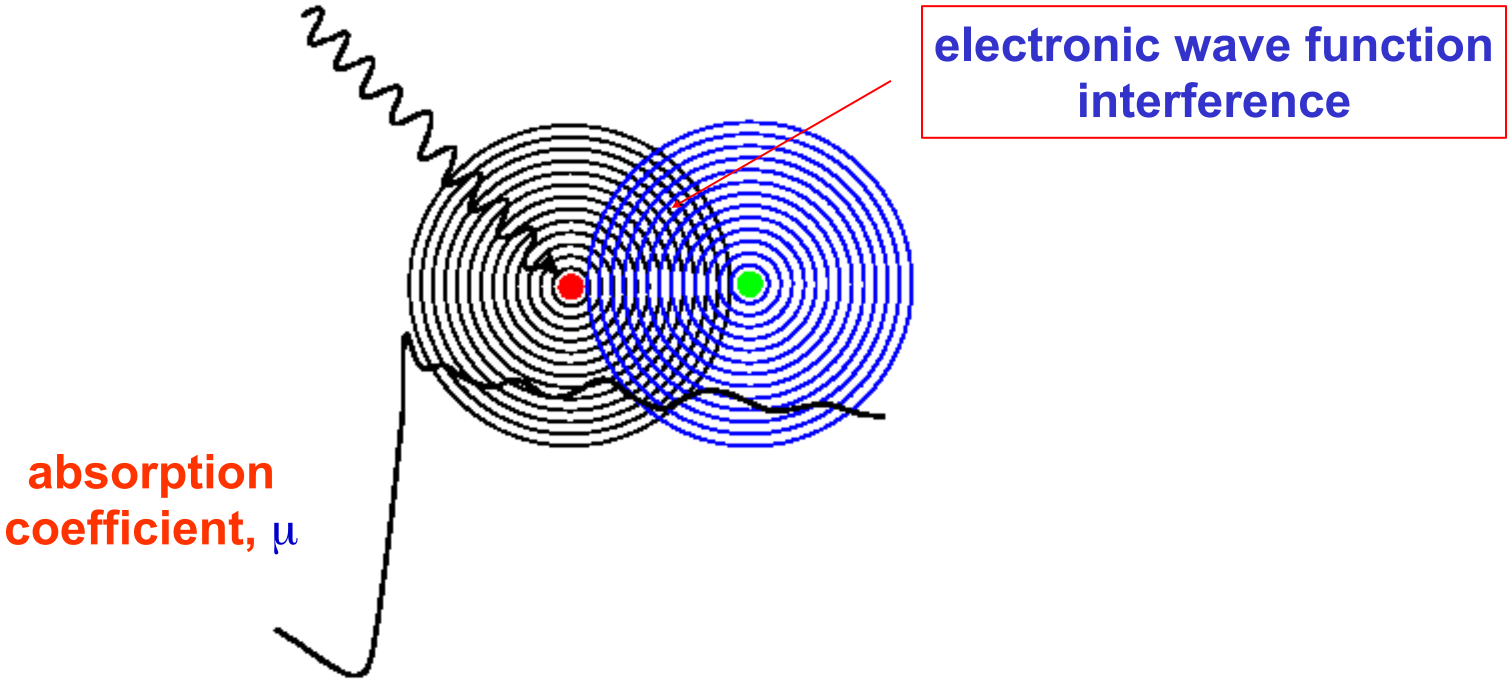
* K-edge: ionization of innermost electrons
L-edges: less strongly bound electrons

- The absorption coefficient, μ, decreases monotonically with the incident photon energy, *h*ν

- When *h*ν = E_0 = photo-ionization energy of an inner electron of the absorbing atom (edge energy*), μ sharply increases.

- It then decreases monotonically soon after the edge

XAS spectrum from a non-isolated atom (e.g. a diatomic molecule)



electronic wave function interference

absorption coefficient, μ

- In a multi-atomic system μ doesn't decrease monotonically after the edge, rather it has an oscillating behaviour

- The absorber (red dot) emits an outgoing spherical wave (the ionized electron, *photo-electron*)

- The scatterer (green dot) acts as diffusion center of the backscattered wave, which interferes (*in phase or out-of-phase*) with the outgoing wave

EXAFS analysis of Cu^{++} site geometry in Prion peptide complexes

(EMBL-DESY) Hamburg

S1. (BoPrP 25-30, 60-70) KKRPKPWGQPHGGGWGQ

S2. (BoPrP 25-30, 60-78) KKRPKPWGQPHGGGWGQPHGGGWGQ

S3. (BoPrP 25-30, 60-94) KKRPKPWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQ

S4. (αBoPrP 24-242) CKKRPKPGGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQ PHGGGWGQPHGGGWGQPHGGGWGQPHGGGGWGQGGTHGQWNKPSKPKTNMKHVAGAAAAG AVVGGLGGYMLGSAMSRPLIHFGSDYEDRYYRENMHRYPNQVYYRPVDQYSNQNNFVHDCVNITV KEHTVTTTTKGENFTETDIKMMERVVEQMCITQYQRESQAYYQRGAS

Cu^{++} stoichiometry

Sample	N	Cu^{++} equivalence	
		E	E/N
S1	1	0.5	0.5
S2	2	1.5	0.75
S3a	4	3.2	0.8
S3b	4	2.0	0.5
S4	4-5	2.0	0.5/0.4

N: number of Cu^{++} coordination sites in the complex = number of octarepeats

E: $[Cu^{++}]$ / [protein (or peptide)]

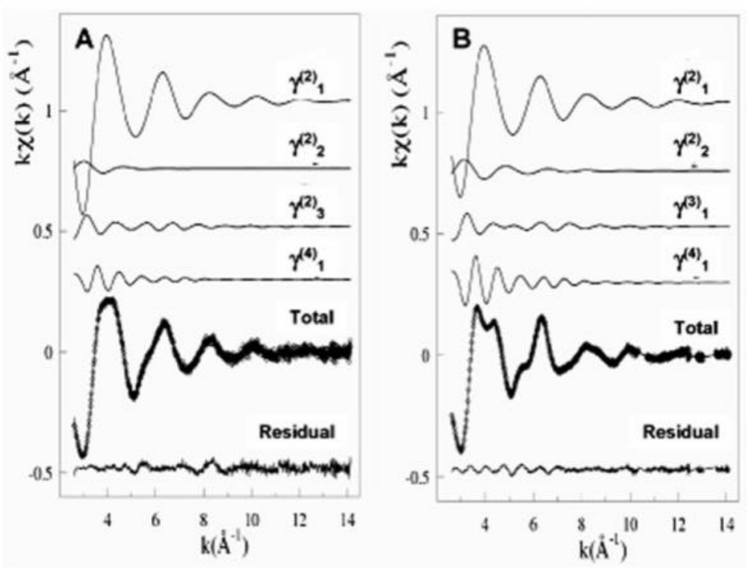
E/N: number of sites saturated with Cu^{++} = $[Cu^{++}]$ / [octarepeat]

← sub-stoichiometric Cu++ concentration

Morante et al., J. Biol. Chem. 279 (2004) 11753

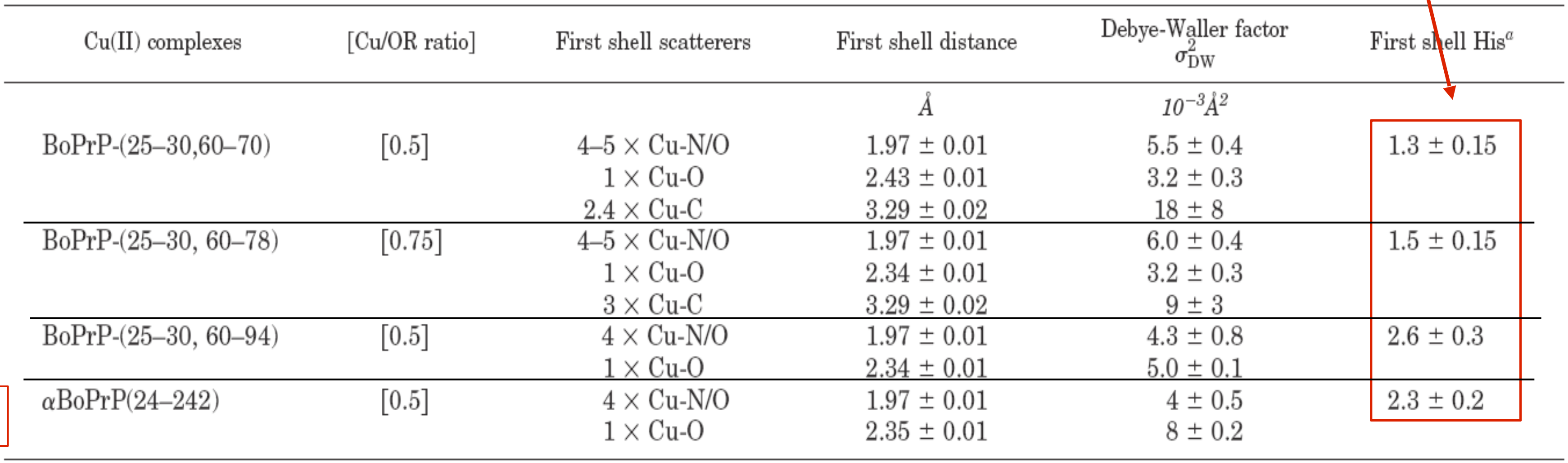
EXAFS data: Single and Multiple Scattering contributions
Fitted curves are within data fluctuations

