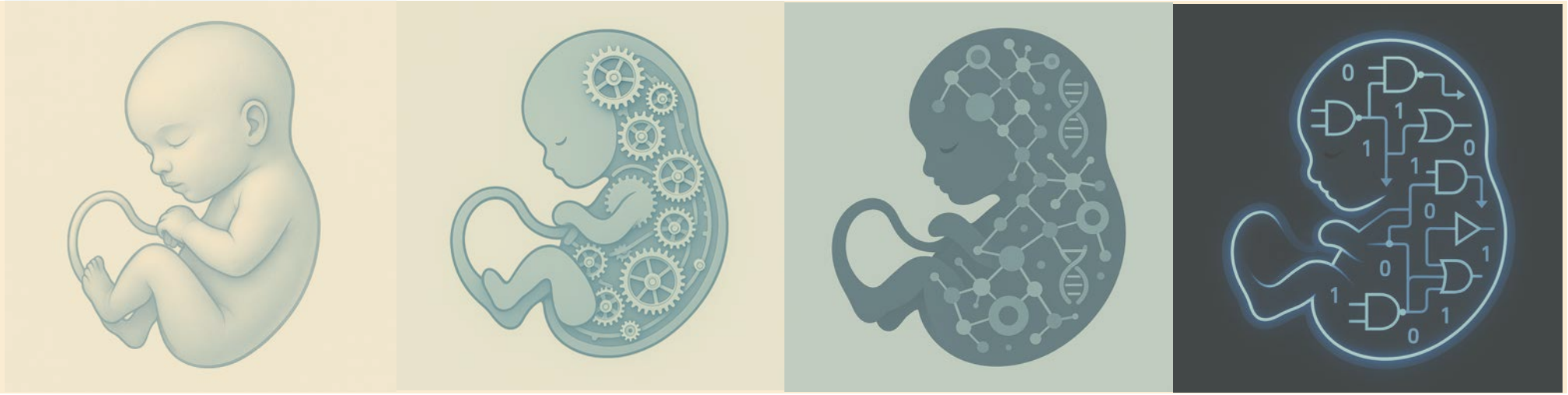
What is biological information? (II)



Course 1: Introduction –

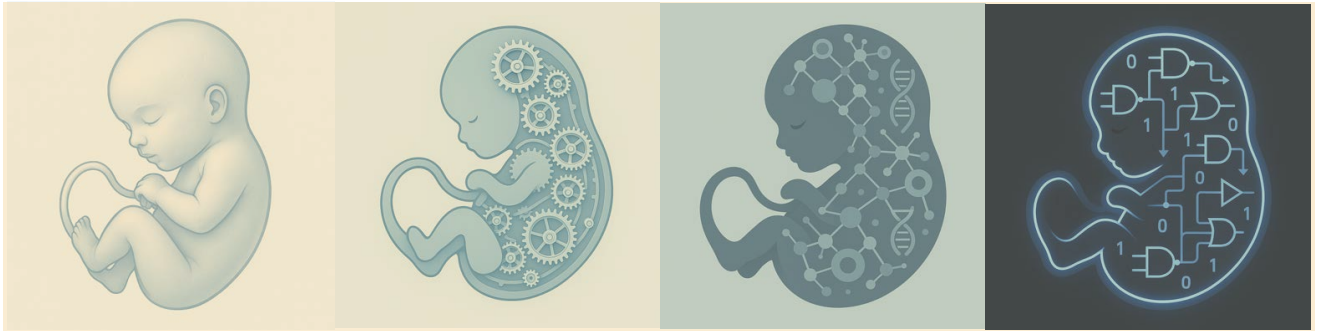
Biological computation

Thomas Lecuit

chaire: Dynamiques du vivant

COLLÈGE DE FRANCE 1530

What is biological information? (II)



- *Organisation et dynamique*
- Flux de matière

« Un courant de matières (…) traverse continuellement l'organisme et le renouvelle dans sa substance en le maintenant dans sa forme. Ce mouvement, qu'on a appelé le *tourbillon vital*, (…) est la condition et la cause immédiate de toutes les autres manifestations vitales. (…) L'universalité d'un tel phénomène, la constance qu'il présente, sa nécessité, en font le caractère fondamental de l'être vivant, le signe plus général de la vie."

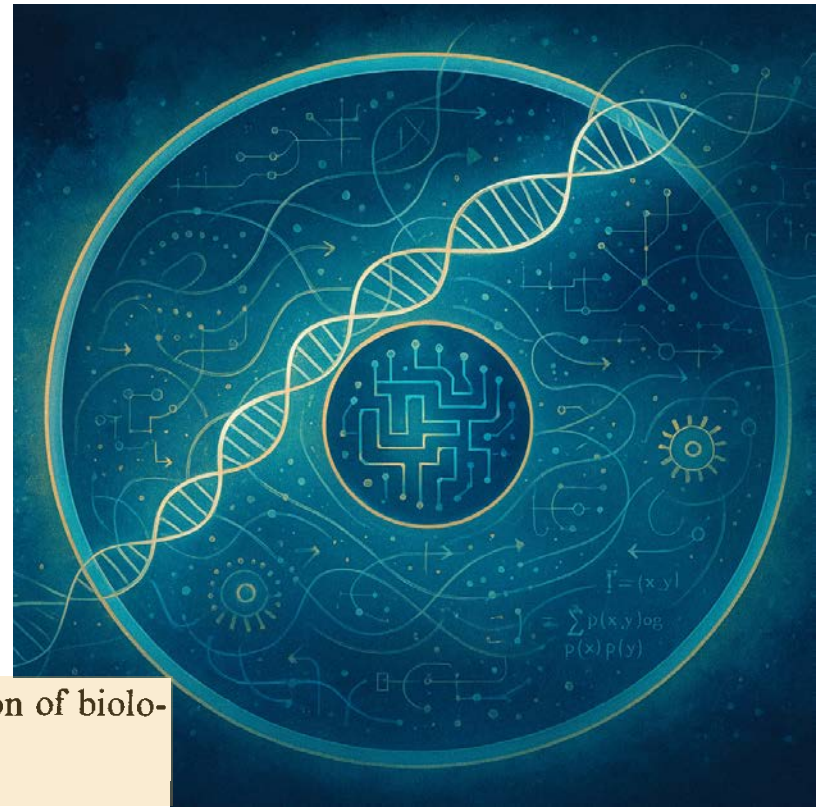
C. Bernard 1875, *Définition de la vie*, partie III.

- Flux d'*énergie: dynamique (hors d'équilibre)*
- Flux d'*information: organisation*

What is biological information? (II)

The power and mystery of life is entangled within the information processing at the heart of all cellular machinery.

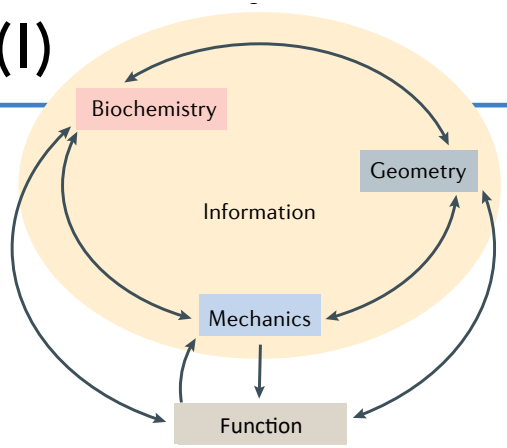
Qian, L. & Winfree, E. *Science* 332, 1196–1201 (2011).



THE PROBLEM of Morphogenesis—broadly understood as the origin and evolution of biological structures—is one of the outstanding questions in present day Biology.

Thom, R. Topological models in biology. *Topology* 8, 313–335. (1969)

Overview – What is Biological Information? (I)



- *Theory of communication*: probabilistic definition of information based on ignorance/knowledge.
- *The content of* information (semantic) is irrelevant.
- Information as entropy is a quantity:
    1. Codes, Information encoding, decoding, recoding
    2. Communication/transmission
    3. Capacity of a channel.

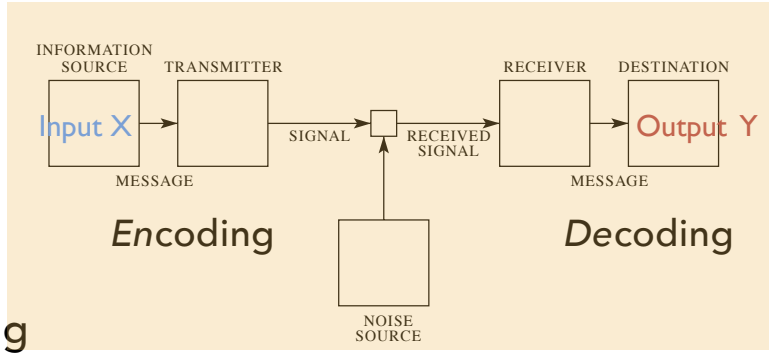


Fig. 1 — Schematic diagram of a general communication system.

- Chemical encoding and decoding of *space* and *time*:
    - New look at embryonic patterning and cell signalling

- But there are other forms of information: Mechanical, and Structural information.

  It is not clear how to account for this.
- Learning and memory.

Why study information?

1. Information provides a framework to decipher the meaningfulness of processes

Intimately linked to functions.

- Chemical characterisation of living processes
- Mechanical description as well
- These allow bottom up models of organisation and dynamics

- Molecules have a context dependent *meaning* for a system.

Information is not absolute but *contextual*.

What is an efficient and relevant encoding? One that has meaning (ie. functional)

- Information as *meaning* is fundamentally *orthogonal* to what Shannon proposed.

But meaning matters in biology. There is a need to address this.

Why study information?

1. Information provides a framework to decipher the meaningfulness of processes

Intimately linked to functions.

2. Information provides a language to decipher the *logic of living systems*

In what sense can it be said that living systems compute?

Consider the *machine metaphor.*

This is nothing but a metaphor. There are many reasons to say that living organisms are not mere complex machines:

- Not exposed to a finite list of instructions. Exposed rather to open environment, including novelty.
- Responses are not invariant. Context dependency.
- Robust to perturbations. Machines are brittle

But this is a useful metaphor

This year

Information provides a language to address the *meaning* and *logic in living systems*

Use Computational metaphor to account for:

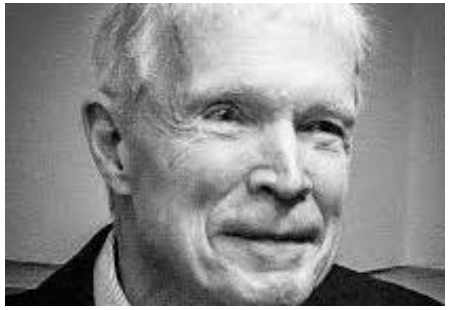
- Purpose/function
- Rules & Logic

Biological processes can be usefully described as **information processing**: they involve the manipulation of symbolic representations according to formal rules to achieve a functional outcome.

This perspective shifts the focus from the specific chemical substrates (e.g., DNA, proteins) to the logical and algorithmic principles governing the system.

Computation as prediction

Link between logic and meaning: prediction



John J. Hopfield

Physics, Computation, and Why Biology Looks so Different

J. J. HOPFIELD

Divisions of Chemistry and Biology, California Institute of Technology, Pasadena, CA 91125, *U.S.A.*

Crudely put, one who can predict the future from the present and make advantageous choices of action on the basis of that prediction will generally win in the game of evolution. Much of the history of evolution can be read as the evolution of systems to make environmental measurements, make predictions, and generate appropriate actions. This pattern has the essential aspects of a computational system, where the inputs are from environmental measurements, the outputs are the signals (chemical or electrical) which modulate the behavior, and the computation represents an appropriate generation of outputs in response to environmental signals.

JJ. Hopfield. *Journal of Theoretical Biology* 171 (1), 53-60
Physics, computation, and why biology looks so different

Living systems as computing machines

- Historical perspective

- Descartes (1662)
- Living organisms are deterministic machines akin to clockworks
- Source of motion is heat and mechanics

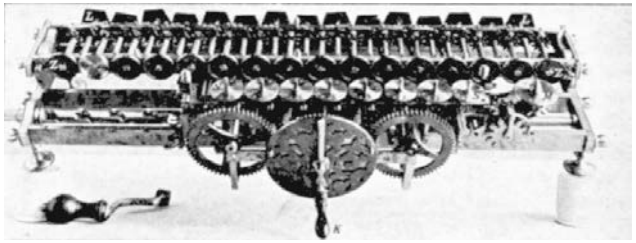
« Je ne reconnais aucune différence entre les machines que font les artisans et les divers corps que la nature seule compose. (…). Et il est certain que toutes les règles des Mécaniques appartiennent à la Physique, en sorte que toutes les choses qui sont artificielles, sont avec cela naturelles. Car, par exemple, lorsqu'une montre marque les heures par le moyen des roues dont elle est faite, cela ne lui est pas moins naturel qu'il est à un arbre… de produire ses fruits. »

Descartes, *Principes de la Philosophie , 1644*.

- Leibniz (1666)
- Logic and rationality: *calculus ratiocinator* using the *characteristica universalis*

"The history of the modern computing machine goes back to Leibniz and Pascal. Indeed, the general idea of a computing machine is nothing but a mechanization of Leibniz's *calculus ratiocinator*. »

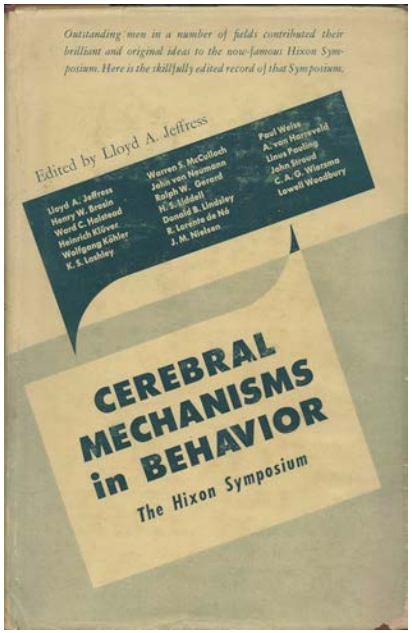
Norbert Wiener (1948)

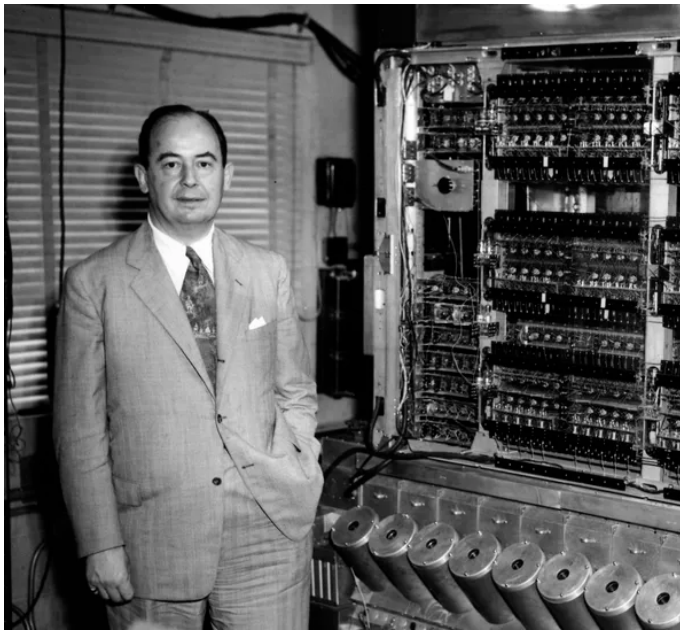


Living systems as computing machines

- Historical perspective

The General and Logical Theory of Automata





John von Neumann (1903-1957)

- The cell is a self-replicating machine
- It contains and processes information

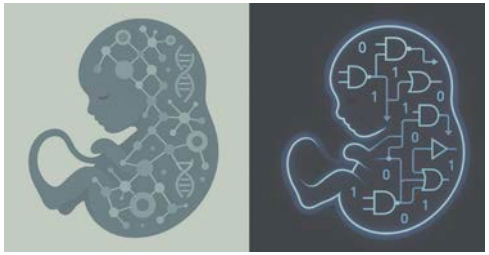
Living systems as computing machines

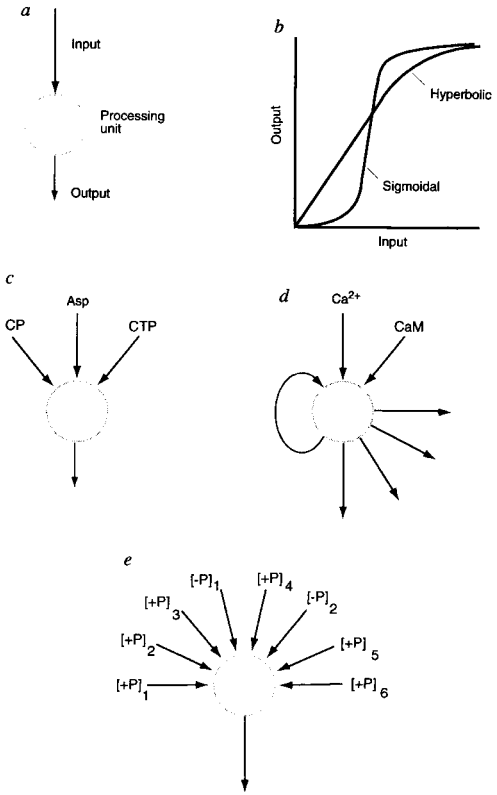
- Historical perspective

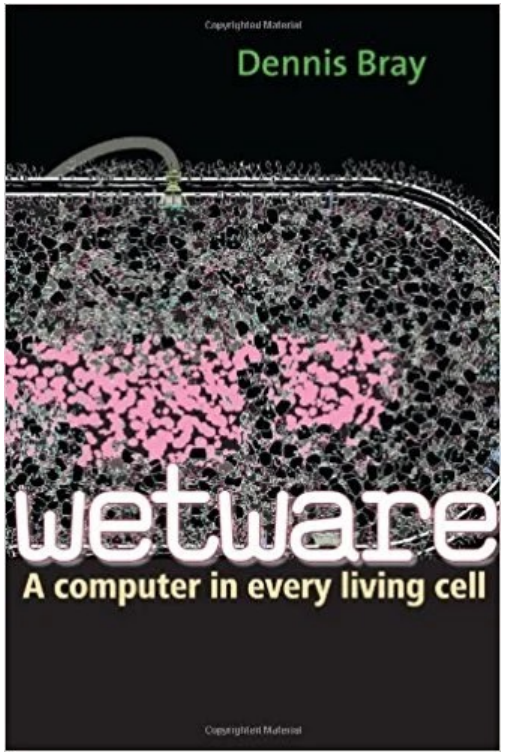
Protein molecules as computational elements in living cells

Dennis Bray

Many proteins in living cells appear to have as their primary function the transfer and processing of information, rather than the chemical transformation of metabolic intermediates or the building of cellular structures. Such proteins are functionally linked through allosteric or other mechanisms into biochemical 'circuits' that perform a variety of simple computational tasks including amplification, integration and information storage.



