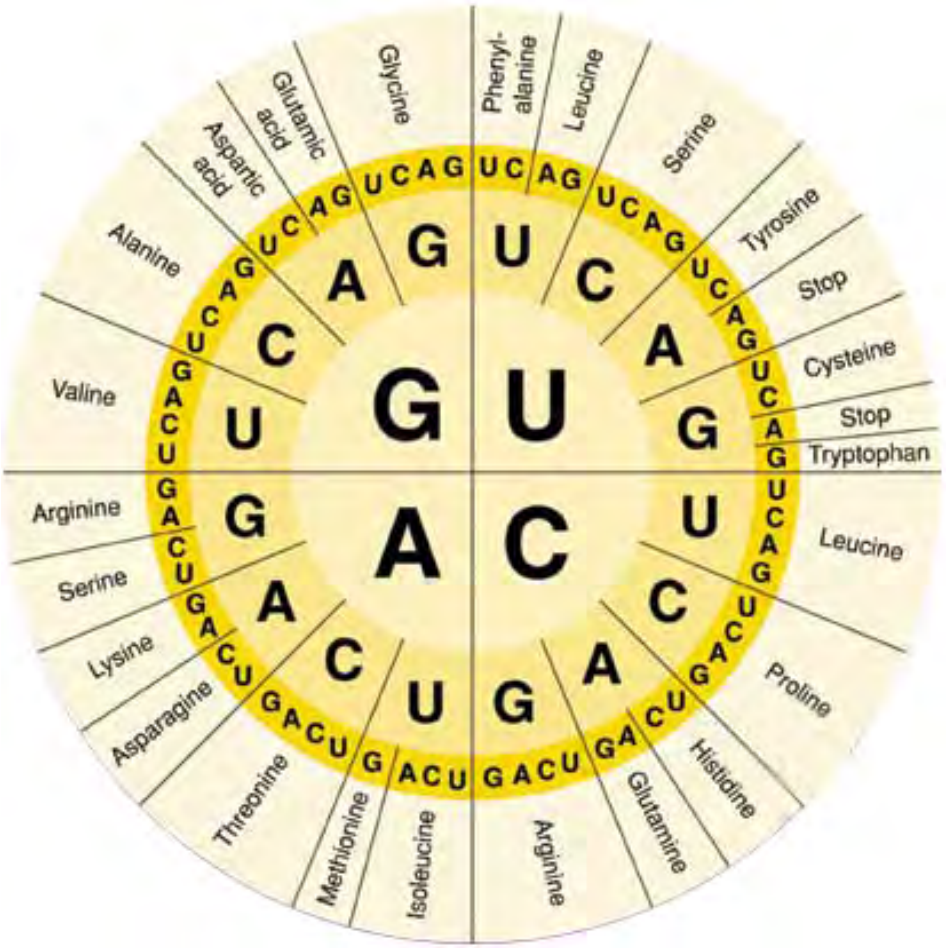
On Formal Representations of Organic Codes