S1



S4

His's are identified from their typical MS contributions

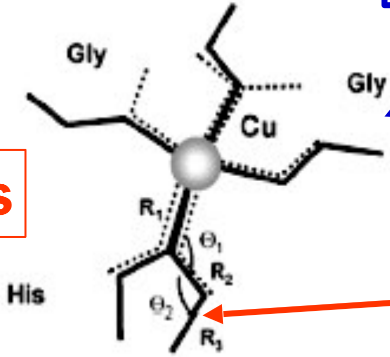


	Cu(II) complexes	[Cu/OR ratio]	First shell scatterers	First shell distance	Debye-Waller factor σ^2_{DW}	First shell His[a]
				Å	10^{-3}Å^2	
S1	BoPrP-(25–30,60–70)	[0.5]	4–5 × Cu-N/O 1 × Cu-O 2.4 × Cu-C	1.97 ± 0.01 2.43 ± 0.01 3.29 ± 0.02	5.5 ± 0.4 3.2 ± 0.3 18 ± 8	1.3 ± 0.15
S2	BoPrP-(25–30, 60–78)	[0.75]	4–5 × Cu-N/O 1 × Cu-O 3 × Cu-C	1.97 ± 0.01 2.34 ± 0.01 3.29 ± 0.02	6.0 ± 0.4 3.2 ± 0.3 9 ± 3	1.5 ± 0.15
S3	BoPrP-(25–30, 60–94)	[0.5]	4 × Cu-N/O 1 × Cu-O	1.97 ± 0.01 2.34 ± 0.01	4.3 ± 0.8 5.0 ± 0.1	2.6 ± 0.3
S4	αBoPrP(24–242)	[0.5]	4 × Cu-N/O 1 × Cu-O	1.97 ± 0.01 2.35 ± 0.01	4 ± 0.5 8 ± 0.2	2.3 ± 0.2

Model interpretation of EXAFS data analysis

M1

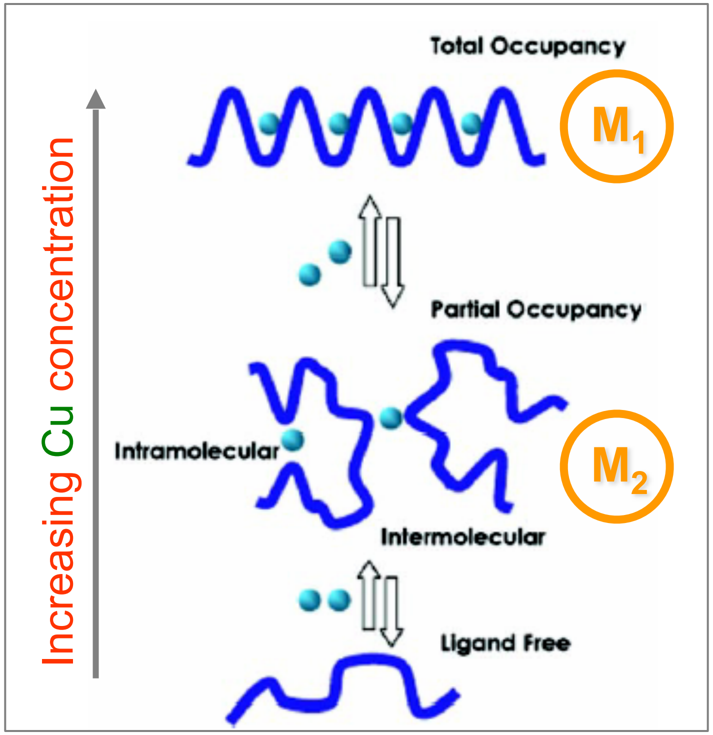
Cu : octarepeat = 1:1



1 His

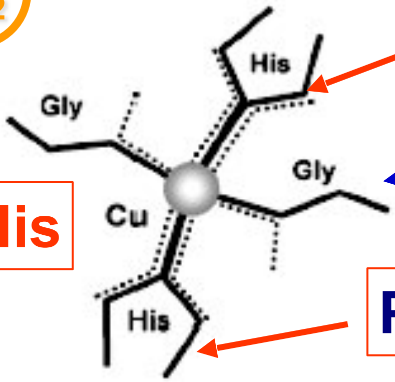
PHGGGWGQ

Intra-octarepeat



M2

PHGGGWGQ



2 His

Cu : octarepeat = 1:2

PHGGGWGQ

Inter-octarepeat

In the actual experimental situation no aggregates form

- EXAFS theory

Extracting structural information from EXAFS data

Data $I = I_0 \exp[-\mu(k)d]$ are expressed and analyzed in terms of

$$\chi(k) = \frac{\sigma_a - \sigma_0}{\sigma_0} = \frac{\mu(k) - \mu_0(k)}{\mu_0(k)} \qquad k = \frac{\sqrt{2m(h\nu - E_0)}}{\hbar}$$

$\mu =$ absorption coefficient

$\sigma_a =$ absorption cross section

$$\mu \propto \sigma_a$$

$$\sigma_a = 4\pi^2 \alpha \hbar \omega \left| \langle f | \hat{\varepsilon} \cdot \vec{r} | i \rangle \right|^2 N(\omega)$$

Fermi's golden rule

ω photon frequency

$N(\omega)$ density of photo-electron final state

$\hat{\varepsilon}$ polarization vector of incidente radiation

$M_{fi} = \left| \langle f | \hat{\varepsilon} \cdot \vec{r} | i \rangle \right|$ matrix element describing the electron transition

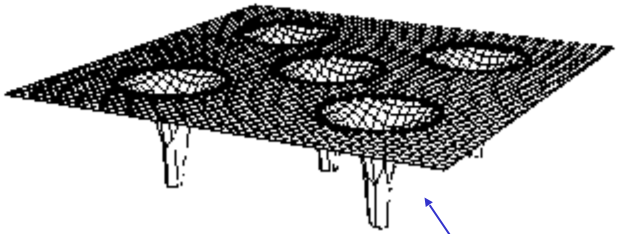
$|i>$: initial bound state

$|f>$: final "free" state

Initial, |*i*>, and the final states, |*f*>, are eigenfunctions of the Hamiltonian

$$H = \frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{r} + V(r)$$

The potential $V(r)$ is (most often) evaluated in the so-called ***muffin tin*** (**MT**) approximation



The **MT** potential consists of non overlapping spherical regions

In the interstitial regions the potential is set to a constant

Computing the transition matrix element, M_{fi}

❑ Electron initial state: one neglects ***V***

The Schrödinger equation for the innermost (K) electron is

$$\left(-\frac{\hbar^2}{2m}\nabla^2+\frac{Ze}{r}\right)|i\rangle=E|i\rangle$$

whose normalized solution is

$$\psi_i(r)=\langle r|i\rangle=\pi^{-1/2}\left(\frac{Z}{a_0}\right)^{3/2}\exp(-Zr/a_0)\equiv\psi_0(r)$$

n=1, ***l=0*** eigenfunction of hydrogen atom

❑ Electron final state: one neglects the Coulomb potential

The Schrödinger equation for the outgoing electron is

$$\left(-\frac{\hbar^2}{2m}\nabla^2+V\right)|f\rangle\equiv(H_0+V)|f\rangle=E|f\rangle$$

H_0 is the free Hamiltonian, ***V*** is the potential due to the presence of scatterers

Iterative solution: let's write $\left|f\right\rangle = \left|k\right\rangle + \left|rest\right\rangle$

$\left|k\right\rangle$ wave function of a free electron of momentum $\vec{k}$

$$\left\langle r\middle|k\right\rangle = N\exp\frac{i\vec{k}\cdot\vec{r}}{\hbar}$$

and satisfies the equation $\quad H_0\left|k\right\rangle = E\left|k\right\rangle \rightarrow (E-H_0)\left|k\right\rangle = 0$

we have $\quad (H_0+V)\left|f\right\rangle = E\left|f\right\rangle \rightarrow (E-H_0)\left|f\right\rangle = V\left|f\right\rangle$

inserting the definition of $\quad \left|f\right\rangle \Rightarrow$

- $(E-H_0)\left|k\right\rangle + (E-H_0)\left|rest\right\rangle = V\left|f\right\rangle \Rightarrow \left|rest\right\rangle = (E-H_0)^{-1}V\left|f\right\rangle$
- $\left|f\right\rangle = \left|k\right\rangle + (E-H_0)^{-1}V\left|f\right\rangle =$

$$= \left|k\right\rangle + (E-H_0)^{-1}V\left|k\right\rangle + (E-H_0)^{-1}V(E-H_0)^{-1}V\left|k\right\rangle + \ldots$$

Introducing the Green function $G_0 = (E-H_0)^{-1}$

where, we recall

$$(E-H_0)\left\langle\vec{r}\middle|G_0\middle|\vec{r}\,'\right\rangle = \delta^{(3)}(\vec{r}-\vec{r}\,') \qquad \left\langle\vec{r}\middle|G_0\middle|\vec{r}\,'\right\rangle = -\frac{m}{2\pi\hbar^2}\frac{e^{i\vec{k}(\vec{r}-\vec{r}\,')/\hbar}}{\left|\vec{r}-\vec{r}\,'\right|}$$

one obtains

$$M_{fi} = \left\langle f\middle|\hat{\varepsilon}\cdot\vec{r}\middle|i\right\rangle = \left\langle k\middle|\hat{\varepsilon}\cdot\vec{r}\middle|i\right\rangle + \left\langle k\middle|G_oV\hat{\varepsilon}\cdot\vec{r}\middle|i\right\rangle + \left\langle k\middle|G_oVG_oV\hat{\varepsilon}\cdot\vec{r}\middle|i\right\rangle + \ldots \equiv$$

$$\equiv A_0 + A_1 + A_2 + \ldots$$

- Stopping the expansion after the first term (single scattering events), one gets