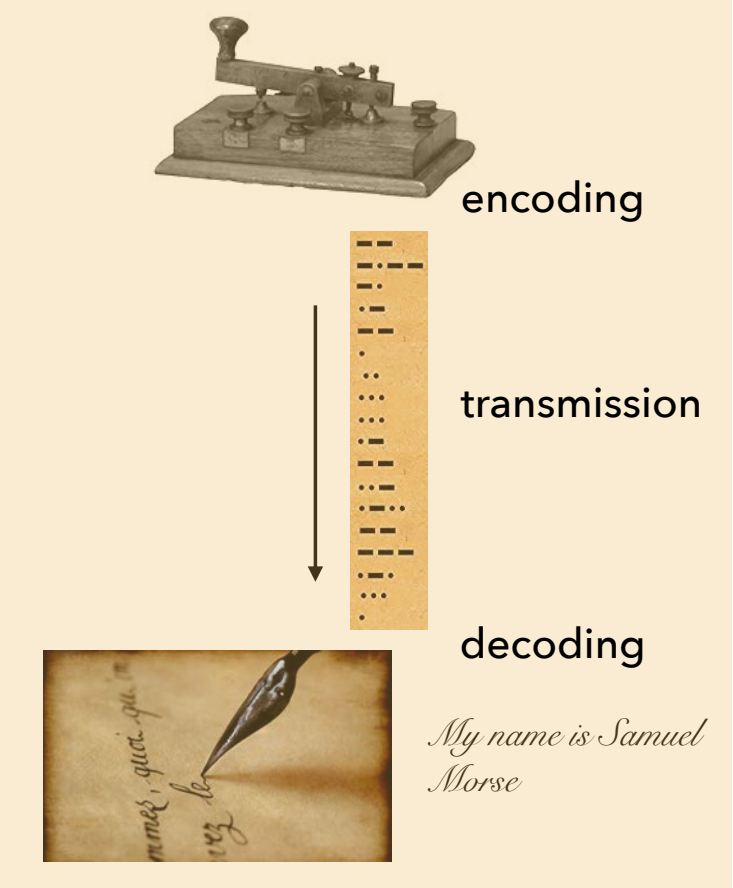




Communication, Information, Codes in Humans

- Information is:
  1. Encoded
  2. Sent (sender)
  3. Transmitted (via electric signals)
  4. Interpreted (receiver)

- A code is used as an *intermediate* between two forms of information
- A code *transforms* an information into another.
- In other words, a code changes a *representation* into another one.

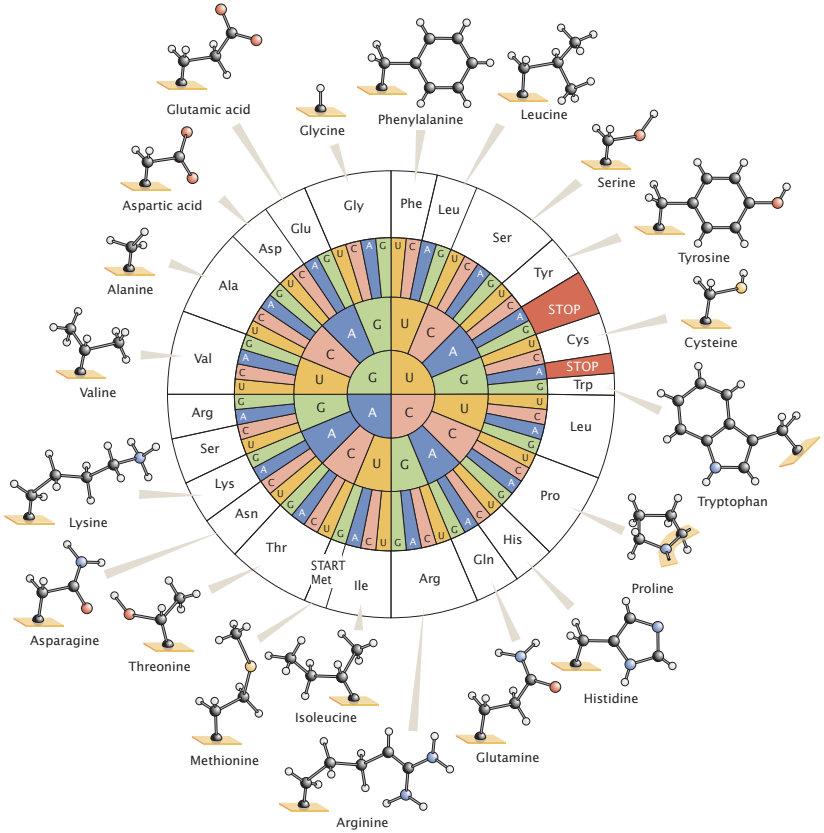


Computation as Translation of symbols

- Symbols are used as *representations* for « things »
- Input Data are *transformed* into Output Data
- Input Symbols are *translated* into output Symbols

My name is
Samuel Morse

System:	Hash Marks	Roman	Decimal	Binary
Zero	n/a	n/a	0	0
One	I	I	1	1
Two	II	II	2	10
Three	III	III	3	11
Four	IIII	IV	4	100
Five	/III/	V	5	101
Six	/III/ I	VI	6	110
Seven	/III/ II	VII	7	111
Eight	/III/ III	VIII	8	1000
Nine	/III/ IIII	IX	9	1001
Ten	/III/ /III/	X	10	1010
Eleven	/III/ /III/ I	XI	11	1011
Twelve	/III/ /III/ II	XII	12	1100
Thirteen	/III/ /III/ III	XIII	13	1101
Fourteen	/III/ /III/ IIII	XIV	14	1110
Fifteen	/III/ /III/ /III/	XV	15	1111
Sixteen	/III/ /III/ /III/ I	XVI	16	10000
Seventeen	/III/ /III/ /III/ II	XVII	17	10001
Eighteen	/III/ /III/ /III/ III	XVIII	18	10010
Nineteen	/III/ /III/ /III/ IIII	XIX	19	10011
Twenty	/III/ /III/ /III/ /III/	XX	20	10100



What is a computational machine?

- **Computational process:**

A formally defined, *step-by-step* *transformation of information* (or state) according to a set of precise, *unambiguous* *rules*.

1. **Transformation of Information/State**: It takes an input (data, a symbol, a representation of a problem) and produces an output. *Eg. calculating a sum, sorting a list, rendering an image etc*

2. **Step-by-Step**: Computation is a sequence of discrete, individual steps that happen one after the other (or sometimes in parallel).

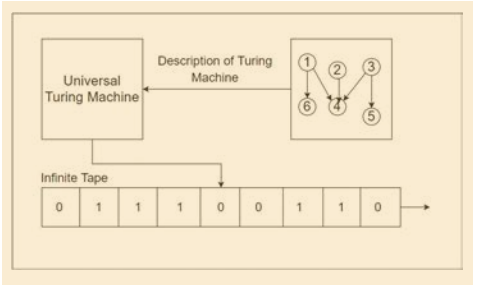
3. **Formally Defined & Unambiguous Rules**: the rules governing each step must be so precise that they can be automatised by a machine. This set of rules is called an algorithm.

What is a computational machine?

- **Computational process:**

A formally defined, *step-by-step* *transformation of information* (or state) according to a set of precise, *unambiguous* *rules*.

- **Tape** (infinite)
- Reading/writing **Head**
- **State register:** variable configuration
- **Transition table: rules**



Turing, A. M. *Proc. Lond. Math. Soc.* s2–42, 230–265 (1936).

- **Requirements** (eg. Turing machine):

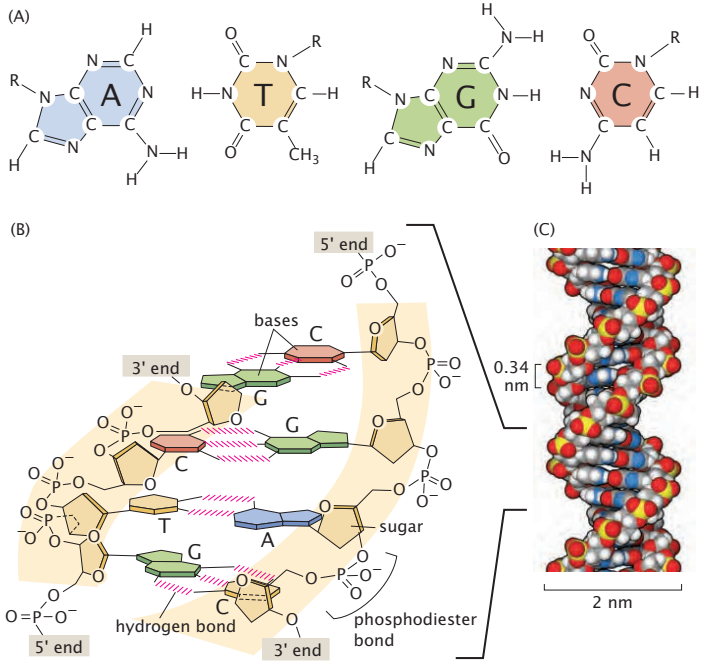
1. Representation of information (Tape)

2. Transformation (Head) : read, write, and change system's state (according to algorithm).
   - **IF** the device is in State `X` **AND** it reads Symbol `Y`,
   - **THEN** it should write Symbol `Z`, change to State `W`, and move the read/write mechanism (e.g., move left or right).
3. Control (Program/Rules): sequence operations

NB: It doesn't matter how it is instantiated (gears, transistors, etc)

Biological molecular computation

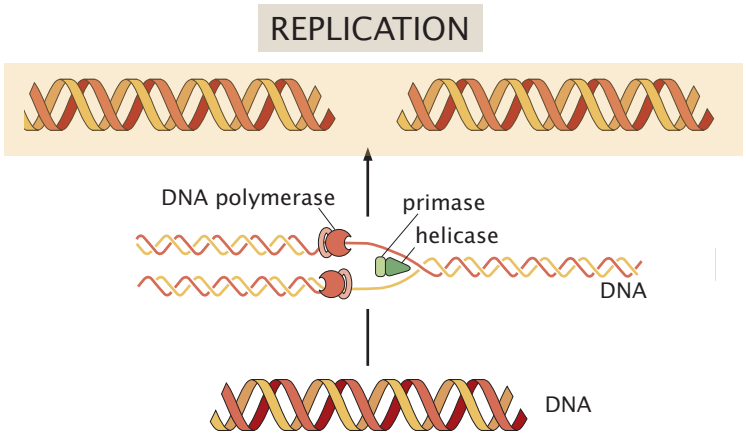
1. **Means of Representation (the "Tape"):** Information is stored in biological structures.
   - **Symbols:** Nucleotides (A, T, C, G) in DNA, amino acids in proteins, ions (Ca^{2+}), or the firing states (spike/not spike) of neurons.
   - **Memory Medium:** DNA (long-term storage), RNA (short-term/temporary storage), concentrations of specific proteins or chemicals inside a cell.

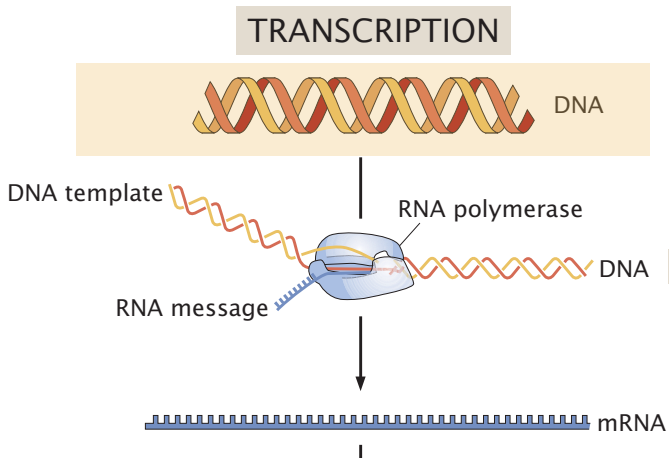


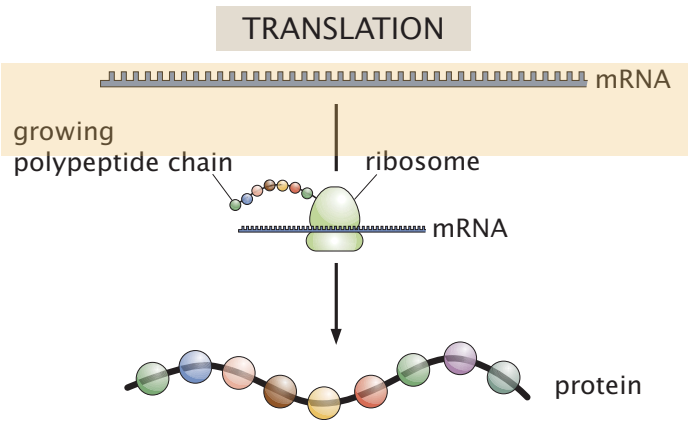
R. Phillips, J. Kondev, J. Thériot & H. Garcia. *Physical Biology of the Cell (Garland Science)* 2012

Biological molecular computation

1. **Means of Representation (the "Tape"):** Information is stored in biological structures.
   - **Symbols:** Nucleotides (A, T, C, G) in DNA, amino acids in proteins, ions (Ca^{2+}), or the firing states (spike/not spike) of neurons.
   - **Memory Medium:** DNA (long-term storage), RNA (short-term/temporary storage), concentrations of specific proteins or chemicals inside a cell.







R. Phillips, J. Kondev, J. Thériot & H. Garcia. *Physical Biology of the Cell (Garland Science)* 2012

Biological molecular computation

2. Means of Transformation ("Head"):

- Central Dogma: Crick 1970

F. Crick (1916-2004)

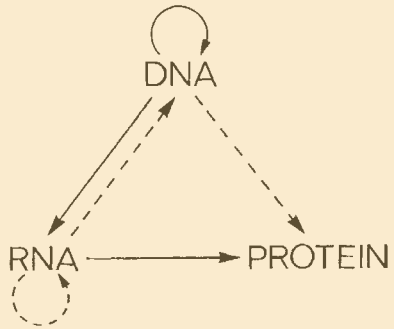
Central Dogma of Molecular Biology

by

FRANCIS CRICK

MRC Laboratory of Molecular Biology,
Hills Road,
Cambridge CB2 2QH

The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid.

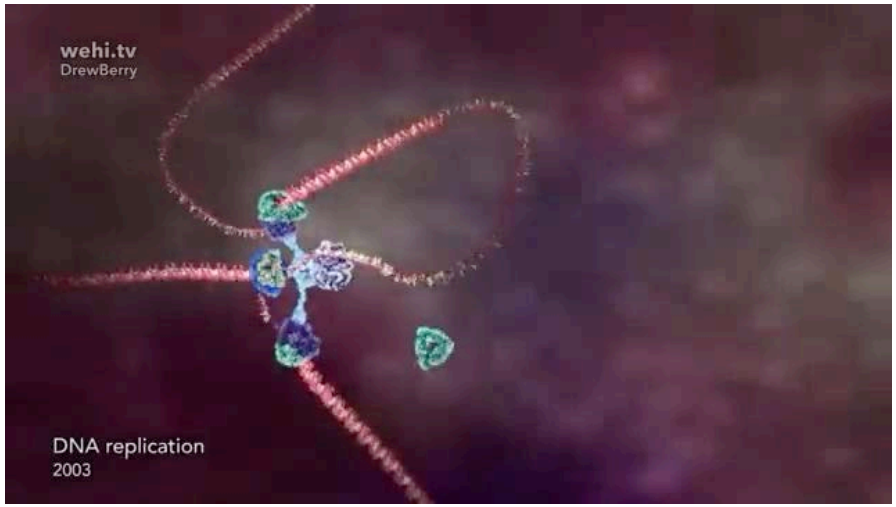


Biological molecular computation

2. Means of Transformation ("Head"):

The "processing" is done by molecular machineries.

- **Read/Write Mechanism:** Enzymes like **DNA polymerase** (reads a DNA template and writes a new strand) and **RNA polymerase** (reads DNA and writes RNA). **Ribosomes** (read an RNA and write polypeptide)



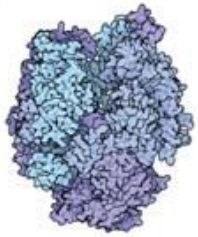
Biological molecular computation

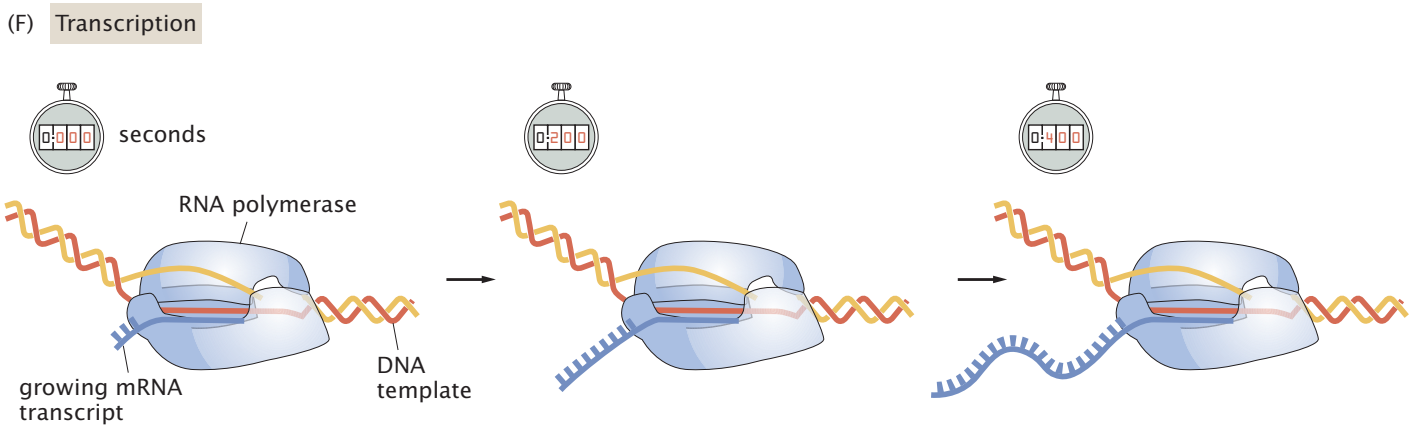
2. Means of Transformation ("Head"):

The "processing" is done by molecular machinery.