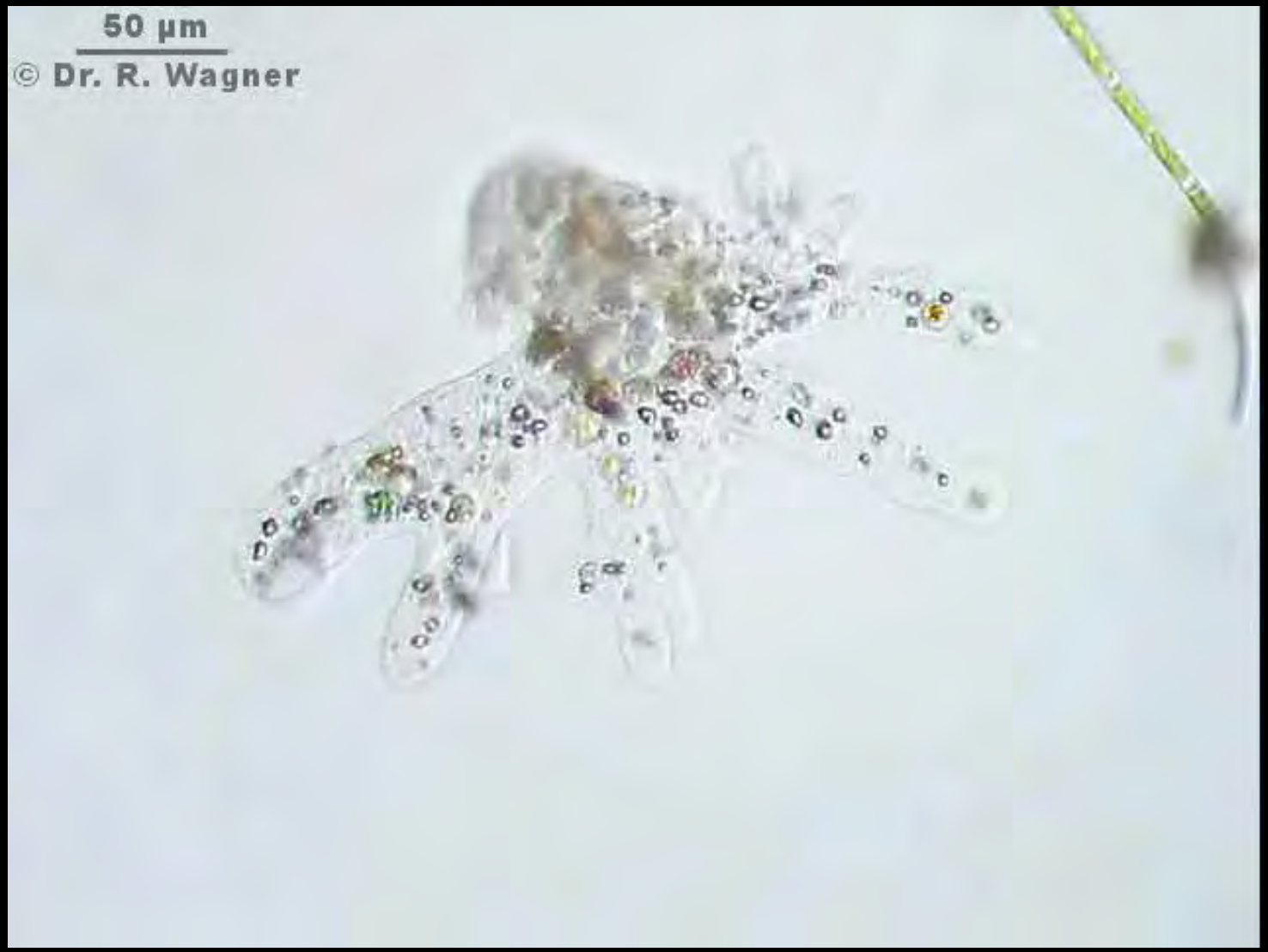
Jannie Hofmeyr

Centre for Studies in Complexity and Dept. of Biochemistry, Stellenbosch University