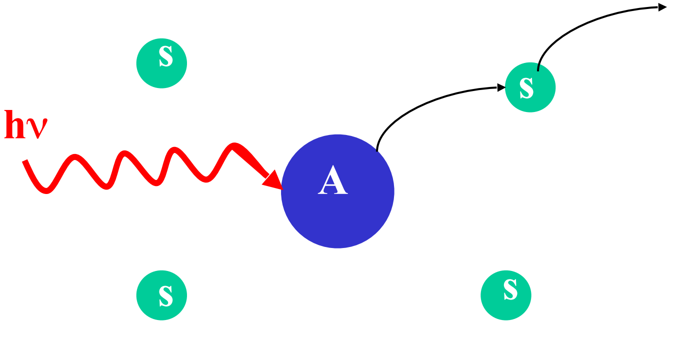
$$\left|\left\langle f\middle|\hat{\varepsilon}\cdot\vec{r}\middle|i\right\rangle\right|^2 = |A_0|^2 + \underbrace{|A_1|^2 + 2\,\mathrm{Re}\left(A_0A_1^*\right) + 2\,\mathrm{Re}\left(A_0A_2^*\right)}$$

atomic absorption contribution (isolated atom)

oscillations of the EXAFS signal

- Including further terms (multiple scattering events), one gets

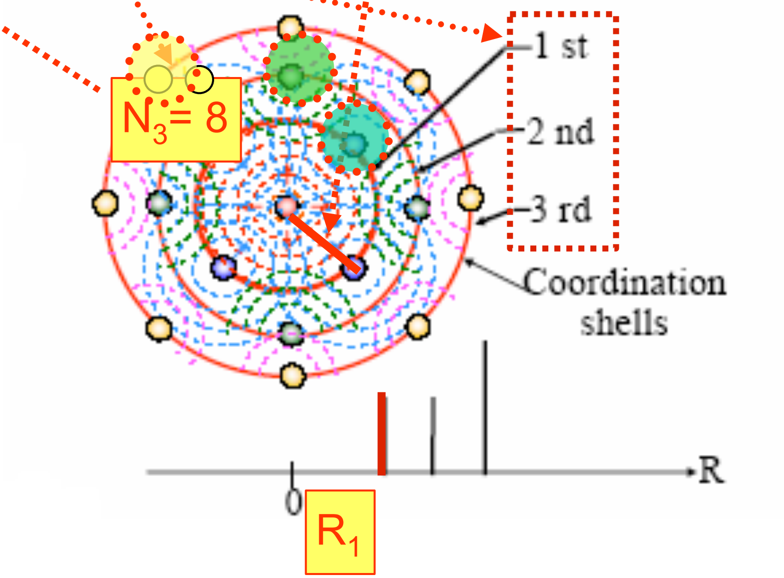
$$\sigma_a = \sigma_0 + \sum_i \sigma_i + \sum_{ij} \sigma_{ij} + \sum_{ijk} \sigma_{ijk} + \ldots$$



In Single Scattering approximation ($\hbar\omega >> E_0$)

Boland Crane Baldeschwieler, JPC 77, 142 (1982)

$$\chi(k) = S_0^2 \sum_i \frac{N_i}{kR_i^2} \left| f_i(k,\pi) \right| e^{-2k^2\sigma_i^2} e^{-2R_i/\lambda(k)} \sin\left[2kR_i + \Phi_i(k)\right]$$



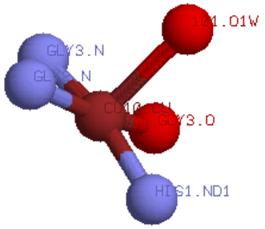
N_i, R_i, σ_i are fit parameters

Single scattering approximation

Information on type, number and mean distance of scatteres

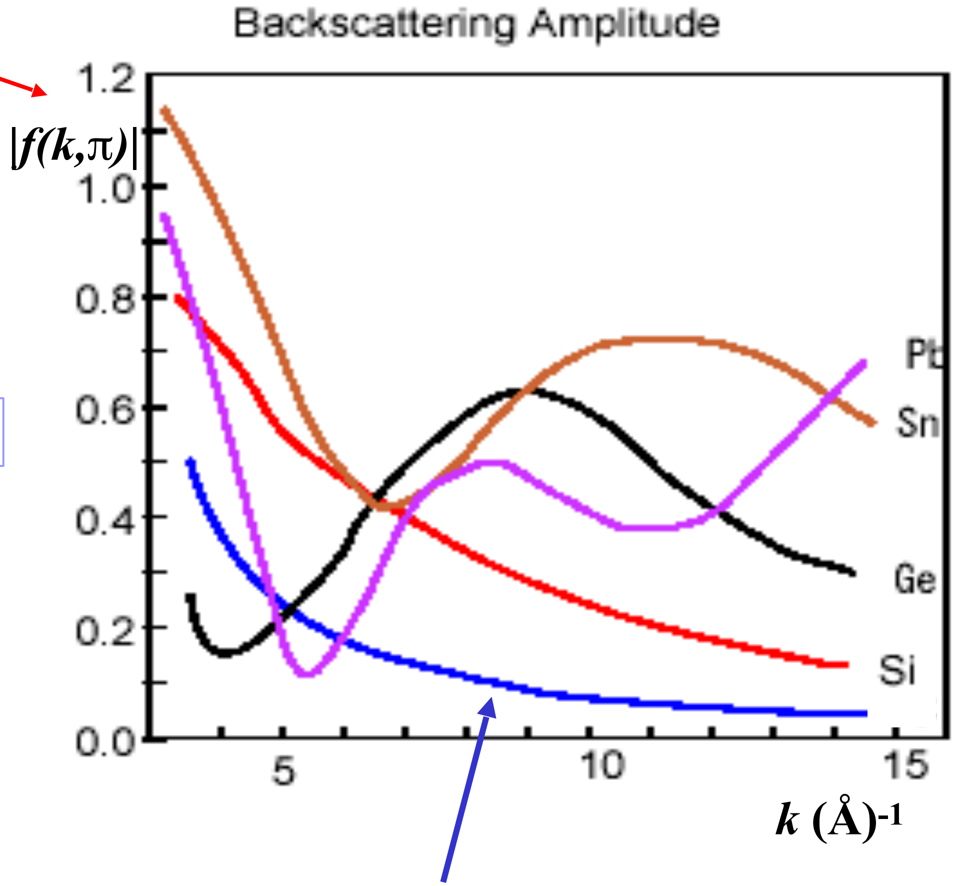
BUT

Copper ligands in the Prion peptide



● nitrogen

● oxigen



Need Multiple Scattering terms in σ_a to disentangle C,O and N contributions

Light atoms: C, O and N

Need to know

- the position of atoms in the vicinity of Cu, as the whole analysis of EXAFS data rests on this knowledge

- which are the actual metal ligands

- how the rest of the peptide is structurally organized

Ab initio calculations are necessary

- Quantum Mechanics to determine the atomic force field (in the Born-Oppenheimer approximation)

- Electrons are dealt with by DFT (Density Functional Theory)
  - Schroedinger equation is solved *à la* Kohn-Sham

- Atoms are treated classically

- Car-Parrinello simulations especially useful
  - Atomic Molecular Dynamics
  - Some dynamics helps in understanding stability

The DFT method

STEP 1

Decoupling of atomic and electronic dof’s (ν_A’s << ν_{el}’s ⇒ BOA)

STEP 2

At fixed atomic coordinates, compute the electronic ground-state w.f. with the help of DFT (Schroedinger eq ⇒ Kohn-Sham eq’s)

STEP 3

“Optimize” atomic coordinates to adapt them to the currently computed inter-atomic potential

STEP 4

Iterate STEP 2 and STEP 3 until you get consistency

The Car-Parrinello idea

Update atomic coordinates while solving Kohn-Sham eq’s

Faster convergence ⇒ CPU-time
Control over configuration stability
Atomic and KS eq’s are made to both look Newtonian (2nd order in time)

In Formulae

Starting point is the Schroedinger equation

electronic coordinates

atomic coordinates

- $$i\hbar \frac{\partial}{\partial t} \Phi[\{\vec{r}\}, \{\vec{R}\}; t] = H\, \Phi[\{\vec{r}\}, \{\vec{R}\}; t]$$

- $$H = -\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + V_A[\{\vec{R}\}] + H_e[\{\vec{r}\}, \{\vec{R}\}]$$
- $$V_A[\{\vec{R}\}] = \sum_{I<J} \frac{Z_I Z_J e^2}{|\vec{R}_I - \vec{R}_J|}$$
- $$H_e[\{\vec{r}\}, \{\vec{R}\}] = -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 + \sum_{i<j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} - \sum_{i,I} \frac{Z_I e^2}{|\vec{R}_I - \vec{r}_i|}$$

STEP 1

BO Molecular Dynamics

- $$M_I \frac{d^2\vec{R}_I(t)}{dt^2} = -\vec{\nabla}_I\left(\langle\Psi_0 \,|\, H_e \,|\, \Psi_0\rangle + V_A[\{\vec{R}\}]\,\right)$$

$$H_e \,|\, \Psi_0\rangle = E_0 \,|\, \Psi_0\rangle$$

$$E_0 = E_0[\{\vec{R}\}]$$

$$\langle\{\vec{r}\} \,|\, \Psi_0\rangle = \Psi_0[\{\vec{r}\}, \{\vec{R}\}]$$

Atoms move classically in the Quantum Mechanical potential generated by the electrons living in their ground-state w.f., Ψ_0

Difficulties

Schroedinger eq. should be solved over and over again at each atomic MD step

Contributions of excited states should be taken into account

One does not really know how to solve the electronic Schroedinger eq.

A useful approximation is Hartree-Fock

- Ψ_0 is written as a Slater determinant (Pauli principle) of N_e single particle trial w.f.'s, $\{\psi_i(r_i)\}$

- The latter are determined by minimizing the total electronic energy

$$\min_{\{\psi\}} \left\langle \Psi_0[\{\psi\}] \right| H_e^{HF} \left| \Psi_0[\{\psi\}] \right\rangle \Big|_{\langle \psi_i | \psi_j \rangle = \delta_{ij}} \qquad \left\langle \{\vec{r}\} \middle| \Psi_0[\{\psi\}] \right\rangle = \det_{ij}[\{\psi_i(\vec{r}_j)\}]$$

$$H_e^{HF} = -\frac{\hbar^2}{2m_e} \sum_i \nabla^2{}_i - \sum_{i,I} \frac{Z_I e^2}{|\vec{R}_I - \vec{r}_i|} + W^{dir}[\{\psi\}] + W^{exch}[\{\psi\}]$$

$$W^{dir}[\{\psi\}] = \sum_j \left[\int d\vec{r}\,' \psi_j^*(\vec{r}\,') \frac{1}{|\vec{r}\,' - \vec{r}|} \psi_j(\vec{r}\,') \right] \psi_i(\vec{r})$$

$$W^{exch}[\{\psi\}] = \sum_j \left[\int d\vec{r}\,' \psi_j^*(\vec{r}\,') \frac{1}{|\vec{r}\,' - \vec{r}|} \psi_i(\vec{r}\,') \right] \psi_j(\vec{r})$$