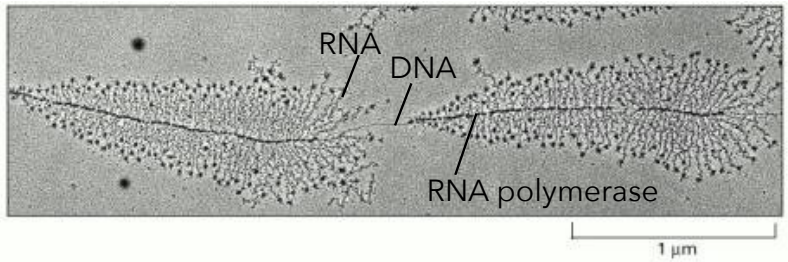
- **Read/Write Mechanism:** Enzymes like **DNA polymerase** (reads a DNA template and writes a new strand) and **RNA polymerase** (reads DNA and writes RNA). **Ribosomes** (read an RNA and write polypeptide)

eukaryotic
RNA polymerase (1i6h)





R. Phillips, J. Kondev, J. Thériot & H. Garcia.
Physical Biology of the Cell (Garland Science) 2012



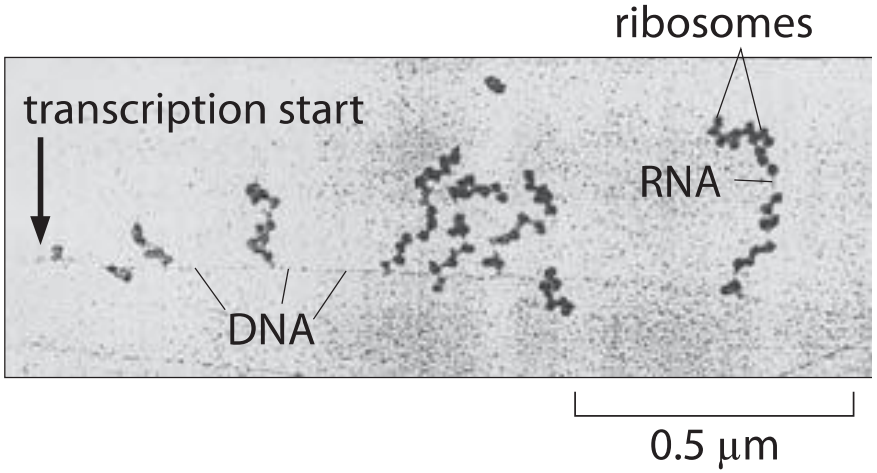
OL. Miller and BR. Beatty, *Science* 1969 May 23;164(3882):955-7

Biological molecular computation

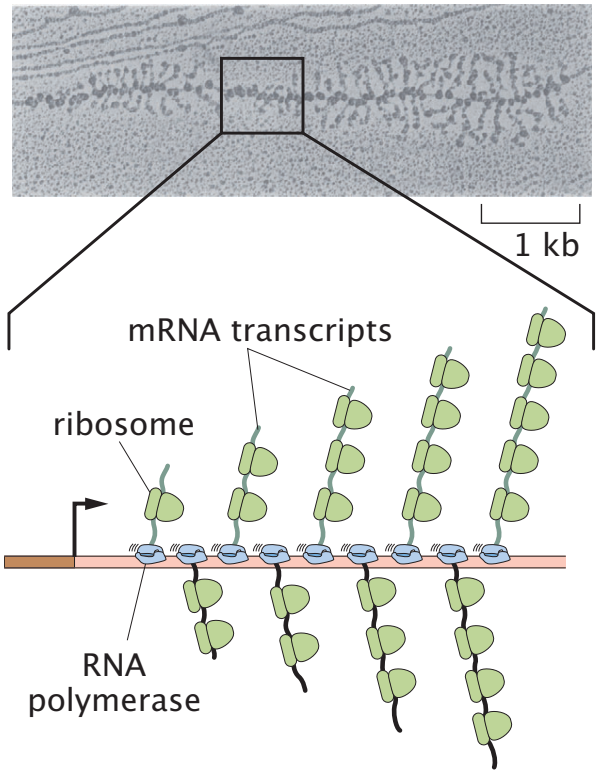
2. Means of Transformation ("Head"):

The "processing" is done by molecular machinery..

- **Read/Write Mechanism:** Enzymes like **DNA polymerase** (reads a DNA template and writes a new strand) and **RNA polymerase** (reads DNA and writes RNA). **Ribosomes** (read an RNA and write polypeptide)



R. Phillips, J. Kondev, J. Thériot & H. Garcia.
Physical Biology of the Cell (Garland Science) 2012

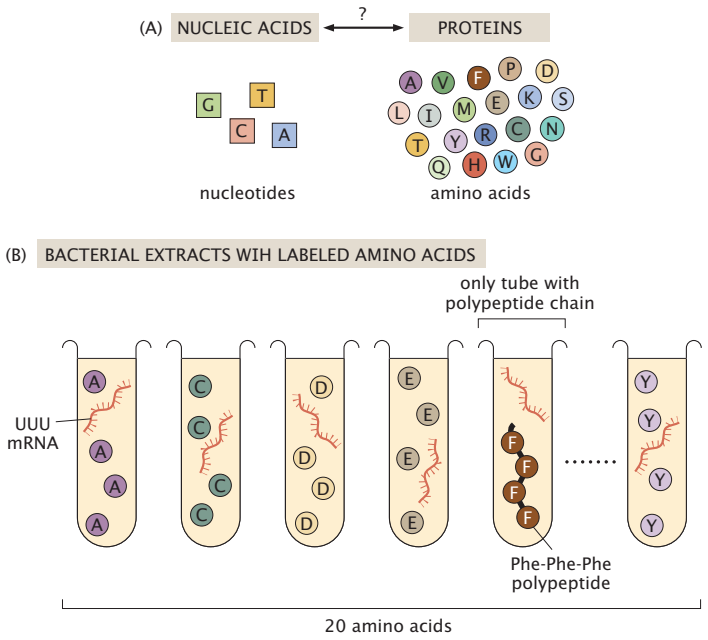


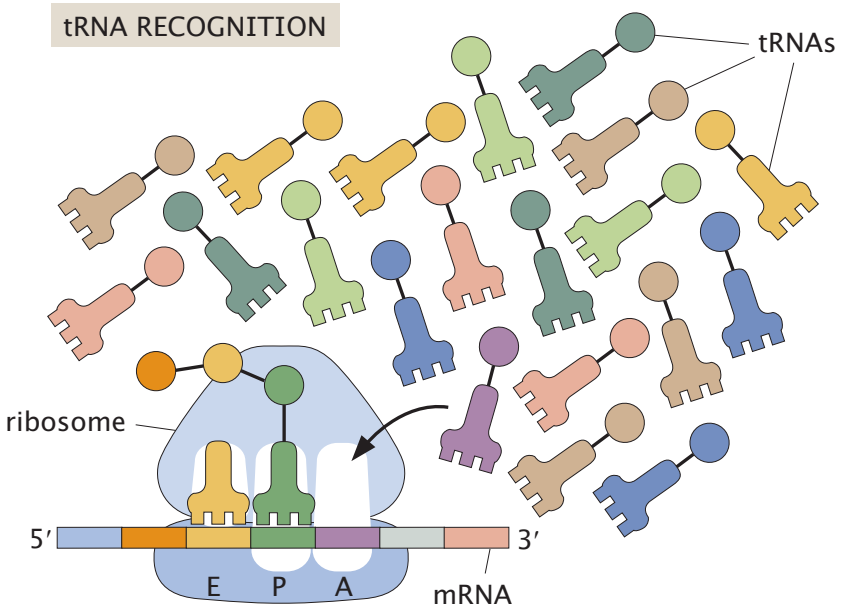
Biological molecular computation

2. Means of Transformation ("Head"):

The "processing" is done by molecular machinery..

- **Read/Write Mechanism:** Enzymes like **DNA polymerase** (reads a DNA template and writes a new strand) and **RNA polymerase** (reads DNA and writes RNA). **Ribosomes** (read an RNA and write polypeptide)





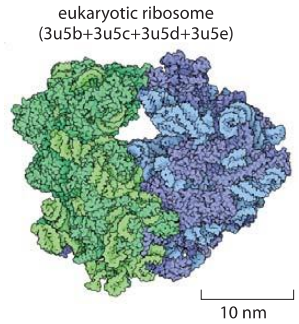
R. Phillips, J. Kondev, J. Thériot & H. Garcia. *Physical Biology of the Cell (Garland Science)* 2012

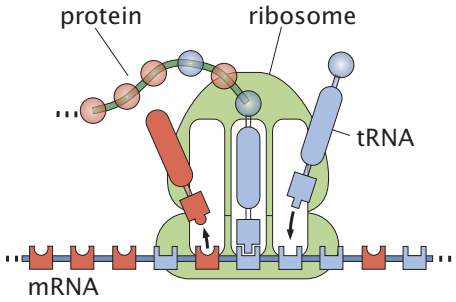
Biological molecular computation

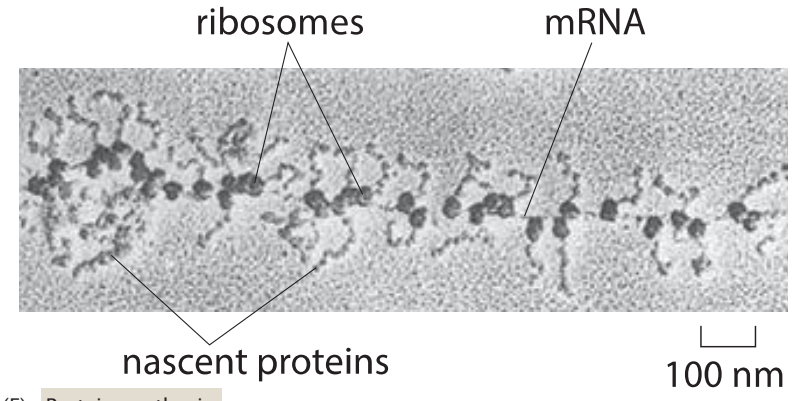
2. Means of Transformation ("Head"):

The "processing" is done by molecular machinery..

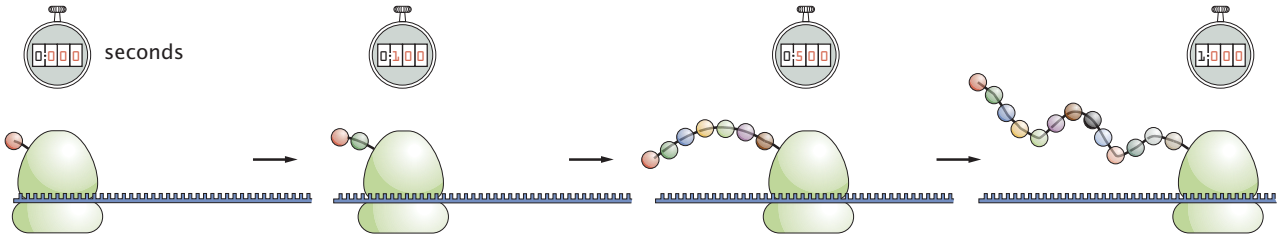
- **Read/Write Mechanism:** Enzymes like **DNA polymerase** (reads a DNA template and writes a new strand) and **RNA polymerase** (reads DNA and writes RNA). **Ribosomes** (read an RNA and write polypeptide)







(E) Protein synthesis

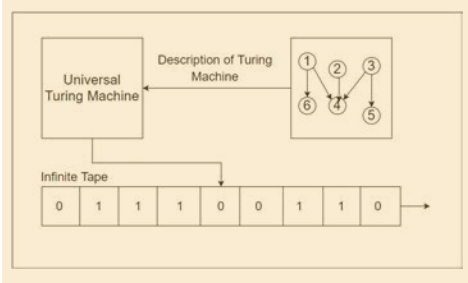


What is a computational machine?

- **Computational process:**

A formally defined, *step-by-step* *transformation of information* (or state) according to a set of precise, *unambiguous* *rules*.

  - **Tape** (infinite)
  - Reading/writing **Head**
  - **State register:** variable configuration
  - **Transition table: rules**



- **Requirements** (eg. Turing machine):

1. Representation of information (Tape)

Turing, A. M. *Proc. Lond. Math. Soc.* s2–42, 230–265 (1936).

2. Transformation (Head) : read, write, and change system's state (according to algorithm).
   - **IF** the device is in State `X` **AND** it reads Symbol `Y`,
   - **THEN** it should write Symbol `Z`, change to State `W`, and move the read/write mechanism (e.g., move left or right).
3. Control (Program/Rules): sequence operations

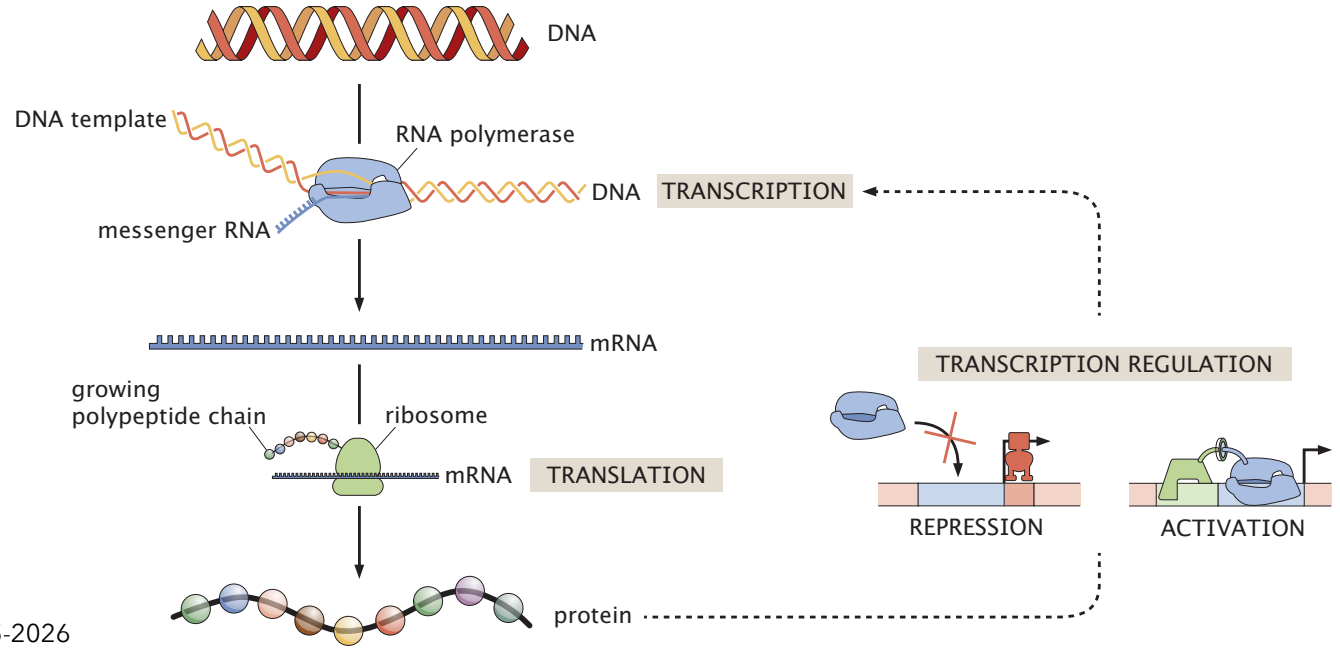
NB: It doesn't matter how it is instantiated (gears, transistors, etc)

Biological molecular computation

3. Means of Control (the "Program"):

IF a specific transcription factor (input) is present (**State X**),

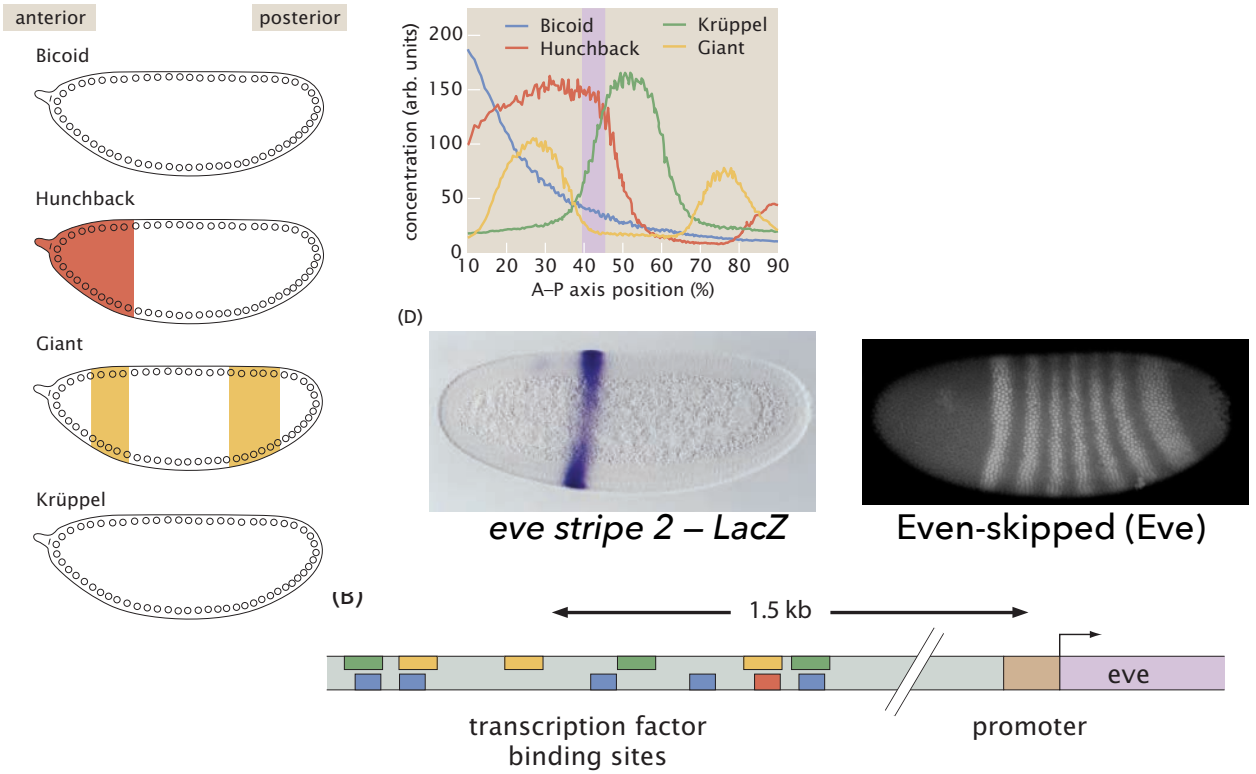
- **AND** it binds to a specific DNA sequence (**reads Symbol Y**),
- **THEN** it recruits RNA polymerase to **write** an mRNA molecule (**Symbol Z**), thereby changing the state of the cell.



Biological computation

3. Means of Control (the "Program"): The sequence of operations is controlled by **regulatory networks.**

- This is governed by gene regulatory networks and signalling pathways.



R. Phillips, J. Kondev, J. Thériot & H. Garcia.
Physical Biology of the Cell (Garland Science) 2012

Marr's Tri-level of analysis

Computational neurosciences

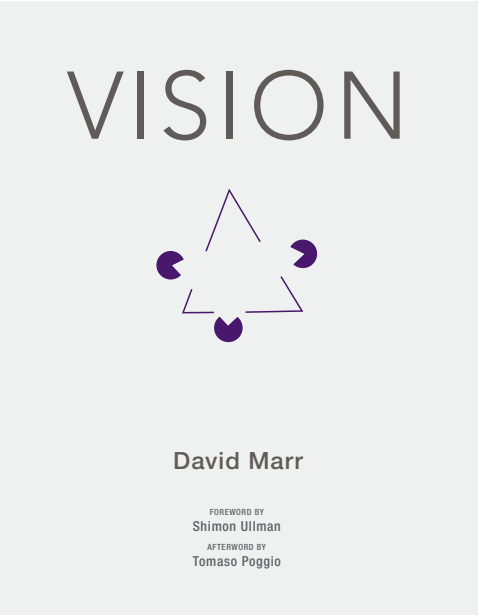
A framework to disentangle in complex processes:

- The *purpose/function* (why): *computational* level
- The *strategy* (how): *algorithmic* level
- The *biology/physics* (what): *implementational* level

VISION

A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr (1945-1980)



1982, Vision, David Marr
W. H. Freeman and Company
2010: *MIT press* (re-published)

Marr's Tri-level of analysis

1. The Computational Level

(« What for » or « Why »)

- **Goal:** Identify the abstract computational problem the system is solving

- **Questions:**
  - What is the task, the goal of the computation? What is the function?
  - What are the constraints imposed on the system (speed, energy consumption, sensitivity to noice etc)

- **Focus:** The **function** associated with a given process.