- H_e^{HF} is a one-body Hamiltonian

- which depends non-locally and non-linearly on all $\{\psi\}$

STEP 2

DFT → provides a way to systematically map the many-body problem (with electron self-interaction, W)

- $H_e\Psi[\{\vec{r}\}] = \left[T + W + U_A\right]\Psi[\{\vec{r}\}] =$

$$= \left[-\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 + \sum_{i<j}\frac{e^2}{|\vec{r}_i - \vec{r}_j|} - \sum_{i,I}\frac{Z_I e^2}{|\vec{R}_I - \vec{r}_i|}\right]\Psi[\{\vec{r}\}] = E[\{\vec{R}\}]\Psi[\{\vec{r}\}]$$

into a single-body problem (without electron self-interaction, W)

DFT → is based on the Hohenberg-Kohn (Phys. Rev **136** (1964) 864) →

Theorem

"There exists a one-to-one mapping between the set of U_A potentials and the set of (admissible) ground-state electronic densities"

Lemma 1

- $\{n\} \leftrightarrow \{U_A\}$ where $n(\vec{r}) = N\int \Pi_{i=1}^{N} d\vec{r}_i \Psi_0^*(\vec{r}, \vec{r}_2, \ldots, \vec{r}_N)\Psi_0(\vec{r}, \vec{r}_2, \ldots, \vec{r}_N)$

Since U_A fixes H_e → Ψ_0 is in turn a unique functional of n, hence of U_A

- $\{n\} \leftrightarrow \{U_A\} \leftrightarrow \{\Psi_0\}$ ← **HK mapping**

Lemma 2

- $F_{HK}[n] = \langle \Psi[n] | T + W | \Psi[n] \rangle$

is a well-defined, universal functional of the (admissible) electronic density

Lemma 3

The functional

- $E_{u_A}[n] = F_{HK}[n] + \int d\vec{r}\, u_A(\vec{r}) n(\vec{r}), \qquad u_A(\vec{r}) = \sum_I \frac{Z_I e^2}{|\vec{R}_I - \vec{r}|}$

1) attains its minimum when $n = n_{u_A}(\vec{r})$, i.e. when the electronic density equals the value which is in correspondence with U_A in the HK mapping

2) at the minimum it equals the total electronic energy

Corollary

We can compute the ground-state electronic density, hence all the ground-state observables, from the minimum equation

- $$\frac{\delta E_{u_A}[n]}{\delta n(\vec{r})} = 0 = \frac{\delta F_{HK}[n]}{\delta n(\vec{r})} + u_A(\vec{r}) \qquad (♥)$$

except that we do not know the HK-functional, F_{HK}

Kohn and Sham have proposed a way to go around this problem

The Kohn-Sham equations

- The key observation is that the HK mapping exists, even if we set the electronic self-interaction term to zero in all the above equations, $W \equiv 0$

  - in this situation the many-body electronic Schroedinger equation separates into N decoupled one-body equations

  - furthermore for any given electronic density, n, there exists a u_A^{NSI} such that one can represent n as the sum of the moduli square of the solutions of the one-body Schroedinger equation

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + u_A^{NSI}[n;\vec{r}]\right]\varphi_i(\vec{r}) = \varepsilon_i\varphi_i(\vec{r}) \qquad i = 1,2,\ldots,N$$

Kohn-Sham equations

n

$$n(\vec{r}) = \sum_{i=1}^{N} |\varphi_i(\vec{r})|^2 \qquad\qquad u_A \leftrightarrow n \leftrightarrow u_A^{NSI}$$

- We are done if we can find the relation between u_A and $u_A{}^{NSI}$
  - Ψ_0 is exactly the Slater determinant of the $\{\varphi_i\}$
  - the NSI HK-functional is simply the kinetic energy
    - $F_{HK}^{NSI}[n] \equiv T_{HK}^{NSI}[n] = \left\langle \Psi_0[n] \,|\, T \,|\, \Psi_0[n] \right\rangle = -\dfrac{\hbar^2}{2m_e} \sum_{1=1}^{N} \int d\vec{r}_i \varphi_i^*(\vec{r}_i) \nabla^2 \varphi_i(\vec{r}_i)$
  - and satisfies the equation
    - $\dfrac{\delta T_{HK}^{NSI}[n]}{\delta n(\vec{r})} + u_A^{NSI} = 0$ (♣)
- We now rewrite $E_{u_A}[n] = F_{HK}[n] + \int d\vec{r}\, u_A(\vec{r}) n(\vec{r})$ in the form
  - $E_{u_A}[n] = T_{HK}^{NSI}[n] + \int d\vec{r}\, u_A(\vec{r}) n(\vec{r}) + \dfrac{e^2}{2} \int d\vec{r} d\vec{r}\,' \dfrac{n(\vec{r}) n(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} + E^{exch}[n]$
  - $E^{exch}[n] \equiv F_{HK}[n] - T_{HK}^{NSI}[n] - \dfrac{e^2}{2} \int d\vec{r} d\vec{r}\,' \dfrac{n(\vec{r}) n(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}$
- Minimizing $E_{v_A}[n]$ and using equations (♥) and (♣), we get
  - $u_A^{NSI}(\vec{r}) = u_A(\vec{r}) - e^2 \int d\vec{r}\,' \dfrac{n(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} + \dfrac{\delta E^{exch}[n]}{\delta n(\vec{r})}$

- Inserting back u_A^{NSI} in the KS equations one ends up with

$$\bullet \left[-\frac{\hbar^2}{2m_e}\nabla^2 + u_A(\vec{r}) - e^2\int d\vec{r}\,'\frac{n(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + \boxed{\frac{\delta E^{exch}[n]}{\delta n(\vec{r})}}\right]\varphi_i(\vec{r}) = \varepsilon_i\varphi_i(\vec{r}) \qquad (\spadesuit)$$

- formally identical to the HF equations, but for
- ε_i are Lagrange multipliers enforcing $<\varphi_i\,|\,\varphi_j> = \delta_{ij}$

- On the solution the total energy reads

$$\bullet\ E_0^{HK} = \sum_{i=1}^{N}\varepsilon_i + \frac{e^2}{2}\int d\vec{r}\,d\vec{r}\,'\frac{n_0(\vec{r})n_0(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + E^{exch}[n_0] - \int d\vec{r}\,\left.\frac{\delta E^{exch}[n]}{\delta n(\vec{r})}\right|_{n_0} n_0(\vec{r})$$

- it is a function of the atomic positions
- it plays the role of inter-atomic potential in MD simulations

- We need an expression for $E^{exch}[n]$ and $\dfrac{\delta E^{exch}[n]}{\delta n(\vec{r})}$
  - for the Free Electron Gas

$$\bullet\ T_{FEG}[n] = \frac{3}{10}\int d\vec{r}\,(3\pi^2 n)^{2/3}n \qquad E_{FEG}^{exch}[n] = -\frac{3}{4\pi}\int d\vec{r}\,(3\pi^2 n)^{1/3}n$$

- LDA / GGA / … ⟶ $E_{LDA/GGA}^{exch}[n] = c\int d\vec{r}\;\eta_{LDA/GGA}^{exch}[n]\,n$

STEP 3 → STEP 4

"Optimization" of atomic coordinates can be achieved in various ways

1) Solve the classical eqs of motion

$$\bullet\ M_I \frac{d^2\vec{R}_I(t)}{dt^2} = -\vec{\nabla}_I \left(E^{HK}[\{\vec{R}\}] + V_A[\{\vec{R}\}]\right)$$

but, need to know $E^{HK}[\{R\}]$ for all values of $\{R\}$

2) Solve simultaneously classical eqs of motion for atoms and the KS eqs for electrons

Car-Parrinello

It can be elegantly done by introducing the effective Lagrangian

$$\bullet\ L_{CP} = \frac{1}{2}\sum_I M_I \left(\frac{dR_I}{dt}\right)^2 + \frac{1}{2}\sum_i \mu_i \int d\vec{r}\, \frac{d\varphi_i^*(\vec{r},t)}{dt}\,\frac{d\varphi_i(\vec{r},t)}{dt} +$$

$$+\sum_{I<J}\frac{Z_I Z_J e^2}{|\vec{R}_I - \vec{R}_J|} - E^{HK}[\{\varphi\},\{\vec{R}\}] + \sum_i \Lambda_{ij}\int d\vec{r}\,\varphi_i^*(\vec{r},t)\varphi_i(\vec{r},t)$$

$\{R\}$ and $\{\varphi\}$ are Lagrangian coordinates, $n(r) = \Sigma_i\, \varphi_i^*(r)\, \varphi_i(r)$ and

$$\bullet\ E^{HK}[\{\varphi\},\{\vec{R}\}] = \sum_{i=1}^{N} \varepsilon_i + \frac{e^2}{2}\int d\vec{r}\,d\vec{r}\,'\,\frac{n(\vec{r})n(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} + E^{exch}[n] - \int d\vec{r}\,\frac{\delta E^{exch}[n]}{\delta n(\vec{r})}\,n(\vec{r})$$

- Rather than the minimum equation (♠), we get for the electronic w.f., the 2nd order equation in the (fictitious) time

$\boxed{0 =}$

ε_i eigenvalues of Λ_{ij}

$$\mu_i \frac{d^2\varphi_i(\vec{r},t)}{dt^2} = \left[-\frac{\hbar^2}{2m_e}\nabla^2 + u_A(\vec{r}) - e^2\int d\vec{r}'\,\frac{n(\vec{r}',t)}{|\vec{r}-\vec{r}'|} + \frac{\delta E^{exch}[n]}{\delta n(\vec{r})}\right]\varphi_i(\vec{r},t) - \Lambda_{ij}\varphi_j(\vec{r},t)$$

- A unique time step for atomic MD and KS eqs, $\Delta t \approx$ femtosecond

- We need to solve the KS eqs by adiabatically lowering the electronic “kinetic energy”
  - “total electronic energy” is (almost) conserved
    we have a Lagrangian system
    little energy transfer between atoms and electrons
  - by progressively lowering $T_e \rightarrow 0$, the system will end in the minimum of the “potential”
  - where the force, hence the acceleration is zero