This level is about the *problem itself*, independent of any specific algorithm or hardware.

Marr's Tri-level of analysis

2. The Algorithmic & Representational Level (« How »)

- **Specify the representation** for the input and output

There is usually a wide range of possible representations.

  - **Decimal (54):** `5` tens + `4` ones = $(5 \times 10^1) + (4 \times 10^0)$
  - **Binary (110110):** `1` thirty-two + `1` sixteen + `0` eights + `1` four + `1` two + `0` ones = $(1 \times 2^5) + (1 \times 2^4) + (0 \times 2^3) + (1 \times 2^2) + (1 \times 2^1) + (0 \times 2^0)$
    So, $\mathbf{54_{10} = 110110_2}$
  - **Continuous** (real #, or physical quantities, eg. Volt, Pa) vs **Discrete** (eg. binary #) variables: analog vs digital computation.

The input and output need not have the same representation: ex. Fourier transform: time domain as input, frequency domain as output.

- **Information processing strategy.**
- **The algorithm (the step-by-step procedure) transforms the input representation into the output representation.**
- Defines the **rule-based processing**.

Different algorithms may exist for a given representation: choice depends on emphasis on eg. speed, robustness etc, or any other constraints defined in the computational level.
Could also be influenced by implementation.

Marr's Tri-level of analysis

3. The Implementational Level

- **Goal:** Explain how the algorithm is physically realized.
- **Questions:** How is the algorithm and the representations physically implemented in hardware? What are the biological or mechanical structures that carry out the computation?
- **Focus:** The physical and chemical substrate (e.g., neurons, synapses, silicon chips, proteins).

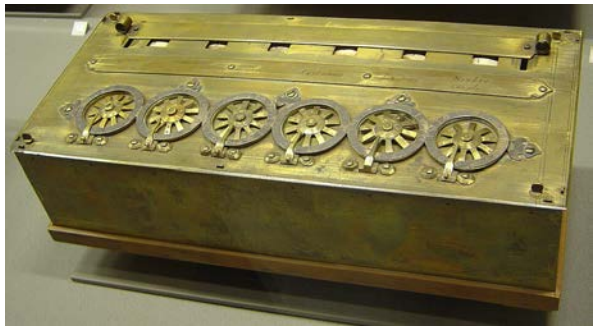
>>The 3 levels are largely independent.

Different physical implementations are possible:

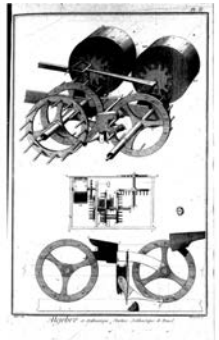
Mechanical computation, vs Electronic computation, vs Molecular computation, vs Cellular computation

Arithmetic
+ - x /

cm scale



Conservatoire national des arts et métiers, Paris

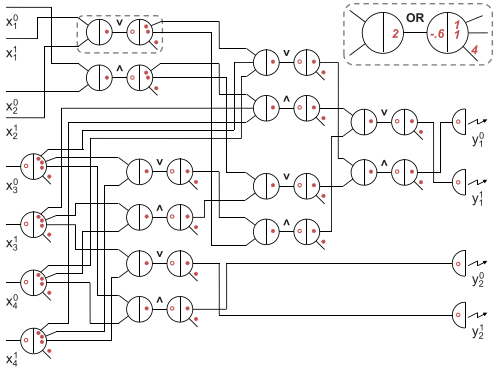


L'Encyclopédie de Diderot et d'Alembert

μm scale

(current transistors: few nm)

nm scale



$$y_2 y_1 = \lfloor \sqrt{x_4 x_3 x_2 x_1} \rfloor$$

Qian, L. & Winfree, E. *Science* 332, 1196–1201 (2011).

Marr's Tri-level of analysis

- Marr's level of analysis: The Cash Register
  - **Computational Level:** The goal is to perform arithmetic addition. The theory is that of addition (e.g., commutativity: 5+2 = 2+5, associativity, existence of a zero and inverse (sum equals zero)). The constraint is that the output must be the sum of the inputs.
  - **Algorithmic/Representational Level:**
    - ***Representation****:* Input numbers are represented as base-10 digits on a keypad or as digital binary values. The output is represented as a set of symbols on a paper tape or on an LCD screen.
    - ***Algorithm****:* The specific step-by-step procedure. For example, it could use the algorithm learned in school (add the rightmost digits, carry the one, etc.) or it could use a different algorithm implemented in binary logic (e.g., a ripple-carry adder circuit).
  - **Implementational Level:** This is the physical hardware: the specific arrangement of silicon transistors, logic gates, wires, and a power source that physically perform the binary addition algorithm.
    A mechanical cash register would use gears and levers.

1982, Vision, David Marr
W. H. Freeman and Company

Marr's Tri-level of analysis

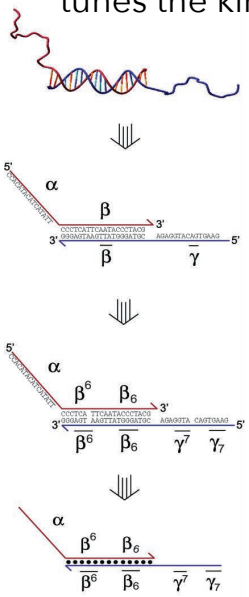
3. The Implementational Level: Molecular scale

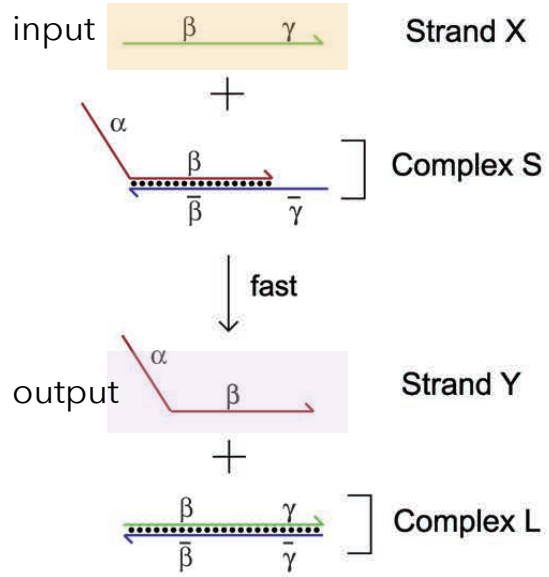
- **Goal:** Rational design of chemical devices (logic gates) capable of chemical computation using DNA.
- **Approach:**
  - Leverages the high predictability of Watson-Crick nucleotide pairing.
  - A DNA strand can serve as a **signal when it is free**, but is **inhibited when it is bound** to a complementary strand.
  - **Strand displacement** allows the production of an **output** ssDNA signal given an **input** ssDNA signal.
  - **Toehold**: a short single-stranded overhang region that initiates the strand displacement reaction and tunes the kinetics of strand displacement (via length of toehold).

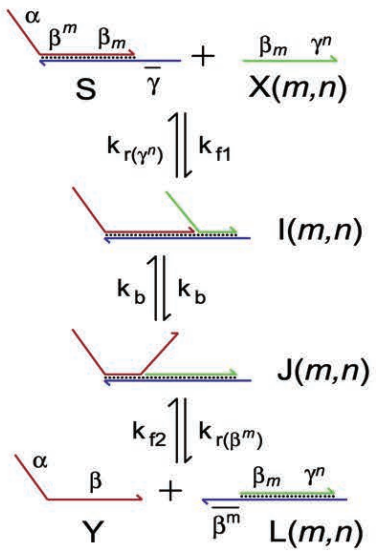
DNA molecule

DNA sequence

DNA abstraction







D. Y. Zhang, E. Winfree, *J. Am. Chem. Soc.* 131, 17303 (2009).

Marr's Tri-level of analysis

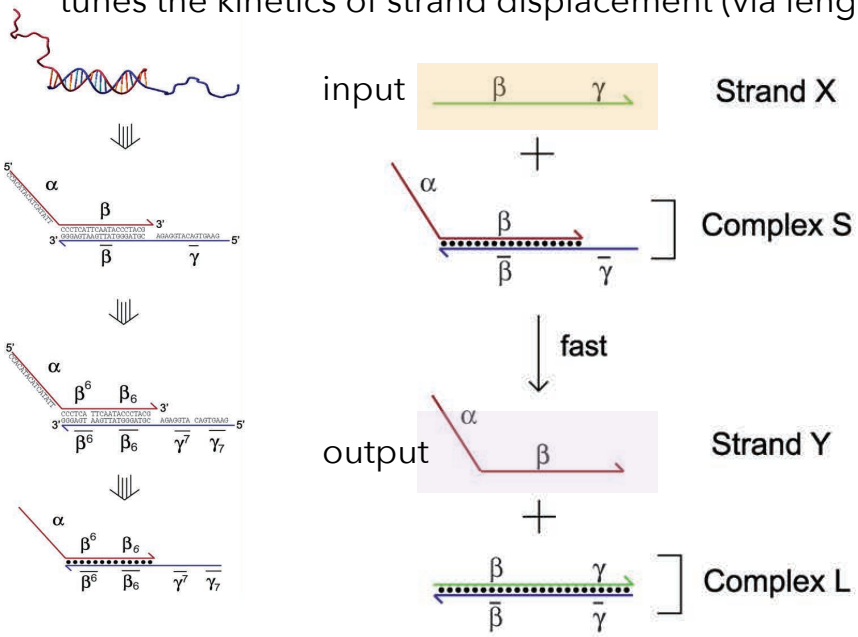
3. The Implementational Level: Molecular scale

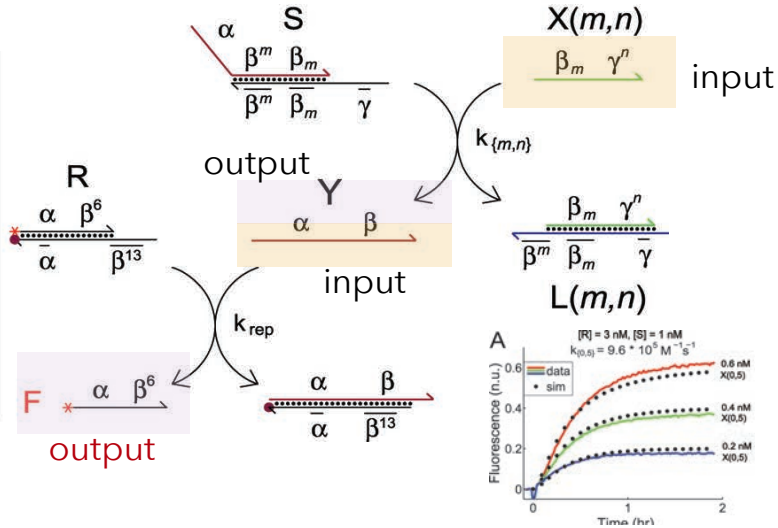
- **Goal:** Rational design of chemical devices (logic gates) capable of chemical computation using DNA.
- **Approach:**
  - Leverages the high predictability of Watson-Crick nucleotide pairing.
  - A DNA strand can serve as a **signal when it is free**, but is **inhibited when it is bound** to a complementary strand.
  - **Strand displacement** allows the production of an **output** ssDNA signal given an **input** ssDNA signal.
  - **Toehold**: a short single-stranded overhang region that initiates the strand displacement reaction and tunes the kinetics of strand displacement (via length of toehold).

DNA molecule

DNA sequence

DNA abstraction



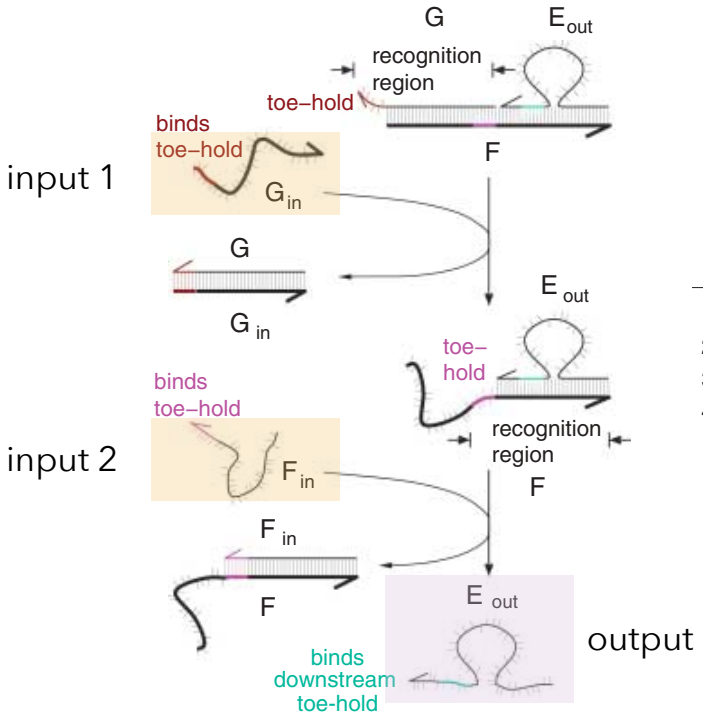


D. Y. Zhang, E. Winfree, *J. Am. Chem. Soc.* 131, 17303 (2009).

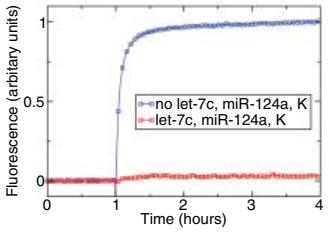
Marr's Tri-level of analysis

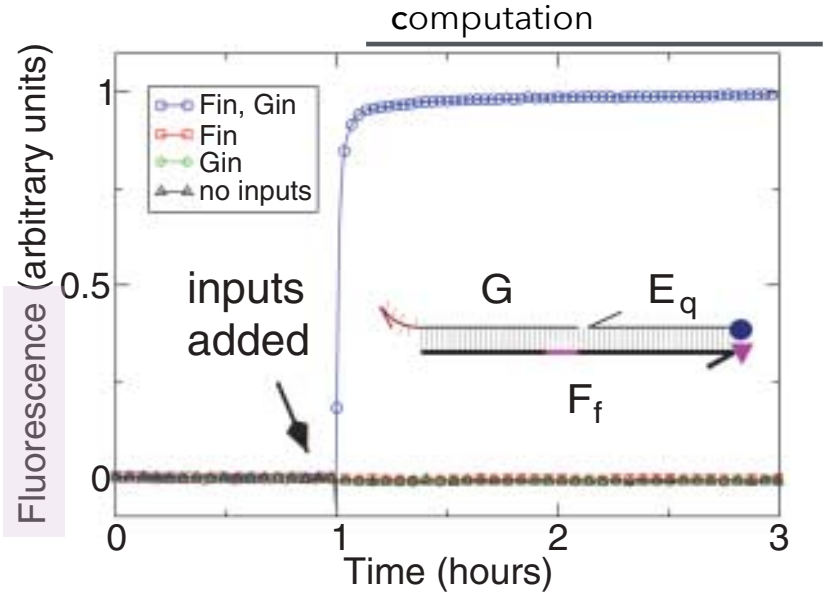
3. The Implementational Level: Molecular scale

- **Design of a Two-input molecular AND gate:** gate function fully determined by base pairing and breaking
- **Implement full set of Boolean-logic functions (AND, OR and NOT): cascading and multilayering.**
- To avoid degradation, signal is restored by threshold and amplifying gates
- Watson-Crick interactions between modular interaction domains determine the connectivity of gates.



	REACTANTS	PRODUCTS
1	EFG, NO INPUTS	SAME AS REACT.
2	EFG + F_{in}	SAME AS REACT.
3	EFG + G_{in}	EF + GG_{in}
4	EFG + F_{in} + G_{in}	E + FF_{in} + GG_{in}

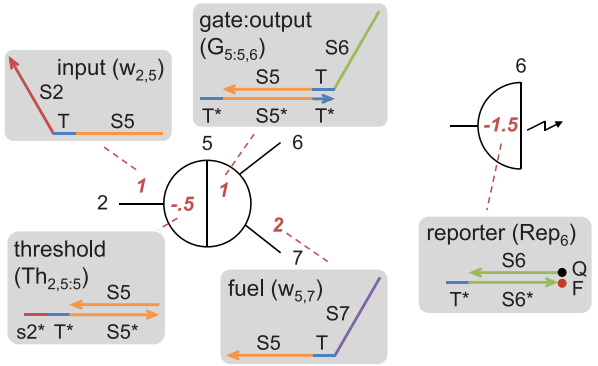




G. Seelig, D. Soloveichik, D. Y. Zhang, E. Winfree, *Science* 314, 1585 (2006).

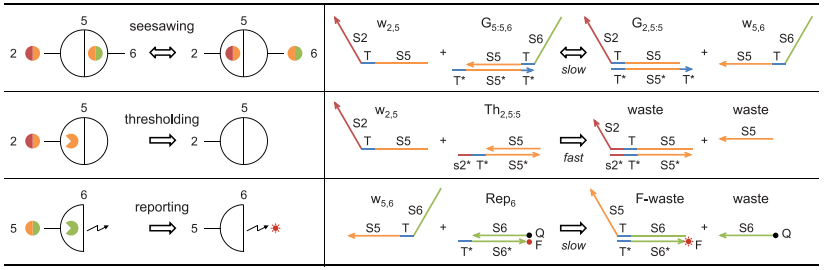
Marr's Tri-level of analysis

3. The Implementational Level: molecular scale

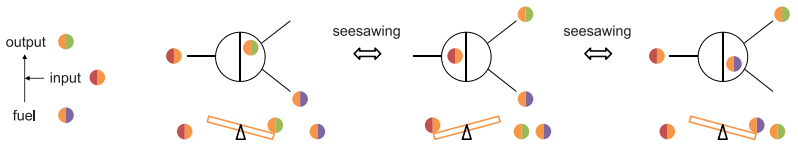


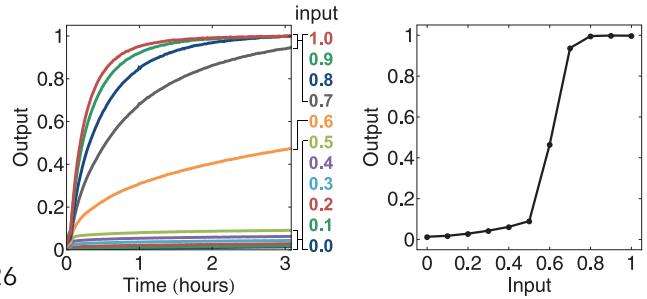
- Design of a seesaw gate:
  - **reversible reaction** by exchanging the activity of DNA signals.
  - a pair of see-saw gates can perform and Boolean logic function.
  - gate represented by 2-sided node and wires represent DNA signals.
  - Signals can be input, output or fuel (gate signals that catalyse output)

- Three basic functions:
  - **Seesawing:** reversible ***slow*** reaction
  - **Thresholding:** ***fast*** dead-end reaction that competes with seesawing.
    Seesawing occurs *if and only if* signal above threshold
  - **Reporting:** does not compete with seesawing



- **Catalytic cycle:** input signal can transform free fuel into output without being consumed
- **Thresholding** transforms analog signal into digital response



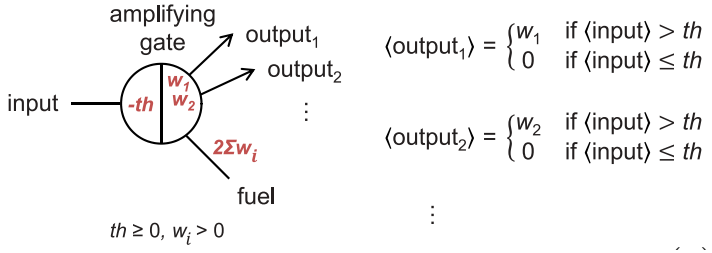


Qian, L. & Winfree, E. *Science* 332, 1196–1201 (2011).
QIan, L. & Winfree, E. *J. R. Soc. Interface* 8, 1281–1297 (2011)

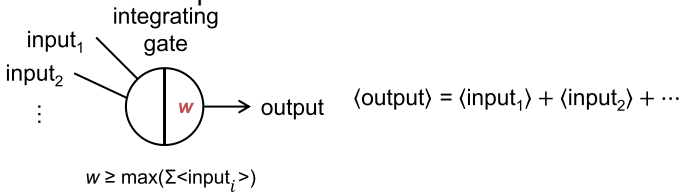
Marr's Tri-level of analysis

3. The Implementational Level: Molecular scale

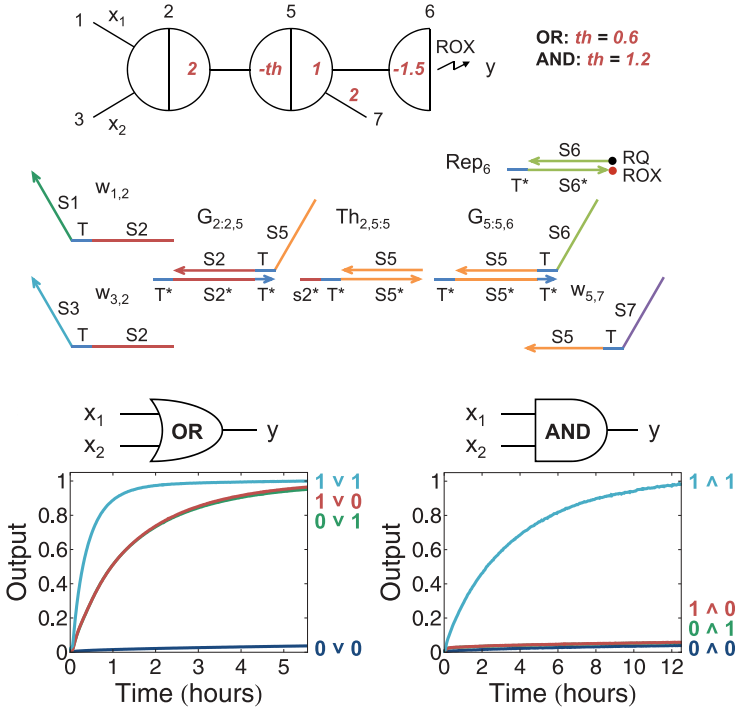
- **Amplifying gate:** fan-out
  output increases up to maximum if input above threshold



- **Integrating gate:** fan-in
  output released stoichiometrically with input.
  output is sum of inputs.



- Building logical AND/OR gates from cascades of seesaw gates
  Integrating gate/Amplifying gate/Reporter gate
  OR gate: low threshold
  AND gate: high threshold: requires both inputs to pass threshold

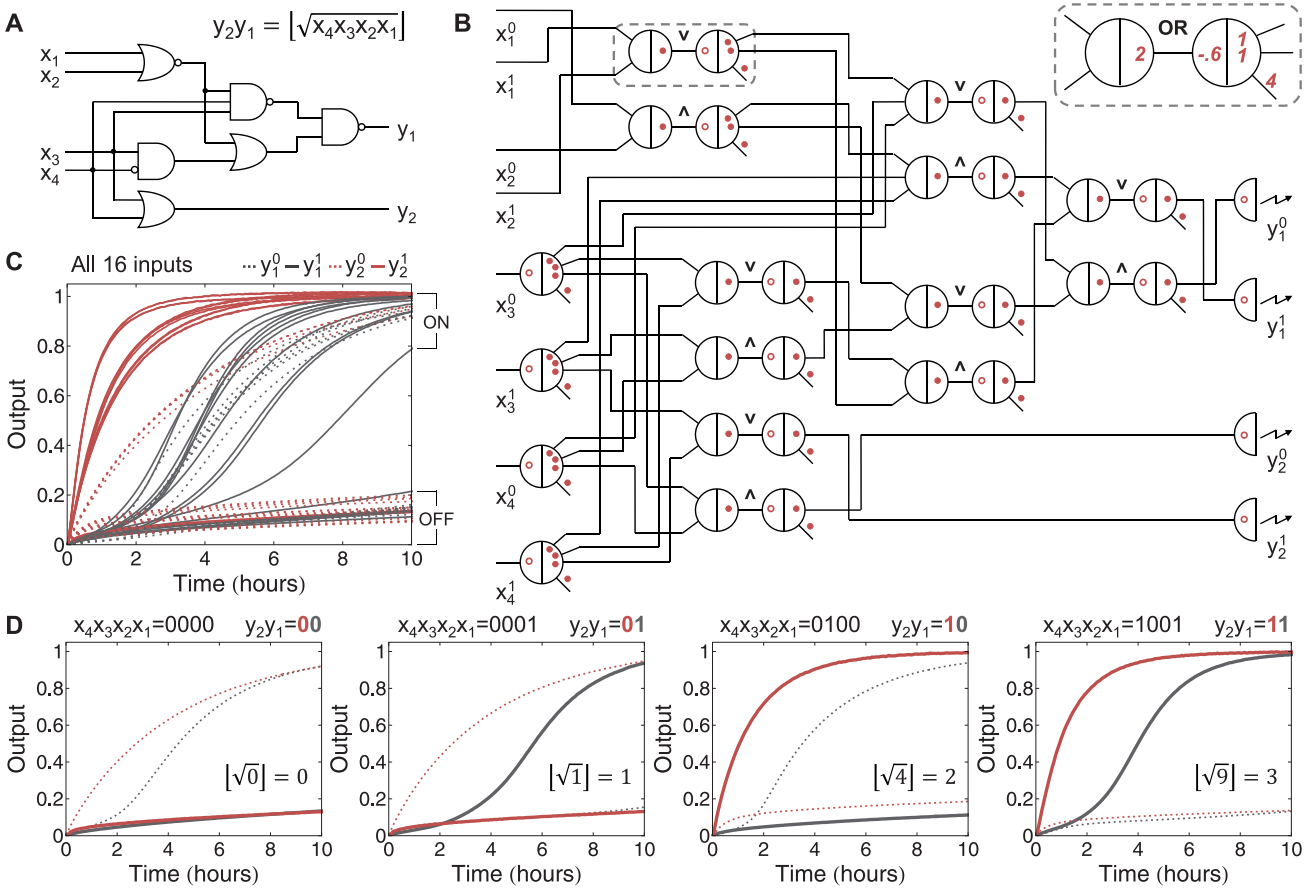


Qian, L. & Winfree, E. *Science* 332, 1196–1201 (2011).

Marr's Tri-level of analysis

3. The Implementational Level: Molecular scale

A digital circuit that computes the floor of the square root of a four-bit binary number!



Qian, L. & Winfree, E. *Science* 332, 1196–1201 (2011).

Marr's Tri-level of analysis

Computational neurosciences

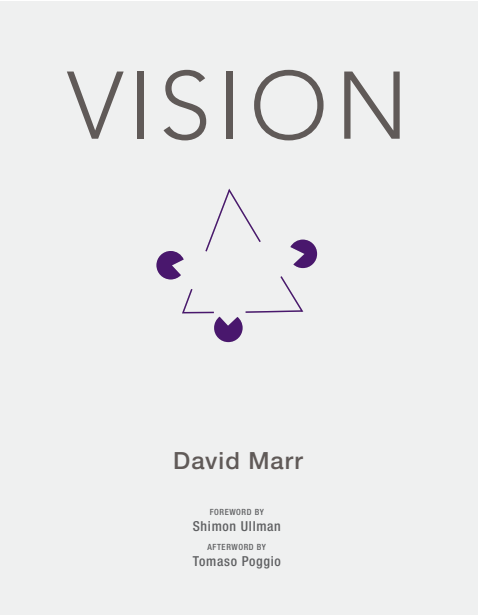
A framework to disentangle in complex processes:

- The *purpose/function* (why): *computational* level
- The *strategy* (how): *algorithmic* level
- The *biology/physics* (what): *implementational* level

VISION

A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr (1945-1980)



1982, Vision, David Marr
W. H. Freeman and Company
2010: *MIT press* (re-published)

Marr's Tri-Level in *Bacteria metabolism*

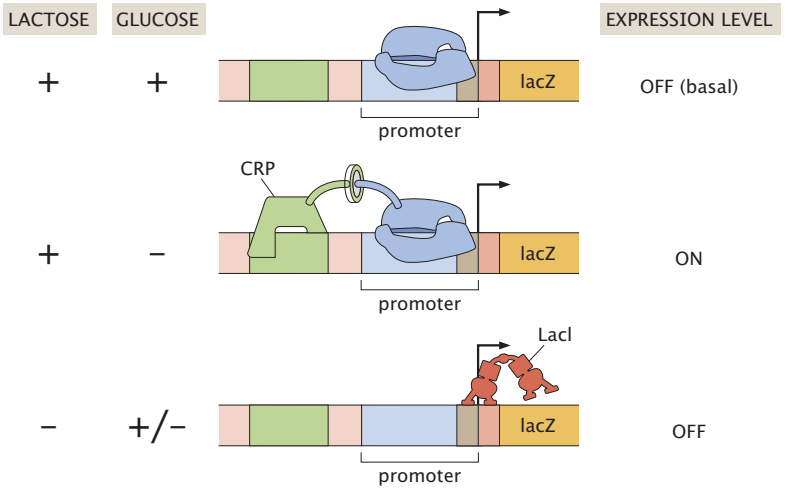
The Lac Operon in *E. coli*, a Molecular Logic Gate.

- **The Problem:** Should a bacterium produce enzymes to digest lactose?
- **The Inputs:**
  - Input 1: Is lactose present?
  - Input 2: Is glucose present?

- **The Algorithm:**

```
IF (Lactose is PRESENT) AND (Glucose is ABSENT) THEN PRODUCE
digestive enzymes.
ELSE, DO NOTHING.
```

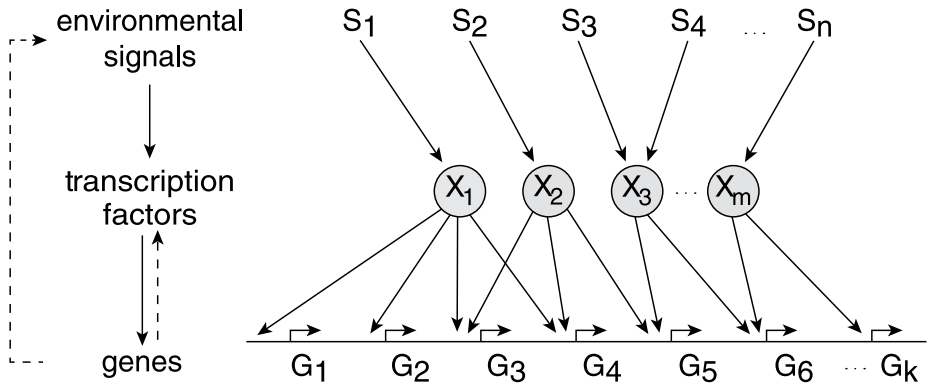
- **The Biological Implementation:**
  - A repressor protein blocks the genes for the lactose enzymes (the "OFF" state).
  - **IF** lactose is present, it binds to the repressor, pulling it off the DNA. This is like a NOT gate: the presence of lactose removes a "stop" signal.
  - **BUT**, the gene still needs a "GO" signal. This comes from the protein CRP, which is only active **IF** glucose is absent (NOT gate on the glucose signal).
  - The system requires both the "stop signal" to be removed *and* the "go signal" to be present—a logical AND operation.



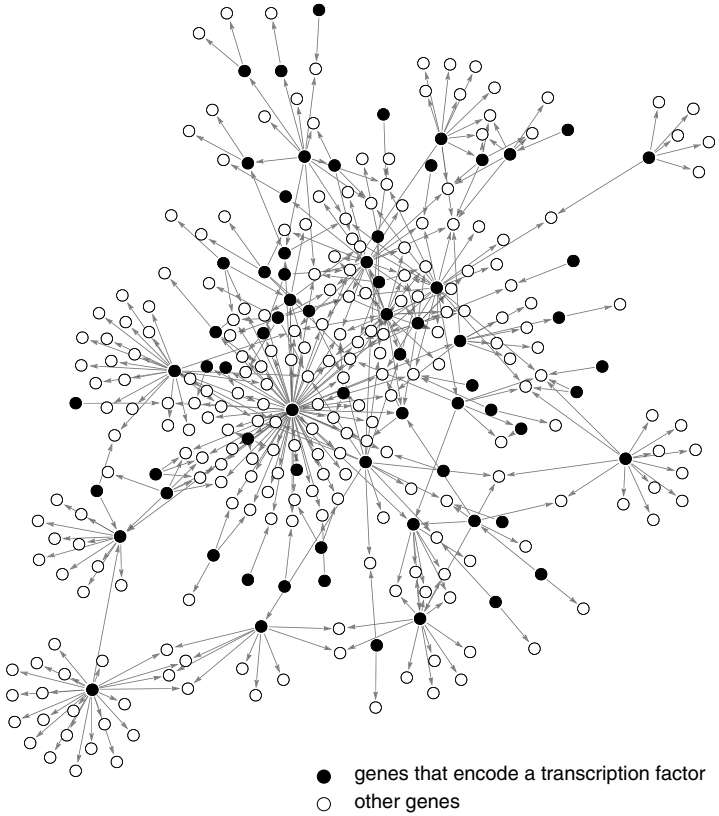
Hernan G. Garcia and Rob Phillips, *Physical genomics*

Marr's Tri-Level in *Bacteria metabolism*

- Transcription factors « symbols » that provide an internal representation of the environment.
- 300 TFs provide a 300-dimensional representation of the world to regulate ~4500 genes.



U. Alon. *An introduction to systems biology*. CRC Press (2020)



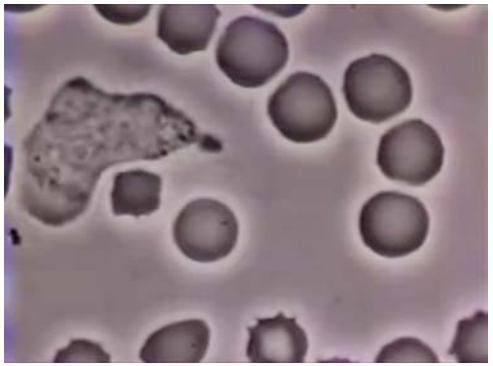
Hernan G. Garcia and Rob Phillips, *Physical genomics*

Marr's Tri-Level in *Chemotaxis*

- Computation: function/purpose
  - Read a gradient of concentration of a molecule
  - Orient motility towards the high concentration of the molecule
  - Keep animal in molecule-rich environment.
  - Or opposite: avoid toxic molecule and move away.
  - Constraints: cell size

- Algorithm: strategy/solution
  - Read a difference in concentration in space (for large enough cells)
    - symmetry breaking (instability via feedback)
  - Read a difference in concentration in time (for small cells eg. Bacteria)
    - cellular sensing, adaptation, memory and comparison.

- Implementation:
  - Molecular detection
  - Amplification, adaptation and memory

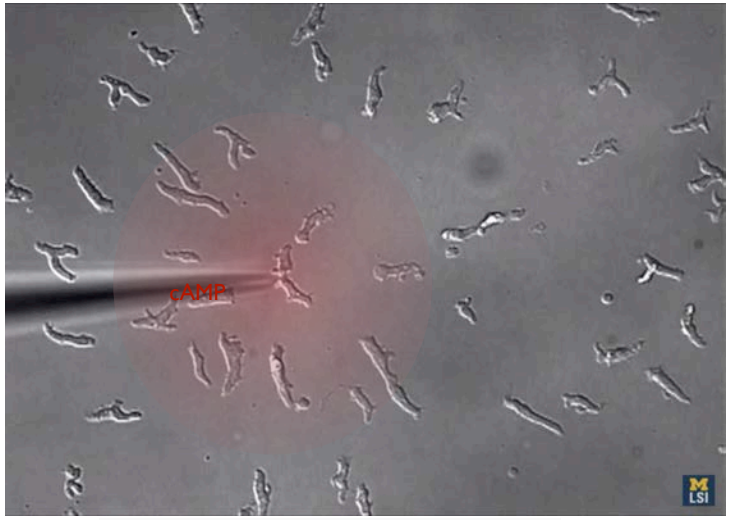


https://www.youtube.com/watch?v=I_xh-bkiv_c

David Rogers at Vanderbilt University.

Marr's Tri-Level in *Chemotaxis:* Algorithm

Eukaryotic cells read and respond to a spatial gradient of chemokine
Cell polarisation in response to gradient

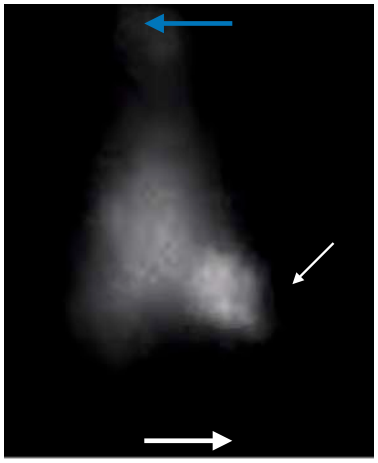


Carole Parent lab at the University of Michigan Life Sciences Institute.