CP dynamics is implemented in a number of codes, among which Quantum ESPRESSO and CPMD

http://www.quantum-espresso.org/ http://www.cpmd.org/

- Quantum ESPRESSO is an initiative of the DEMOCRITOS National Simulation Center (Trieste) and of its partners.
- In collaboration with
  - CINECA, the Italian National Supercomputing Center in Bologna
  - Ecole Polytechnique Fédérale de Lausanne
  - Princeton University
  - Massachusetts Institute of Technology
  - Many other individuals…
- Integrated computer code suite for electronic structure calculations and materials modeling at the nanoscale
  - Released under a free license (GNU GPL)
  - Written in Fortran 90, with a modern approach
  - Efficient, Parallelized (MPI), Portable
- Suite components
  - PWscf (Trieste, Lausanne, Pisa): self-consistent electronic structure, structural relaxation, BO molecular dynamics, linear-response (phonons, dielectric properties)
  - CP (Lausanne, Princeton): (variable-cell) Car-Parrinello molecular dynamics

The Quantum-ESPRESSO Software Distribution

- Car-Parrinello variable-cell molecular dynamics with Ultrasoft PP’s.
- Developed by A. Pasquarello, K. Laasonen, A. Trave, R. Car, N. Marzari, P. Giannozzi, C. Cavazzoni, G. Ballabio, S. Scandolo, G. Chiarotti, P. Focher.
- Verlet dynamics with mass preconditioning
- Temperature control: Nosé thermostat for both electrons and ions, velocity rescaling
- Variable-cell (Parrinello-Rahman) dynamics
- Damped dynamics minimization for electronic and ionic minimization
- Modified kinetic functional for constant-pressure calculations
- “Grid Box” for fast treatment of augmentation terms in Ultrasoft PP’s
- Metallic systems: variable-occupancy dynamics
- Nudged Elastic Band (NEB) for energy barriers and reaction paths
- Dynamics with Wannier functions

A *first principle* study
of the Cu-HGGG interactions
A - the monomer
B - the dimer