Dictyostelium discoideum cells are attracted by cAMP released in a gradient from a pipette

PIP3 sensor
(PH-GFP)

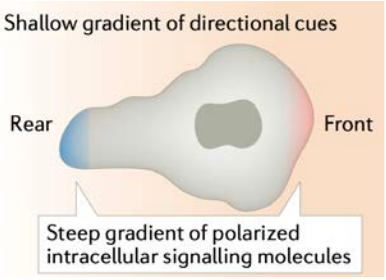
cAMP wave



cell movement

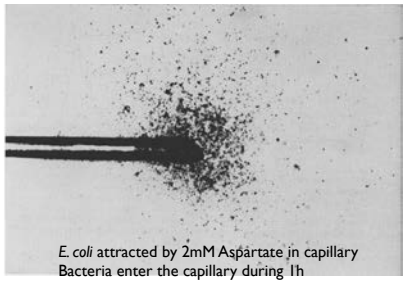
P. Devreotes lab

signal sensing
receptor
transduction network
cell polarization
mechanical response



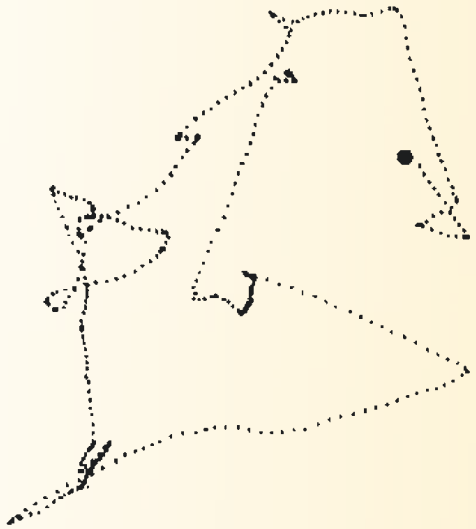
Marr's Tri-Level in *Chemotaxis:* Algorithm

Chemotaxis in Bacteria entails detection of a temporal gradient



E. coli attracted by 2mM Aspartate in capillary
Bacteria enter the capillary during 1h

R. Macnab. D.E. Koshland. *PNAS*. 69:2509-2512 (1972)

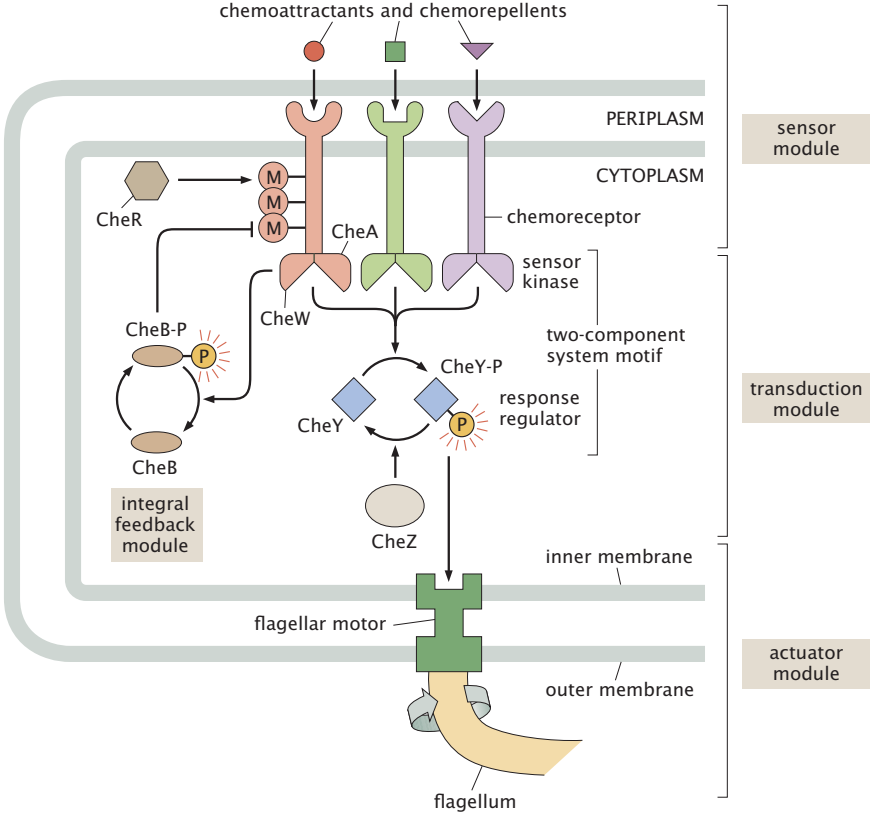


Problem: Bacteria can go up an exponential gradient, over 20mm.
For a 2µm cell to detect such a gradient, they would need to detect 0.0001% difference on both ends

- Bacteria are too small to detect spatial gradient of concentration
- Bacteria detect a temporal change in concentration of chemoattractant
- As they navigate in space, they detect in time different concentrations
- This requires comparison of 2 measurements and memory

Marr's Tri-Level in *Chemotaxis:* Algorithm

Chemoreceptor physiology

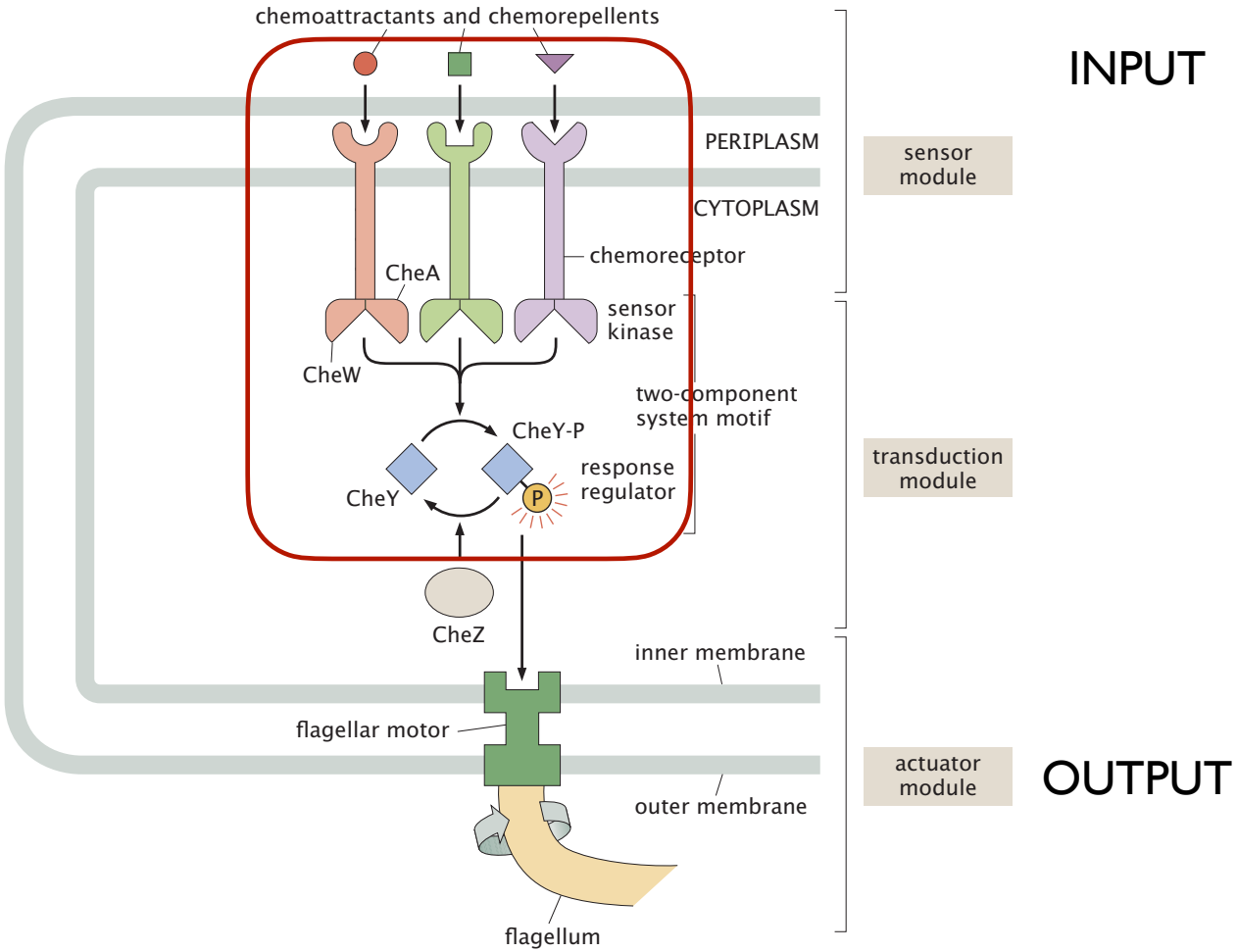


Key properties of chemotactic network

- Sensitivity - Gain : output/input ratio
- Adaptation: reset after input
- High amplitude range

Molecular circuit driving chemotaxis

- Sensitivity



R. Phillips, The Molecular Switch: signaling and allostery. *Princeton Univ. Press*. 2020

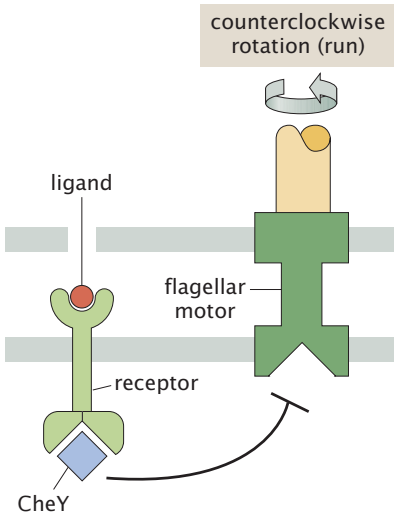
Molecular circuit driving chemotaxis

Regulation of tumbling frequency

- **Ligand binding:** CheY is unphosphorylated and does not signal: Receptor is OFF, rotation is CCW
- **No ligand:** CheY is phosphorylated, signals: Receptor is ON, rotation is CW
  CheY-P binds the motor and reverses rotation

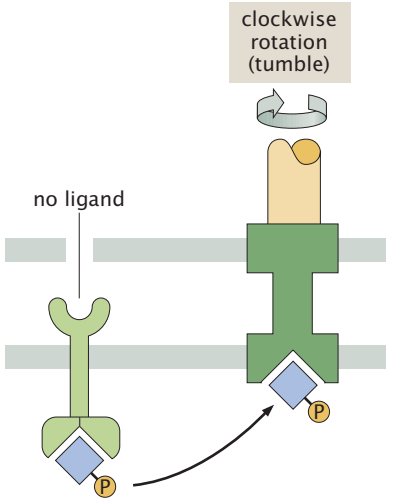
Activity of receptor (R and CheA) $a(t) = f[L(t)]$ L : Ligand

	Ligand: bound
INPUT	Receptor OFF
OUTPUT	Motor CCW
	Run



Ligand: no
Receptor ON

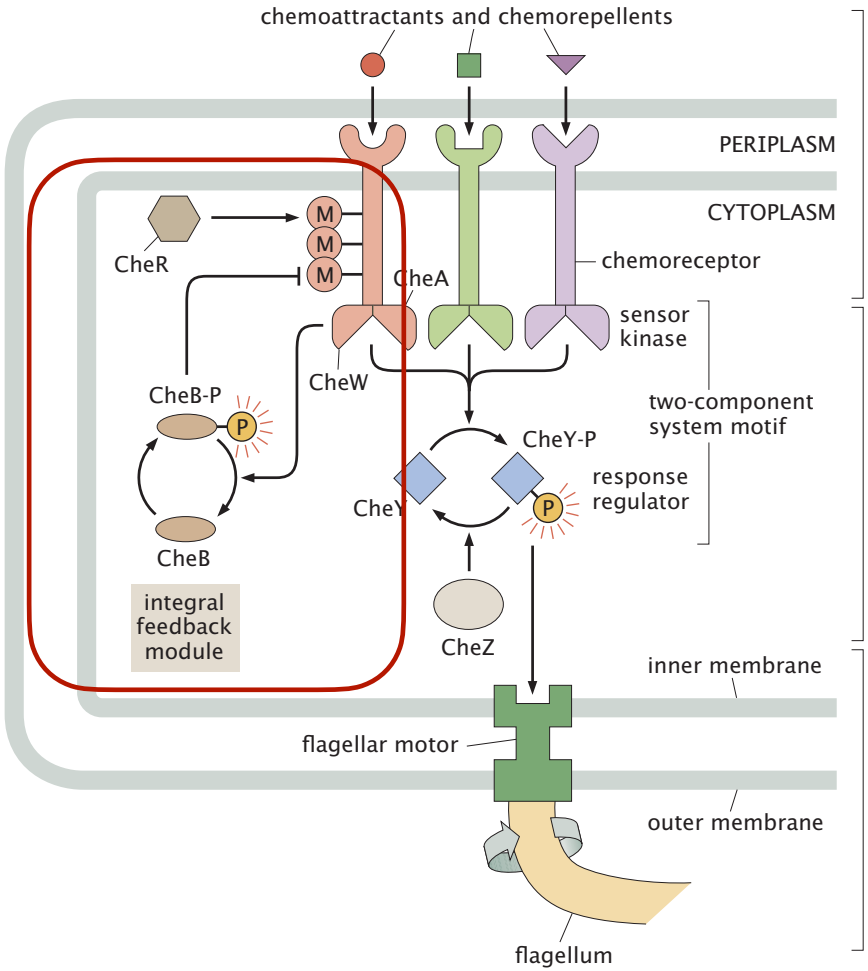
Motor CW
Tumble



R. Phillips, The Molecular Switch: signaling and allostery. *Princeton Univ. Press*. 2020

Molecular circuit driving chemotaxis

- Adaptation



INPUT

Activity of receptor (R and CheA)

$$a(t) = f[L(t), m(t)]$$

L : Ligand

m: methylation state

More methylation → higher CheA autophosphorylation rate

OUTPUT

R. Phillips, The Molecular Switch: signaling and allostery. *Princeton Univ. Press*. 2020

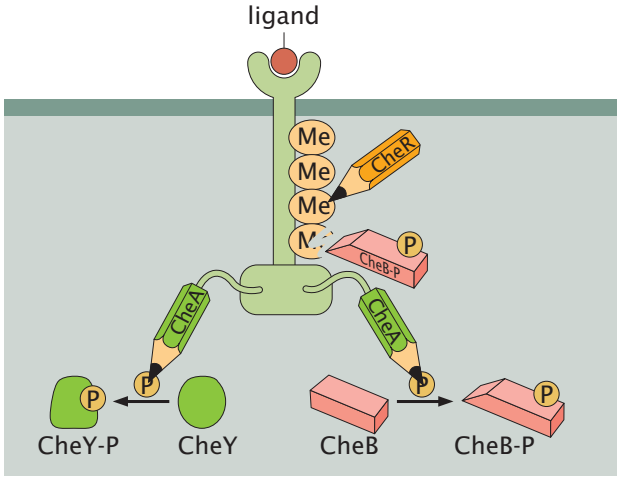
Logic of adaptation

- Cells have a built-in short term memory to compare present and recent past and thereby read the concentration gradient
- Methylation and demethylation take a few seconds and thus reflect receptor activity a few seconds ago (« memory »).
- Receptor occupancy by ligand influences the current activity state (which takes a fraction of a second).
- By comparing the activity state of the cell (CheA) and methylation, the cell can compute how signal evolved in a few seconds, whether it increased, or decreased.

$$a(t) = f[L(t), m(t)]$$

$L(t)$

Fast (<<1s)



$m(t)$

Longer (2-3 s)

Logic and implementation of adaptation

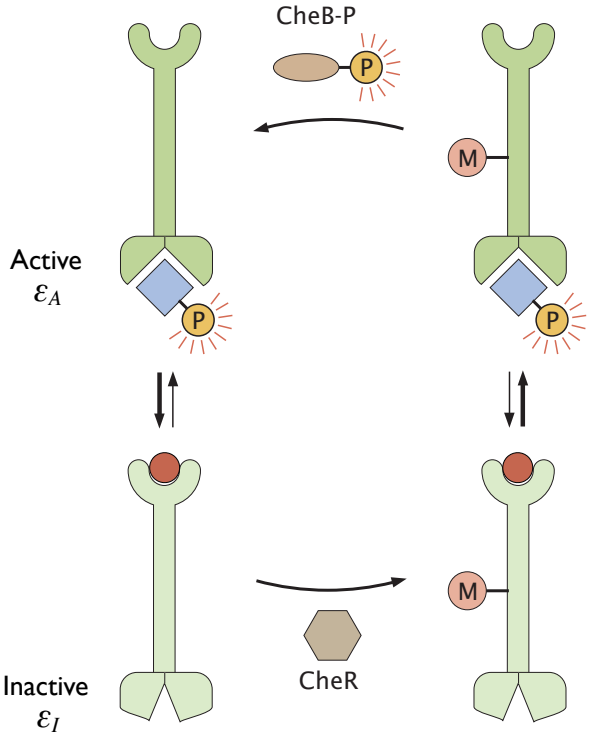
- Core of adaptation: Resetting

  - **Methylation is updated by receptor activity (and ligand binding):** Methylation is mediated by CheR.

    The demethylase CheB is activated via phosphorylation by CheA, and thus by ligand binding to the receptor.

  - Without ligand L, the receptor is on, CheA is active and so CheB is more active as a demethylase, and will convert the receptor into the off state. **The response is damped: the activity *a(t)* decreases.**

  - Conversely, with ligand L, the receptor is off, CheB is less active, methylation accumulates, and the receptor becomes on. Its **activity *a(t)* increases.**



  - **Integral feedback**
  - Error between activity and set point.

$e(t)=a(t)-a^*$ a^* (set-point, corresponds to average methylation state. Encoded in CheR/CheB activities on R)

It integrates this error over time into the methylation level $m(t)$: slow variable that measures the past history.

- If activity $a(t)$ too high → decrease $m(t)$ $\quad dm/dt = k_R\,(1-a) - k_B\,a$
- If $a(t)$ too low → increase $m(t)$

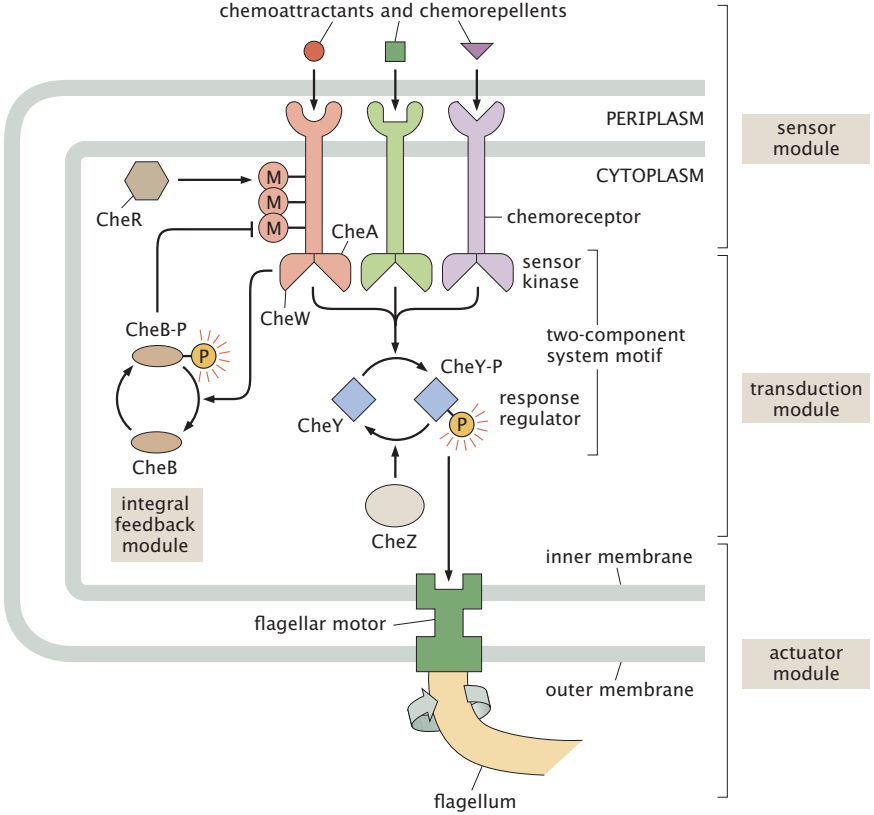
$a(t) = f[L(t), m(t)]$

instantaneous slow integral of past

R. Phillips, The Molecular Switch. *Princeton Univ. Press*. 2020

Marr's Tri-Level in *Chemotaxis:* Algorithm

Chemoreceptor physiology



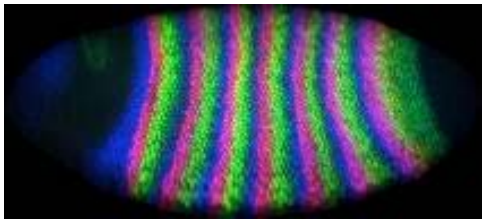
Key properties of chemotactic network

- Sensitivity - Gain : output/input ratio
- Adaptation: reset after input
- High amplitude range

Spatial patterning across scales

- Embryo segmentation

0.5 mm – 1 hour



Shinji Takada

- Plumage/pigmentation pattern

0.5 mm – 2 days

Interface Focus (2012) 2, 433–450 doi:10.1098/rsfs.2011.0122

Tony Hisgett/Wikipedia

5 mm – 10 days

Marr's Tri-Level in developmental ***Patterning***

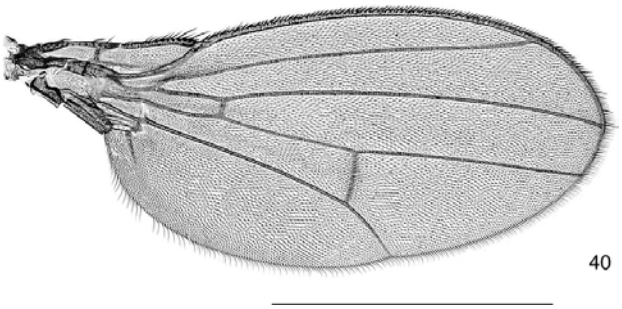
- **Morphogen gradient**

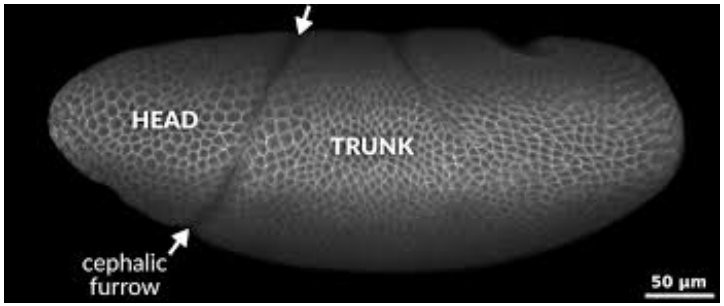
- **Computation: function/purpose**
  - Define robustly discrete regions with clear spatial boundaries in a field of cells
  - *Patterning has the property of scaling (keep proportions as size varies)*
- **Algorithm: strategy/solution**
  - Produce a spatial gradient of a continuous input variable across the field
  - Read the value of this variable
  - Respond non-linearly to input to produce a sharp spatial boundary in output value
  - *Scaling algorithm: eg. expansion-repression integral feedback control* M $\overset{D}{\dashv}$ Expander
- **Implementation:**
  - Chemical gradient: eg. Exponential decay profile with local source, diffusion and degradation.
  - Control of molecule diffusion/transport and degradation to tune the length scale of gradient
  - Cooperative effect in molecular response network: eg. Transcription factor binding on DNA promoter/enhancer sequences.
  - *Scaling: not clear yet…*

Marr's Tri-Level in developmental *Patterning*

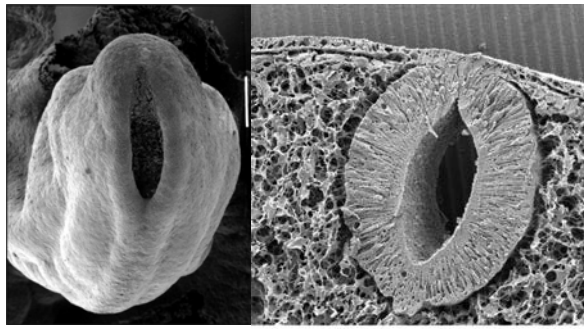
- **Morphogen gradient: Computational level**

  Define discrete regions with clear spatial boundaries in a field of cells

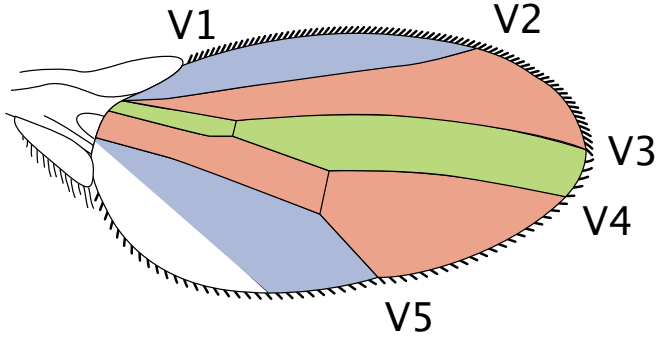


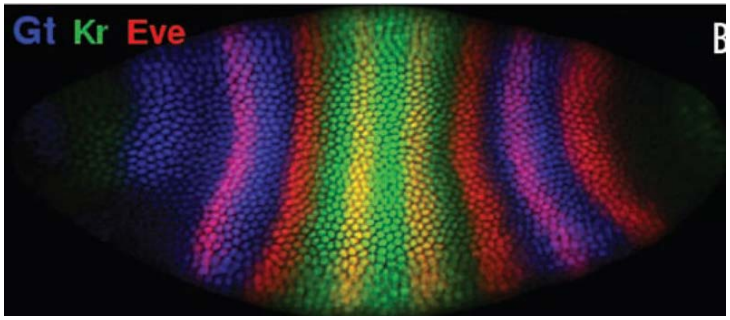


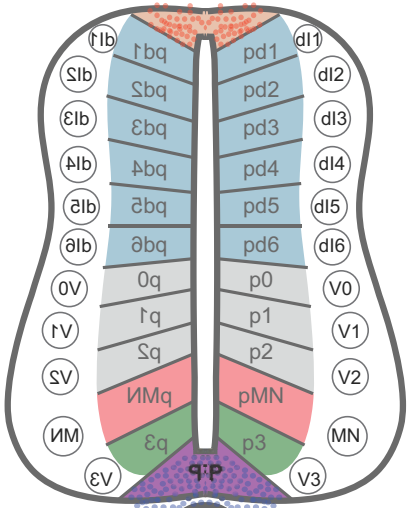
https://brunovellutini.com/posts/cephalic-furrow-thread/



embryology.med.unsw.edu.au/ Eric Théveneau Toulouse CBI



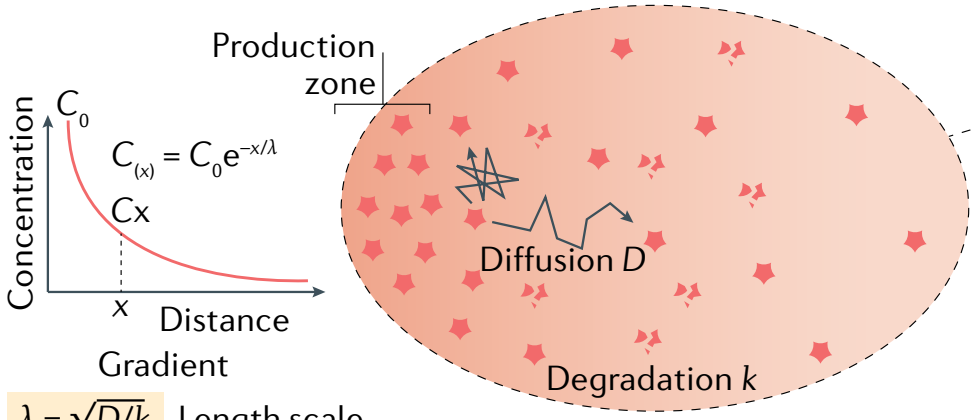




Marr's Tri-Level in developmental ***Patterning***

- Morphogen gradient: Algorithmic level

Gradient of continuous variable

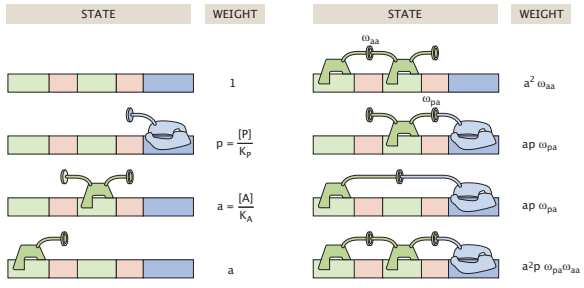


$\lambda = \sqrt{D/k}$ Length scale

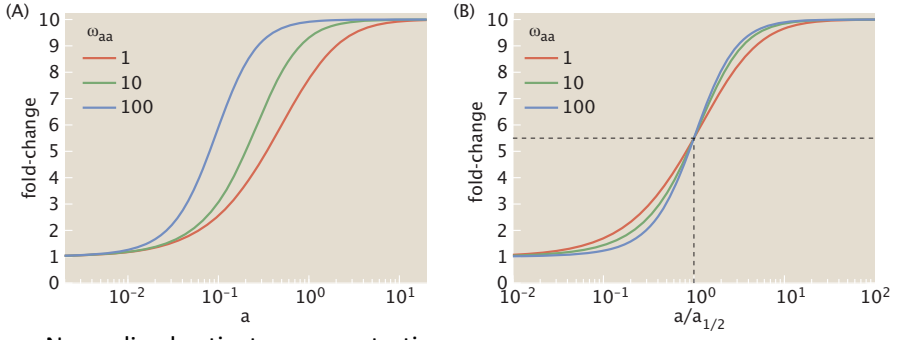
$\tau = \lambda^2/D$ Timescale

Collinet C. & Lecuit T. *Nature Rev. Mol. Cell Biol.*, 2021
doi.org/10.1038/s41580-020-00318-6

Non-linear response and partitioning



$$\omega = e^{-\beta\varepsilon_{int}}$$



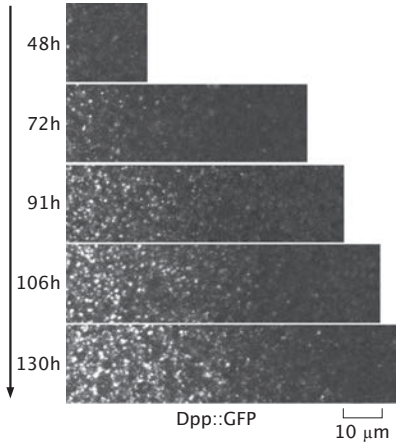
Normalised activator concentration a

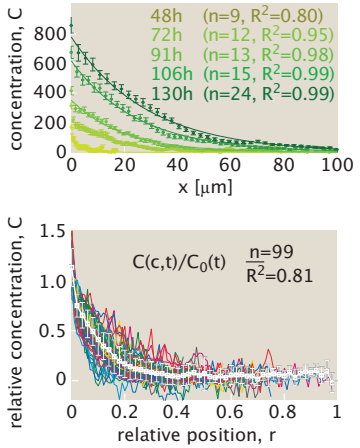
Hernan G. Garcia and Rob Phillips, *Physical genomics*

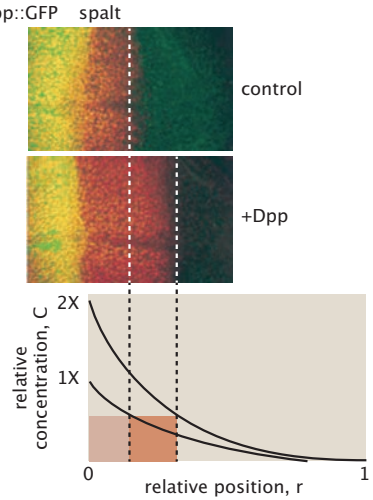
Marr's Tri-Level in developmental *Patterning*

- Morphogen gradient: Implementational level

Wing: BMP molecular gradient and response







mass conservation	$\partial_t C = \underbrace{D\nabla^2 C}_{\text{diffusion}} - \underbrace{kC}_{\text{degradation}} + \underbrace{2j_0\delta(x)}_{\text{production}}$
steady state solution	$C(x) = C_0 e^{-\frac{x}{\lambda}}$
decay length	$\lambda = \sqrt{D/k}$
source concentration	$C_0 = j_0\sqrt{D/k}$

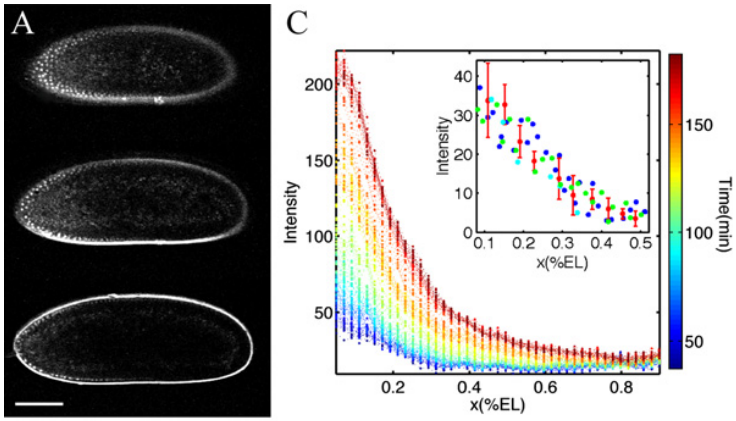
$D = 0.10 \pm 0.05\ \mu m^2/s$

$k = 2.52 \times 10^{-4} \pm 1.29 \times 10^{-4}\ s^{-1}$

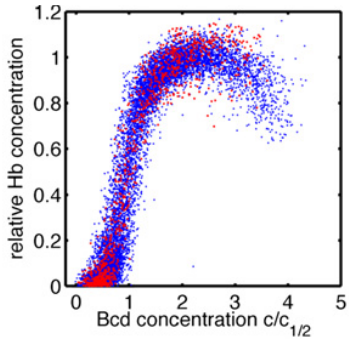
$j_0 = 3.98 \pm 2.34$ molecules/(µm × s)

A. Kicheva et al, *Science* 2000
A. Teleman et al, *Cell* 2000

Embryo: Bicoid gradient and response



T. Gregor et al and D. Tank. *Cell* 130, 141–152 (2007)

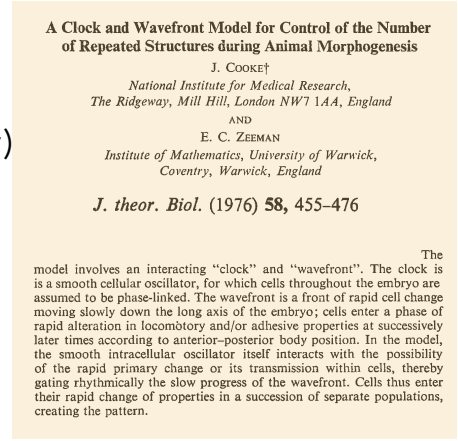


T. Gregor et al and W. Bialek. *Cell* 130, 153–164, 2007

Marr's Tri-Level in developmental *Patterning*

- Other algorithms: clock and wave front

- **Computation: function/purpose**
  - Define discrete regions with clear spatial boundaries in a field of cells
- **Algorithm: strategy/solution**
  - Clock and wavefront model
    - **Wave front** of *sudden* cell state changes (discontinuity)
    - **Clock: smooth oscillation** of phase-linked cells
    - **Slow posterior movement** of the wave front

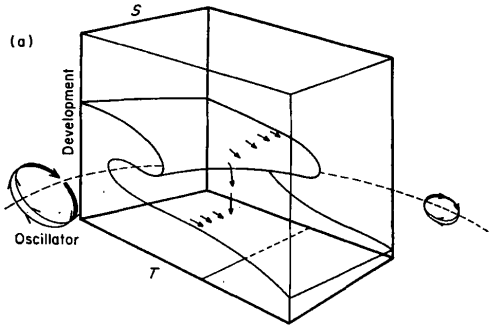


A Clock and Wavefront Model for Control of the Number of Repeated Structures during Animal Morphogenesis

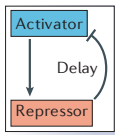
J. Cooke†
National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England
AND
E. C. Zeeman
Institute of Mathematics, University of Warwick, Coventry, Warwick, England

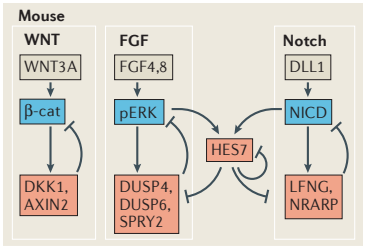
J. theor. Biol. (1976) **58,** 455–476

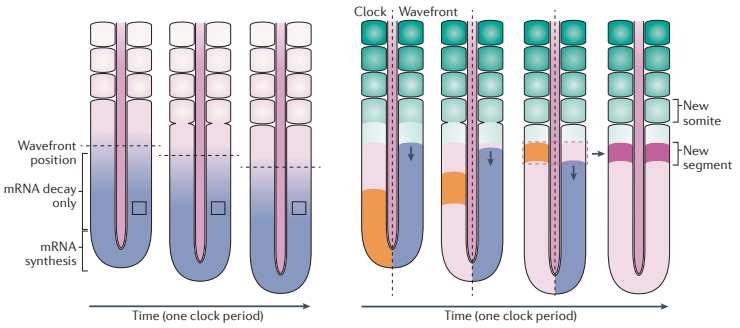
The model involves an interacting "clock" and "wavefront". The clock is is a smooth cellular oscillator, for which cells throughout the embryo are assumed to be phase-linked. The wavefront is a front of rapid cell change moving slowly down the long axis of the embryo; cells enter a phase of rapid alteration in locomotory and/or adhesive properties at successively later times according to anterior–posterior body position. In the model, the smooth intracellular oscillator itself interacts with the possibility of the rapid primary change or its transmission within cells, thereby gating rhythmically the slow progress of the wavefront. Cells thus enter their rapid change of properties in a succession of separate populations, creating the pattern.