Furlan, La Penna, Guerrieri, Morante, GCR, JBIC **12(4)** (2007) 571

A - Initial Cu$^{(+2)}$ HG$^{(-)}$G$^{(-)}$G configuration

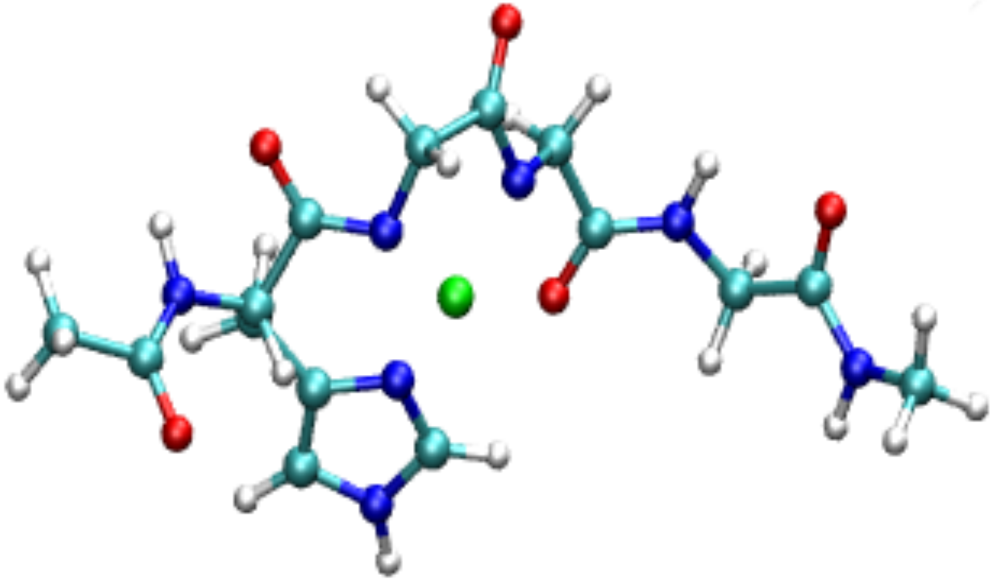
Cu$^{(+2)}$

O

N

C

H



B - Initial 2 x [$Cu^{(+2)}$ $HG^{(-)}G^{(-)}G$] configuration

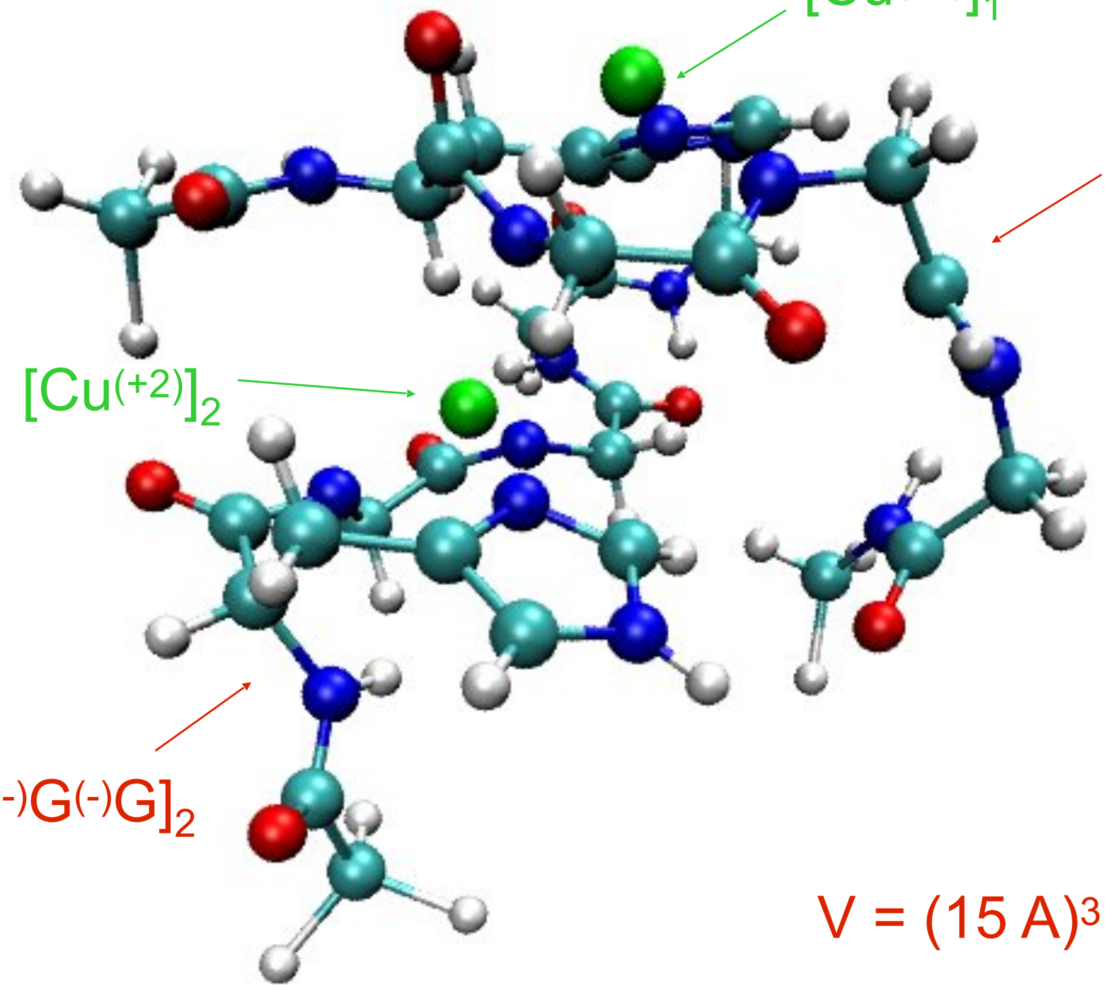
$Cu^{(+2)}$

O

N

C

H



$[Cu^{(+2)}]_1$

$[HG^{(-)}G^{(-)}G]_1$

$[Cu^{(+2)}]_2$

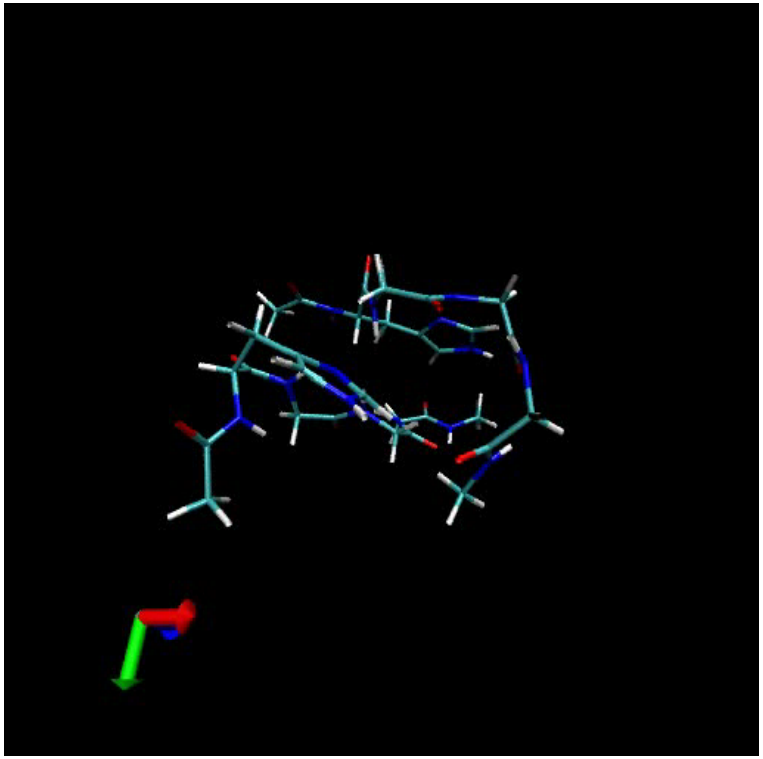
$[HG^{(-)}G^{(-)}G]_2$

V = (15 A)^3

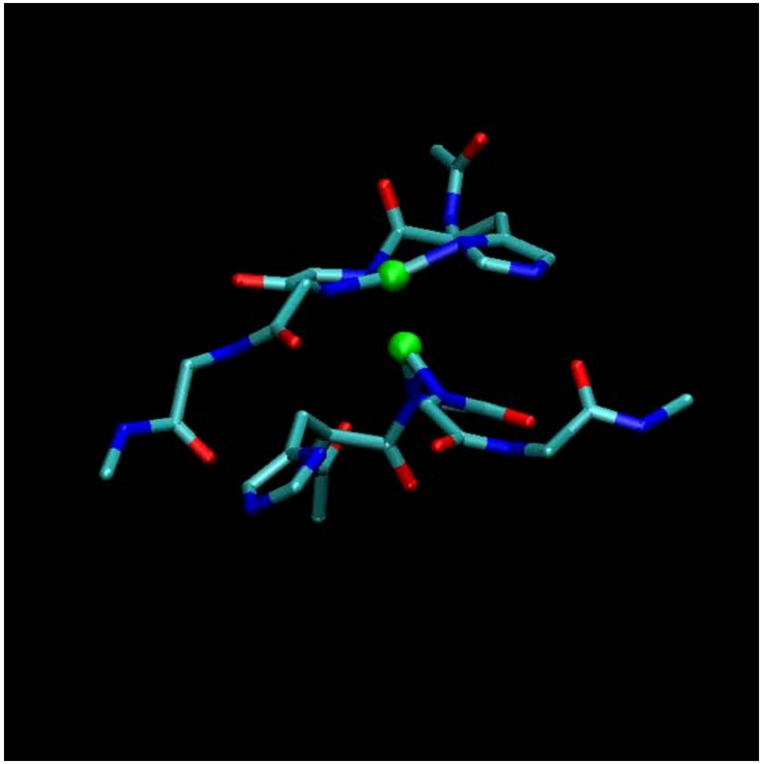
1.8 ps trajectory @ 300K

no Cu → no binding

Cu bonds with Gly and His are dynamically formed and destroyed



Cu
O
N
C



Quantum Mechanics at work

Car-Parrinello ab initio simulations

A *first principle* study of the influence of pH on the geometry of the Cu binding site in the HGGG + H(Im) peptide

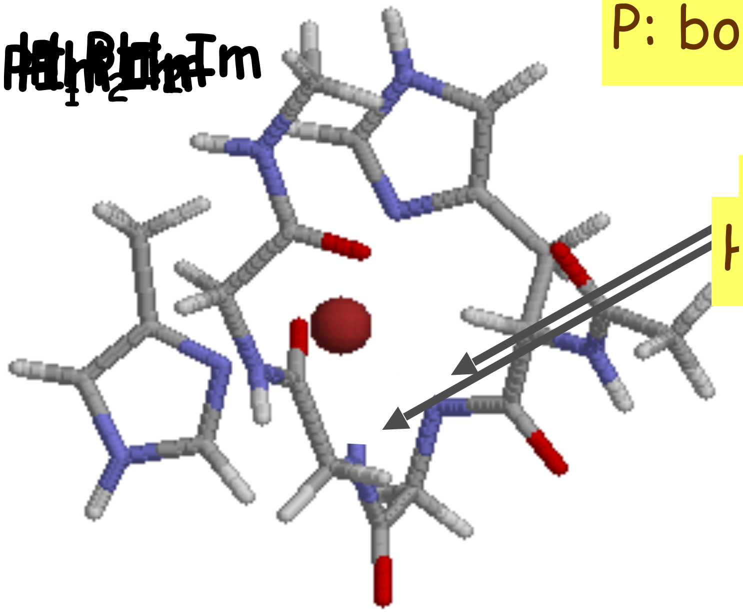
Furlan, La Penna, Guerrieri, Morante, GCR, JBIC

System 1, S1: $Cu^{2+}(HisG_1\text{-}G_2\text{-}G_3)$ + Im + 83 (H_2O)

System 2, S2: $Cu^{2+}(HisG_1G_2\text{-}G_3)$ + Im + 105 (H_2O)

System 3, S3: $Cu^{2+}(HisG_1\text{-}G_2G_3)$ + Im + 92 (H_2O)

System 4, S4: $Cu^{2+}(HisG_1G_2G_3)$ + Im + 92 (H_2O)



P: both Gly_1 and Gly_2 deprotonated

PH_2: only Gly_2 protonated

protonated

S1: HisG$_1$-G$_2$-G$_3$ + Im

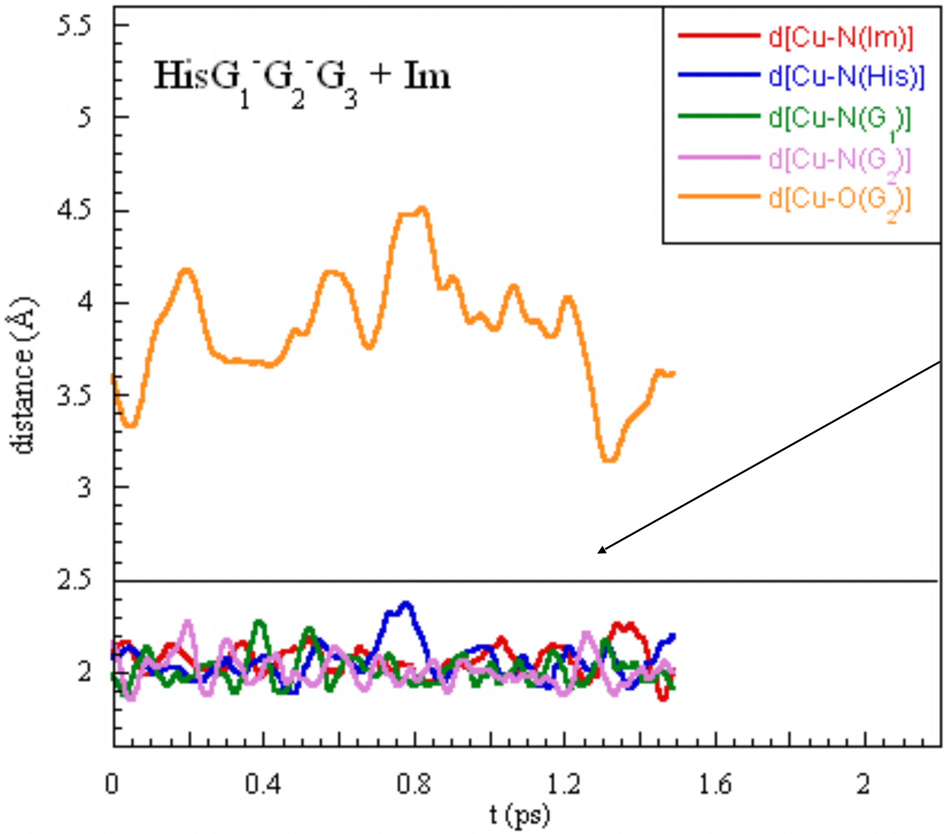
N_δ of isolated imidazole → N(Im)

N_δ of His → N(His)

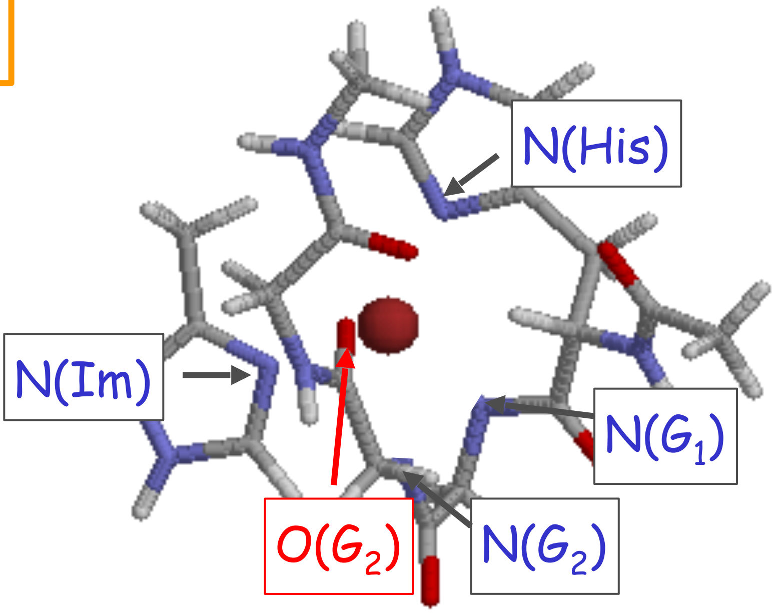
N of first Gly → N(G$_1$)

N of second Gly → N(G$_2$)

Carbonil O of second Gly → O(G$_2$)



line of “coordination sphere”



Atom	<d> (Å)	σ (Å)
		.08
N(His)	2.10	0.10
N(G$_1$)	2.01	0.08
N(G$_2$)	2.01	0.08
O(G$_2$)	3.80	0.30

S2: HisG$_1$G$_2$$^-G_3$ + Im

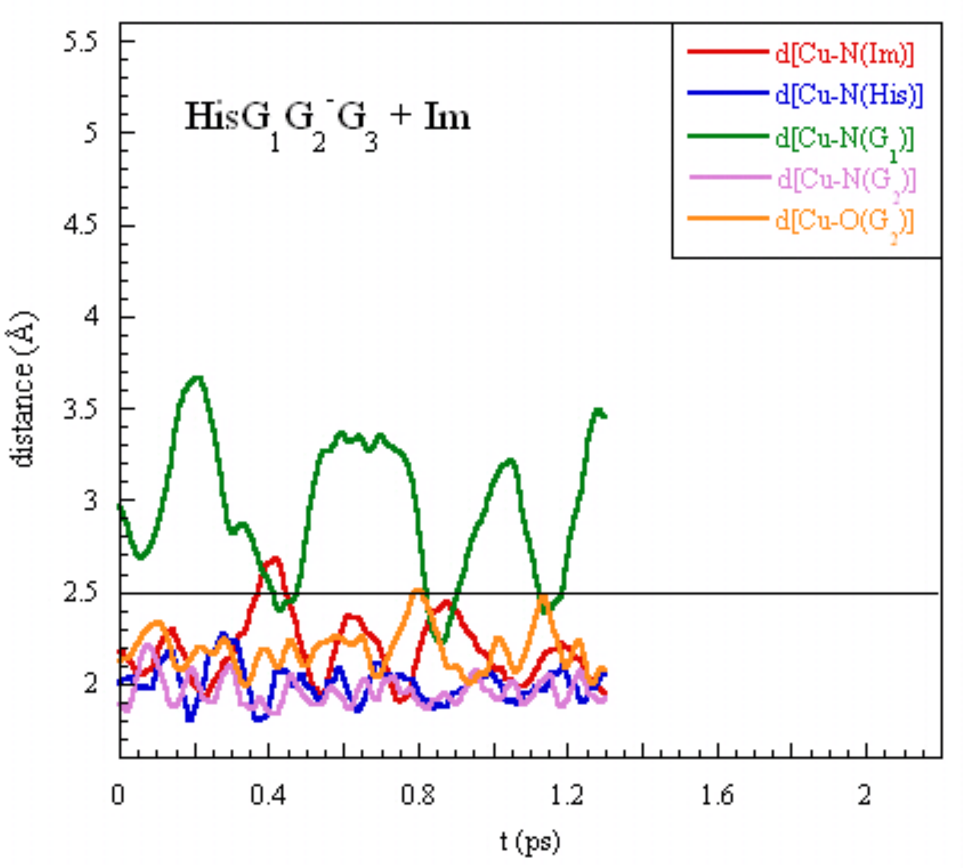
N$_\delta$ of isolated imidazole → N(Im)

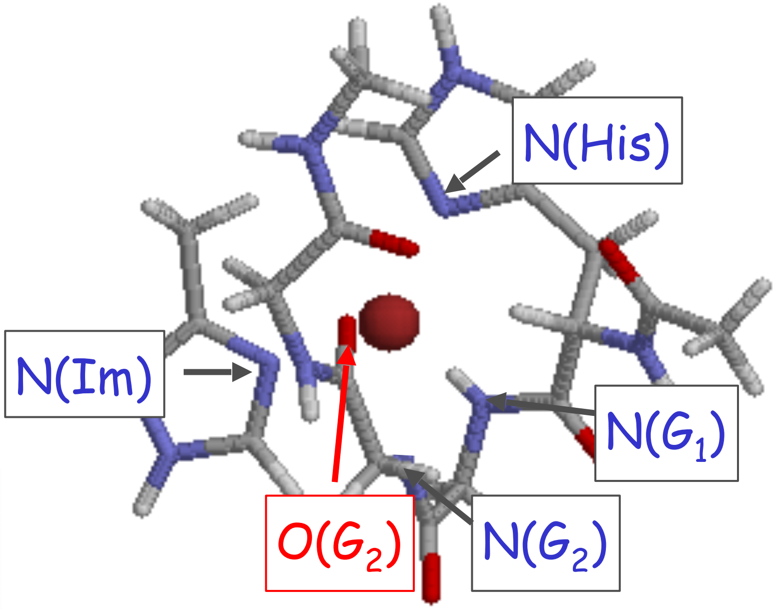
N$_\delta$ of His → N(His)

N of first Gly → N(G$_1$)

N of second Gly → N(G$_2$)

Carbonil O of second Gly → O(G$_2$)





Atom	<d>(Å)	σ (Å)
N(Im)	2.20	0.20
N(His)	2.00	0.10
N(G$_1$)	3.00	0.40
N(G$_2$)	1.96	0.07
O(G$_2$)	2.20	0.10

S3: HisG$_1$$^-G_2G_3$ + Im

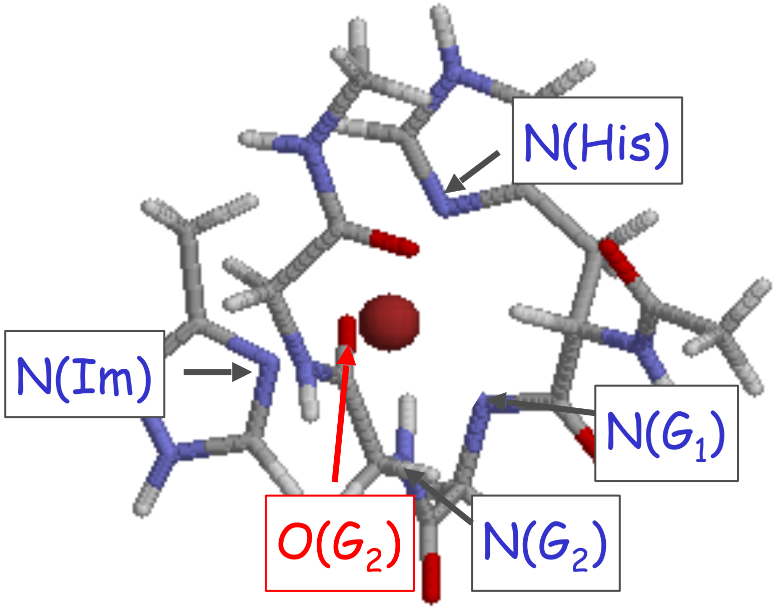
Nδ of isolated imidazole → N(Im)

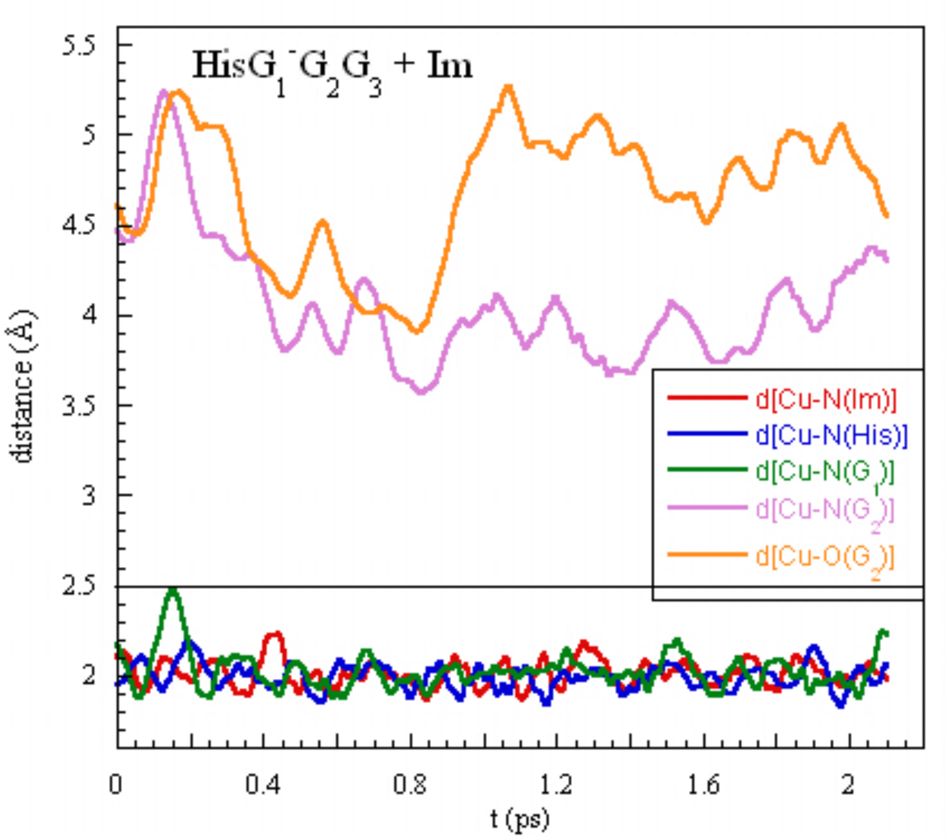
Nδ of His → N(His)

N of first Gly → N(G$_1$)

N of second Gly → N(G$_2$)

Carbonil O of second Gly → O(G$_2$)





Atom	<d>(Å)	σ (Å)
N(Im)	2.01	0.07
N(His)	1.99	0.07
N(G$_1$)	2.00	0.10
N(G$_2$)	4.10	0.30
O(G$_2$)	4.70	0.40

S4: HisG$_1$G$_2$G$_3$ + Im

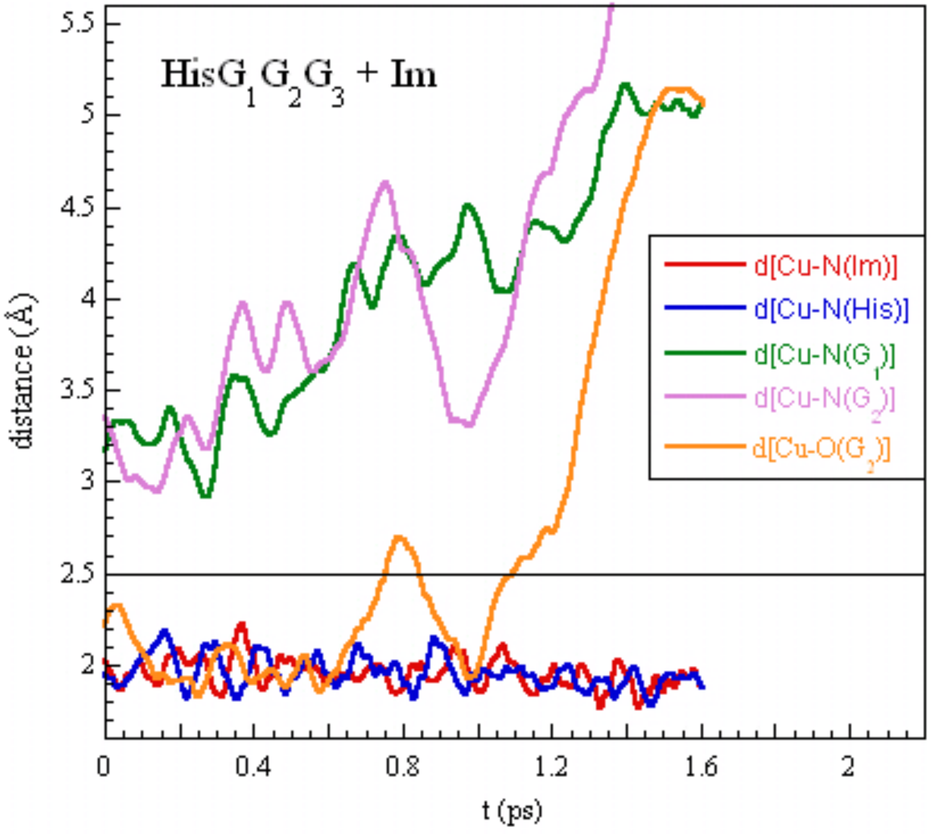
Nδ of isolated imidazole → N(Im)