- **Implementation:**
  - Chemical oscillator
  - Molecular gradient
  - Tissue growth

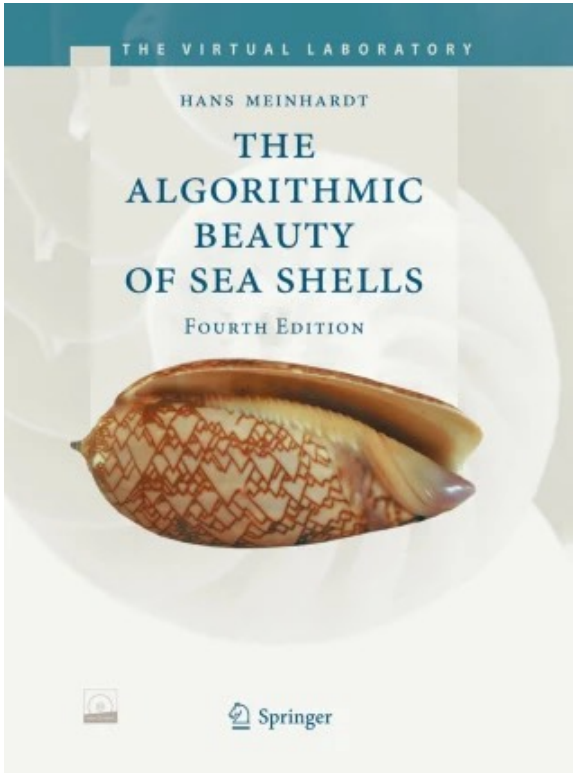




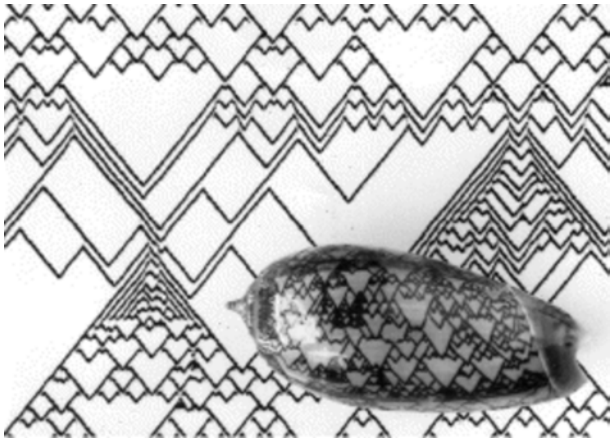


Marr's Tri-Level in developmental *Patterning*

- Self-organised spatial and temporal instabilities



Hans Meinhardt (1938-2016)



Marr's Tri-Level in developmental *Patterning*

- Turing-like spatial instabilities

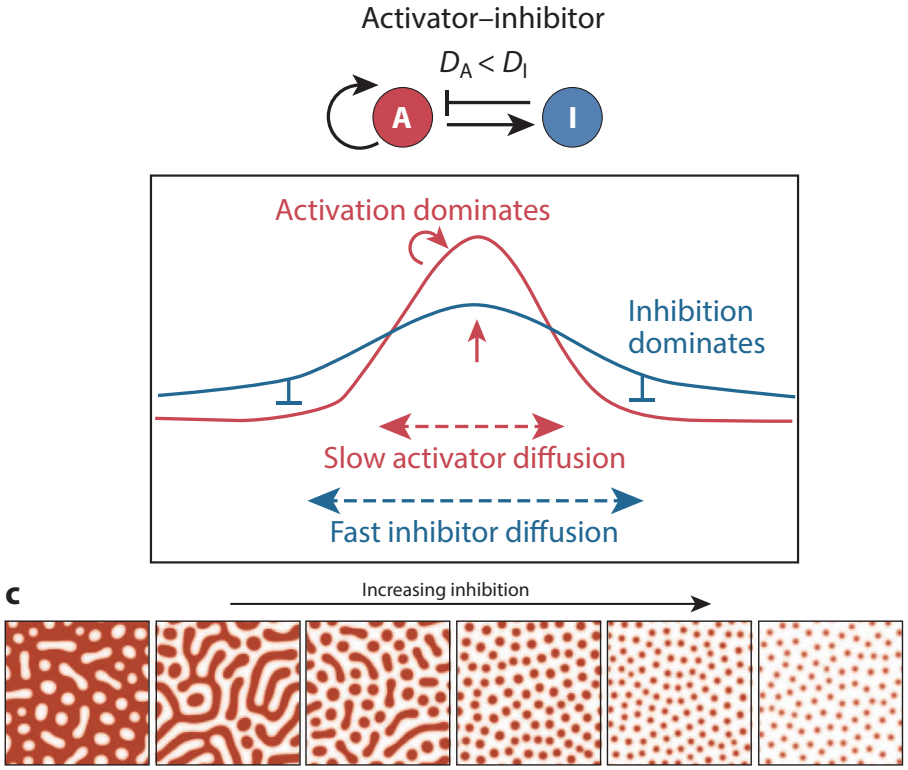
- Computation: function/purpose
  - Generate patterns across a field without scaling
  - Do not specify specific location but rather *spacing* properties

- Algorithm: strategy/solution
  - Local excitation/Non-local inhibition
  - Local positive feedback and long-range repression

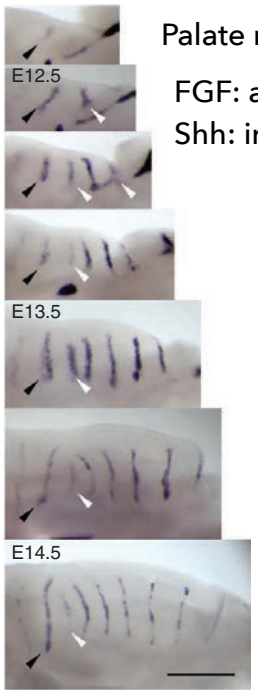
- Implementation:
  - Chemical: reaction diffusion (Turing patterns)
  - Mechanical: Local positive feedback of contractility
    Long range repression: friction, viscosity, or tension
  - Neural network: Local activation, Inhibition at a distance

Marr's Tri-Level in developmental ***Patterning***

- Chemical implementation: Turing reaction-diffusion



Bailles A, Gehrels EW, Lecuit T. *Annu Rev Cell Dev Biol*. 38:321-347 (2022)

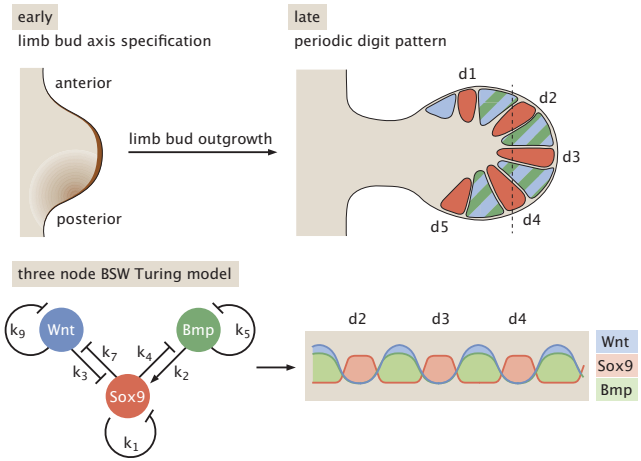


Palate ridges

FGF: activator

Shh: inhibitor

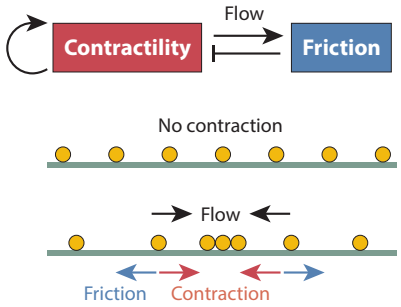
Economou AD, et al. & JBA. Green *Nat Genet*. 44(3):348–51 (2012)

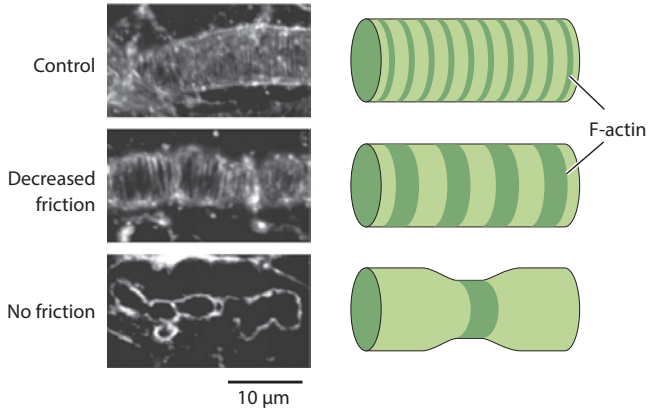


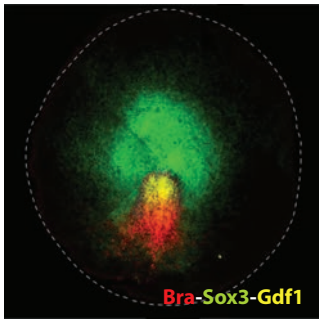
J. Raspopovic et al. and J. Sharpe. *Science* 345, 566 (2014)

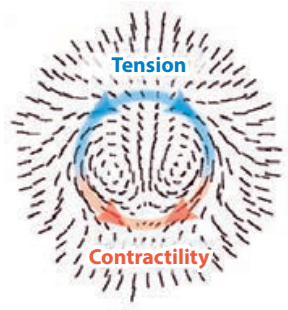
Marr's Tri-Level in developmental *Patterning*

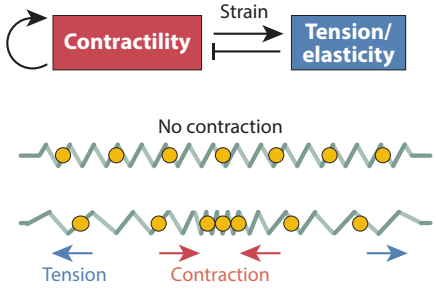
- Mechanical implementation: Turing-like instabilities

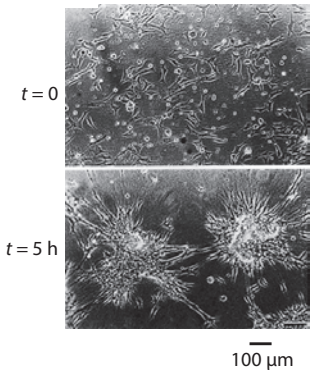


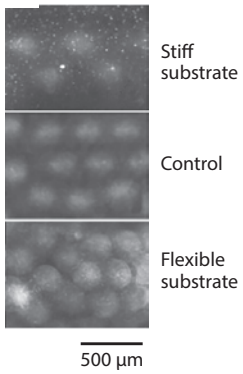


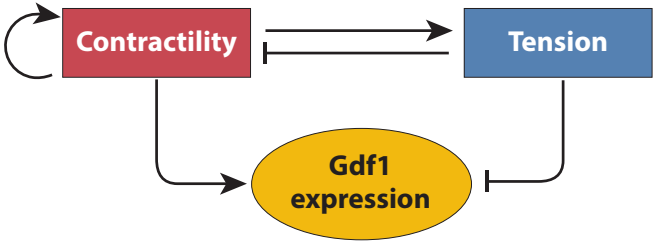












Bailles A, Gehrels EW, Lecuit T. *Annu Rev Cell Dev Biol*. 38:321-347 (2022)

Marr's Tri-Level in developmental *Patterning*

- Neuronal implementation: Turing-like instabilities

A Model for Shell Patterns Based on Neural Activity

by

BARD ERMENTROUT

Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, U.S.A.

JOHN CAMPBELL

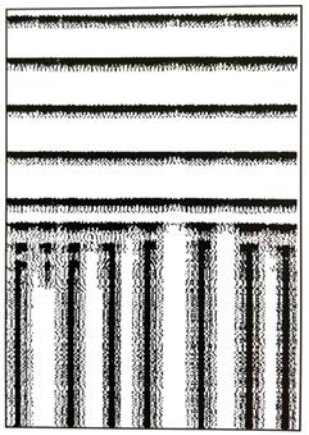
Department of Anatomy, University of California, Los Angeles, California 90024, U.S.A.

AND

GEORGE OSTER

Departments of Biophysics and Entomology, University of California, Berkeley, California 94720, U.S.A.

The Veliger 28(4):369-388 (April 1, 1986)



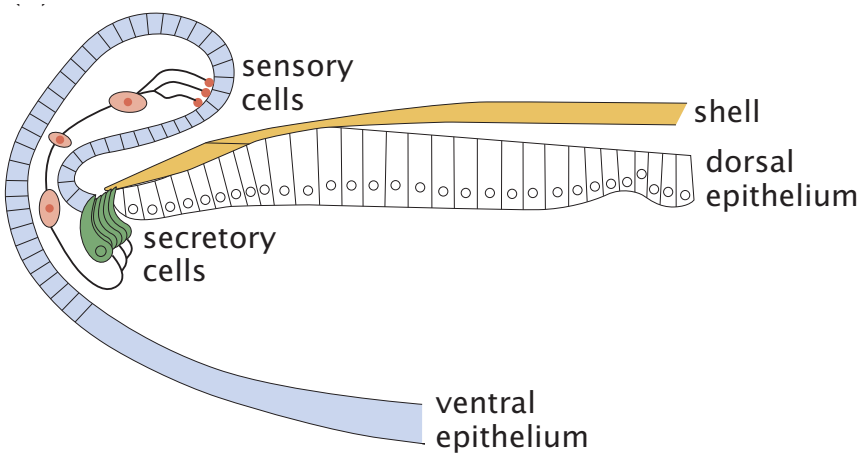
Synchronous oscillations:
Delayed negative feedback
+ spatial coupling

Spatial patterns: local excitation
+ long range inhibition

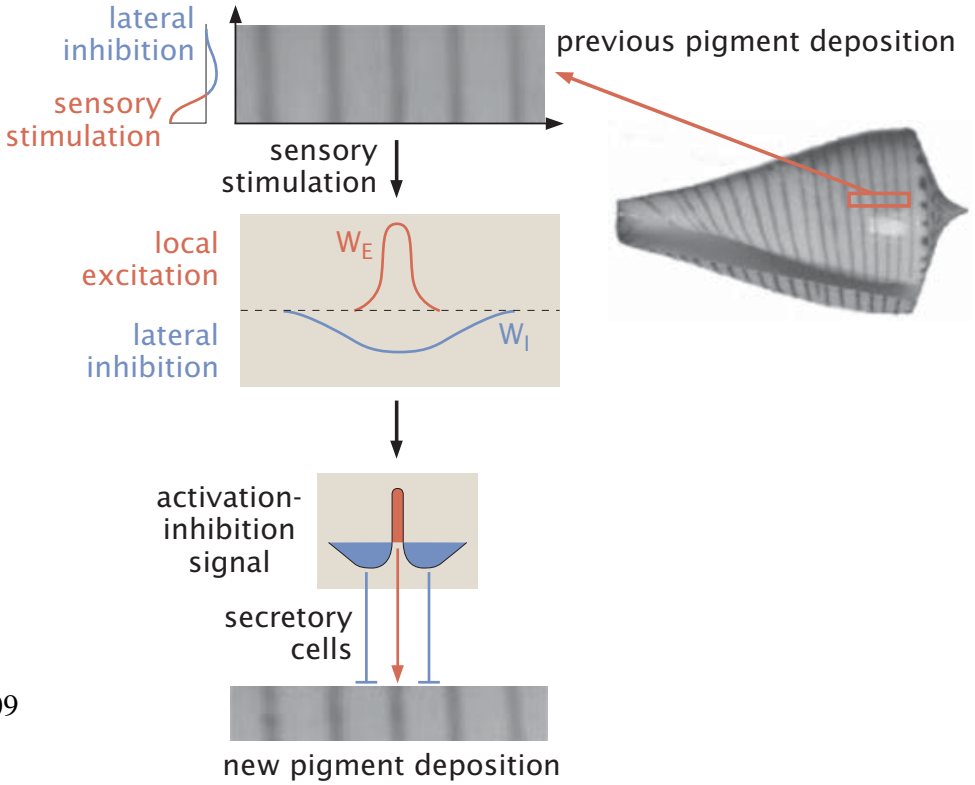
Hans Meinhardt
The algorithmic beauty of seashells (Springer)

Marr's Tri-Level in developmental *Patterning*

- Neuronal implementation: Turing-like instabilities



A. Boettiger, B. Ermentrout and G. Oster. *P.N.A.S* 106:6837-6842. 2009



R. Phillips, J. Kondev, J. Thériot & H. Garcia.
Physical Biology of the Cell (Garland Science) 2012

Conclusion

A framework to disentangle:

- *Purpose*(why): *computation*
- *Strategy* (how): *algorithm*
- *Biology/physics* (what): *implementation*

- Meaning: necessity
- Logic: necessity
- contingency (history)

Unity of algorithmic level for different implementations
Diversity of algorithmic level representations and solutions for a given function

Information provides a language to decipher the *meaning* and *logic of living systems*

VISION

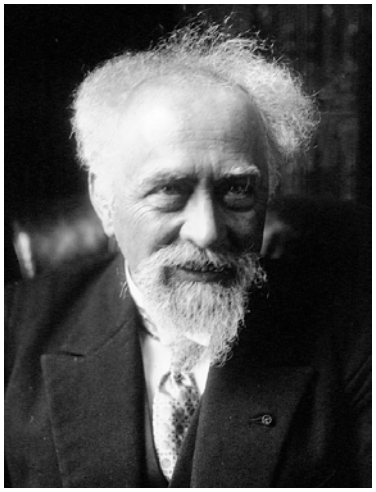
A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr (1945-1980)

1982, Vision, David Marr
W. H. Freeman and Company
2010: *MIT press* (re-published)

In search of principles in biology

Henri Poincaré (1854-1912)



Jean Perrin (1870-1946)

on being a Bio-logician

« Le savant doit *ordonner*;

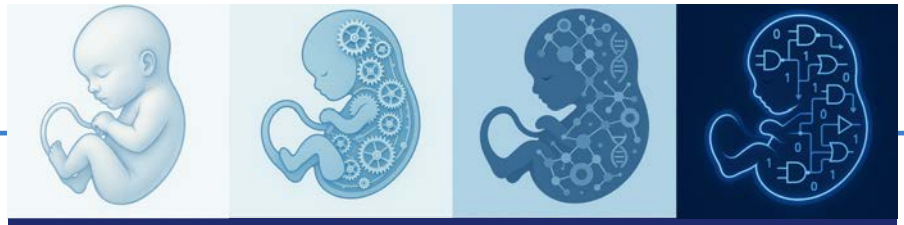
on fait la Science avec des faits comme une maison avec des pierres ;

mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison. »

La science et l'hypothèse (1902)

« La Science remplace du visible compliqué par de l'*invisible simple*. »

Les Atomes (1913)



Thomas LECUIT, chaire Dynamiques du vivant

Qu'est-ce que l'information biologique ? (II)

COURS : 20 novembre > 18 décembre 2025

COURS

Le jeudi de 10 h à 12 h
Amphithéâtre Guillaume Budé

Jeudi 20 Novembre 2025
Introduction :
approche computationelle du vivant

Jeudi 27 Novembre 2025
Complexité et information
au cours du développement

Jeudi 4 Décembre 2025
Vision logique des flux d'information

Jeudi 11 Décembre 2025
Vision dynamique des flux d'information

Jeudi 18 Décembre 2025
Apprentissage non neuronal
dans un système biologique

COLLOQUE

De 9h à 18h
Amphithéâtre Maurice Halbwachs

Vendredi 26 juin 2026
Information flow and computation in living systems

Les cours et colloques
sont gratuits, en accès libre,
sans inscription préalable.

Image générée par une I.A. © T. Lecuit.

COLLÈGE DE FRANCE 1530

Thomas Römer, administrateur du Collège de France
11, place Marcelin-Berthelot, 75005 Paris
www.college-de-france.fr

Année académique 2025/2026