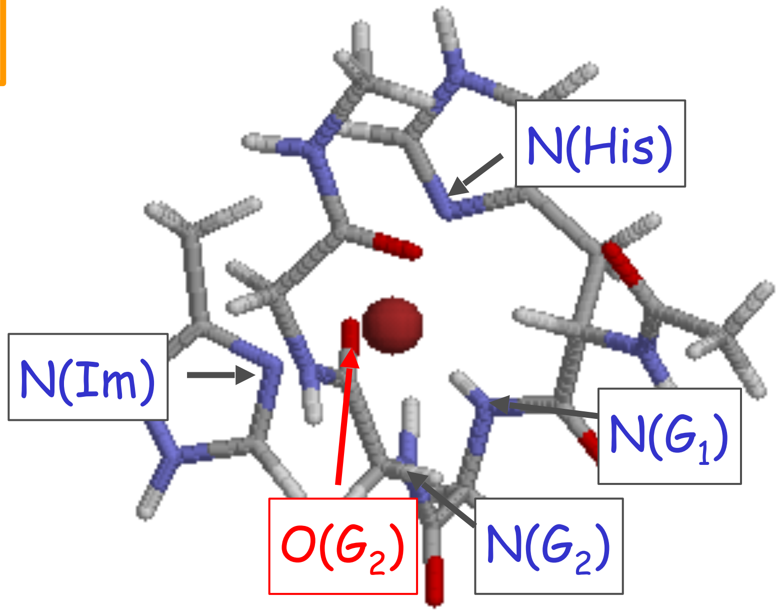
Nδ of His → N(His)

N of first Gly → N(G$_1$)

N of second Gly → N(G$_2$)

Carbonil O of second Gly → O(G$_2$)





Atom	<d> (Å)	σ (Å)
N(Im)	1.95	0.08
N(His)	1.95	0.08
N(G$_1$)	---	---
N(G$_2$)	---	---
O(G$_2$)	---	---

Gly$_2$ protonated → low coordination number

3N

2N

BUT…

Atom	S1 HisG$_1$-G$_2$-G$_3$		S2 HisG$_1$G$_2$-G$_3$		S3 HisG$_1$-G$_2$G$_3$		S4 HisG$_1$G$_2$G$_3$	
	<d>	σ	<d>	σ	<d>	σ	<d>	σ
N(Im)	2.07	0.08	[illegible]	0.20	[illegible]	0.07	1.95	0.08
N(His)	2.10	0.10	2.00	0.10	1.99	0.07	1.95	0.08
N(G$_1$)	2.01	0.08	3.00	0.40	2.00	0.10	---	---
N(G$_2$)	2.01	[illegible]	[illegible]	0.07	4.10	0.30	---	---
O(G$_2$)	3.80	0.30	2.20	0.10	4.70	0.40	---	---

3N1O

Imidazole ring is always in Cu coordination sphere independently of Gly's protonation state

Gly protonation state and Imidazole binding

A stability study

Is the dimeric (two octarepeats) compound more/less stable than the monomeric one?

Compute energies of products of the virtual chemical reactions:

1. $H_1PH_2Im \rightarrow H_1PH_2 + Im \rightarrow H^+_1 + PH_2 + Im \rightarrow H^+_1 + H^+_2 + P + Im$

2. $H_1PH_2Im \rightarrow H_1PH_2 + Im \rightarrow H^+_1P + H^+_2 + Im \rightarrow H^+_1 + H^+_2 + P + Im$

3. $PH_2Im \rightarrow PH_2 + Im \rightarrow H^+_2 + P + Im$

4. $H_1PIm \rightarrow H_1P + Im \rightarrow H^+_1 + P + Im$

5. $PIm \rightarrow P + Im$

P: both Gly_1 and Gly_2 deprotonated

H_1P: only Gly_2 deprotonated

PH_2: only Gly_1 deprotonated

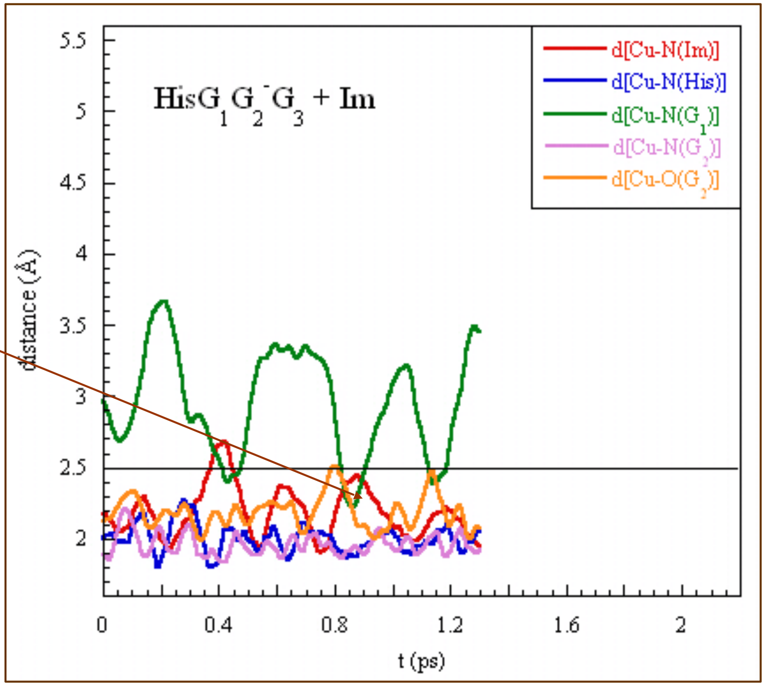
H_1PH_2: both Gly_1 and Gly_2 protonated

Two types of Conclusions

Methodological

we have seen the power of using CP-MD in combination with DFT optimization

“unstable” structures can be recognized and, if needed, discarded

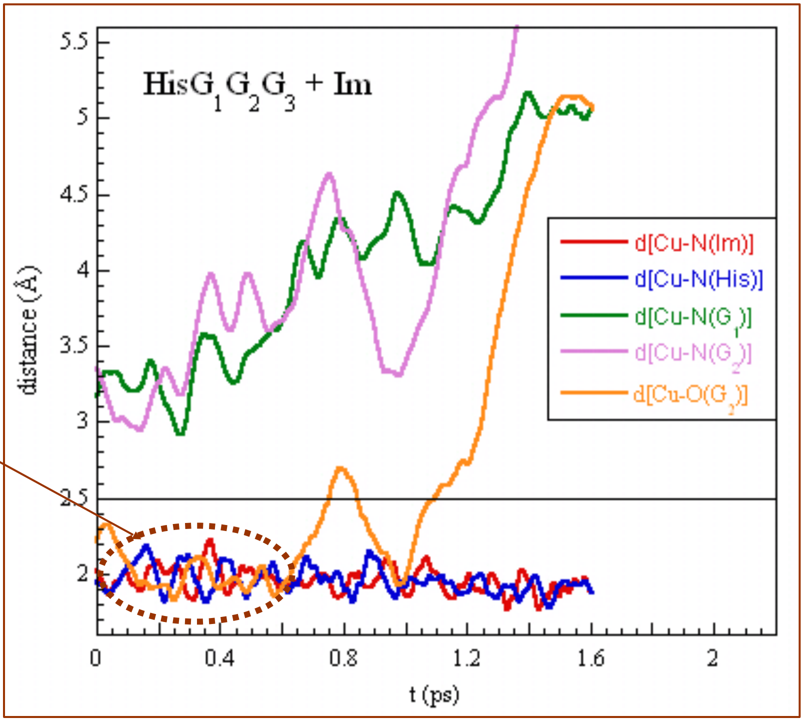


Two types of Conclusions

Methodological

we have seen the power of using CP-MD in combination with DFT optimization

"unstable" structures can be recognized and, if needed, discarded



Biological

Multiple Histidine coordination can occur in the presence of deprotonated Glycines

⇓

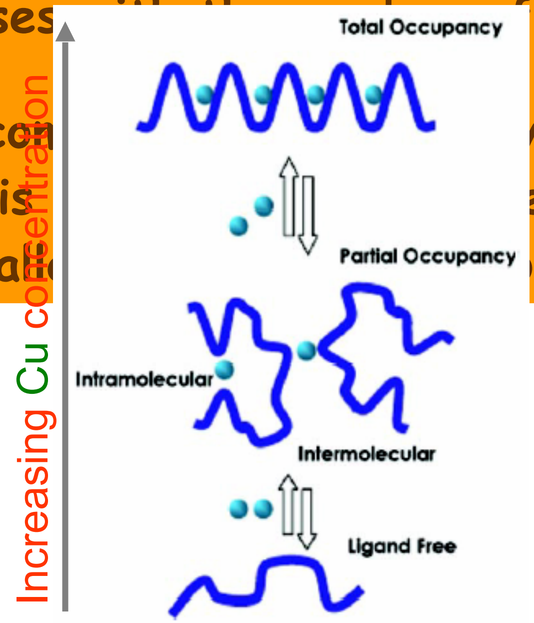
The hypothesis that low copper concentration favors inter-repeat binding is confirmed

The presence of the extra His stablize the Cu peptide complex

⇑

The binding energy decreases [illegible] deprotonated Glycines

The energy of the co[illegible]'s nitrogens are deprotonated, P, is [illegible]e find for the crystal[illegible]n



VI. Conclusions and outlook

Conclusions

Very many difficult problems

But there is hope to successfully attack some of them

Extremely exciting research field

An arena where biology, mathematics, physics, computer science meet

Amazing experimental methods are being developed

Fantastic applications are in view

New positions are foreseeable!

Conclusions

Very many difficult problems

But there is hope to successfully attack some of them

Extremely exciting research field

An arena where biology, mathematics, physics, computer science meet

Amazing experimental methods are being developed

Fantastic applications are in view

New positions are foreseeable!

Outlook?

Outlook

This is not the end.

It is not even the beginning of the end.

But it is, perhaps, the end of the beginning

But for today

Thank you all for listening!