ISCB, Paris 2014

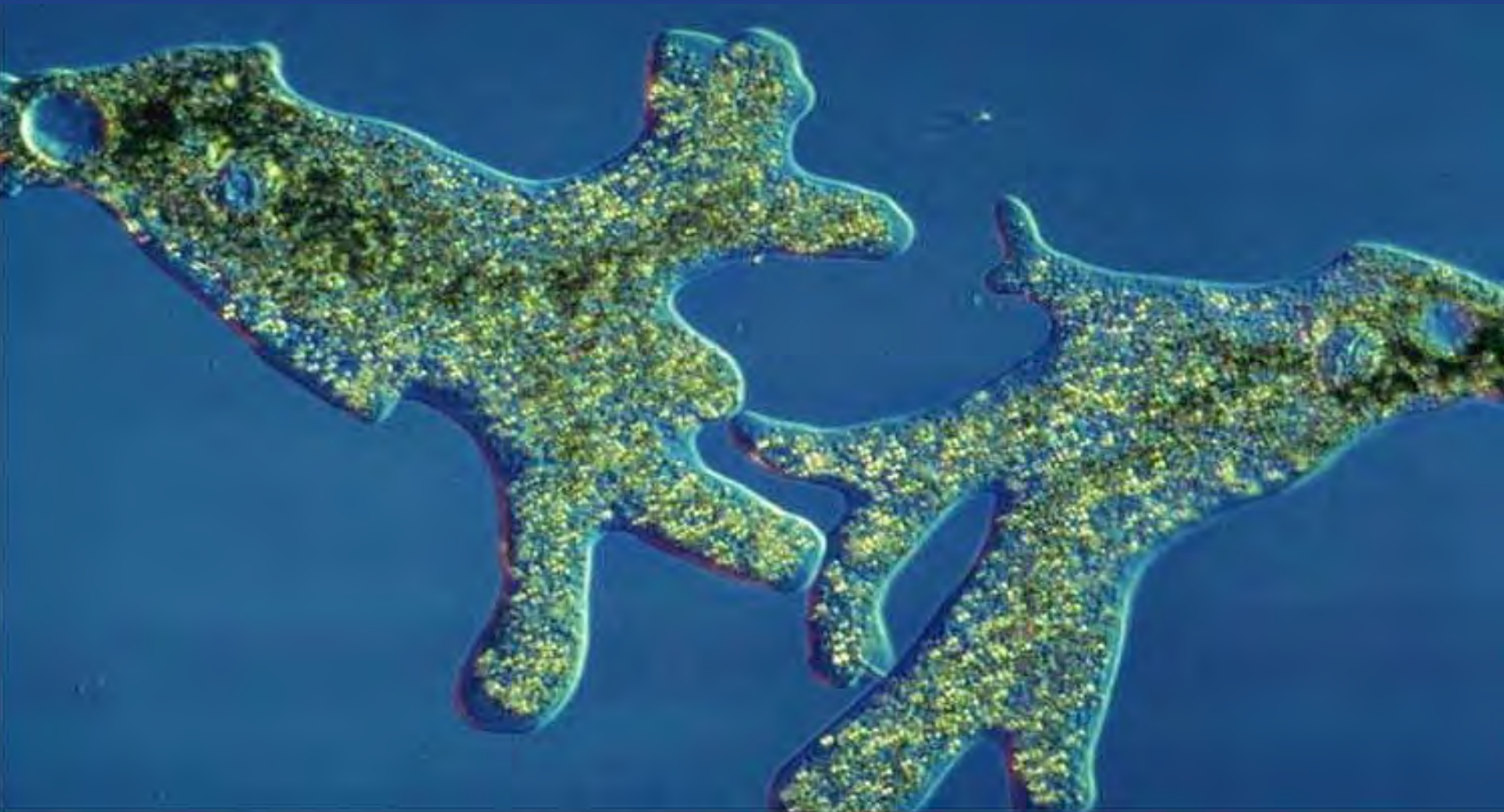
A bit of personal history…

What is the most fundamental problem any organism faces?




How to persist as an entity despite the fragility of all its components
How to fabricate itself

A linguistic model of self-fabrication



Question

What is the set of internal relations (the entailment structure) that makes self-fabrication possible?

or

Is there a relational logic underlying self-fabrication?

Strategy of cells as we know them

Use a single, conceptually straightforward chemical process — polymerisation — to create large, linear molecules

that

fold themselves into functional, three-dimensional structures

that

can autonomously self-assemble into higher-order structures

Central Question

Does the choice of
sequence construction by concatenation
(polymerisation) as the mode of fabrication

have logical consequences for self-fabrication?
i.e.,
does it logically entail the other features?

or are there other equivalent ways to ensure closure?

My aim

To create a formal model of self-fabrication based on sequence construction

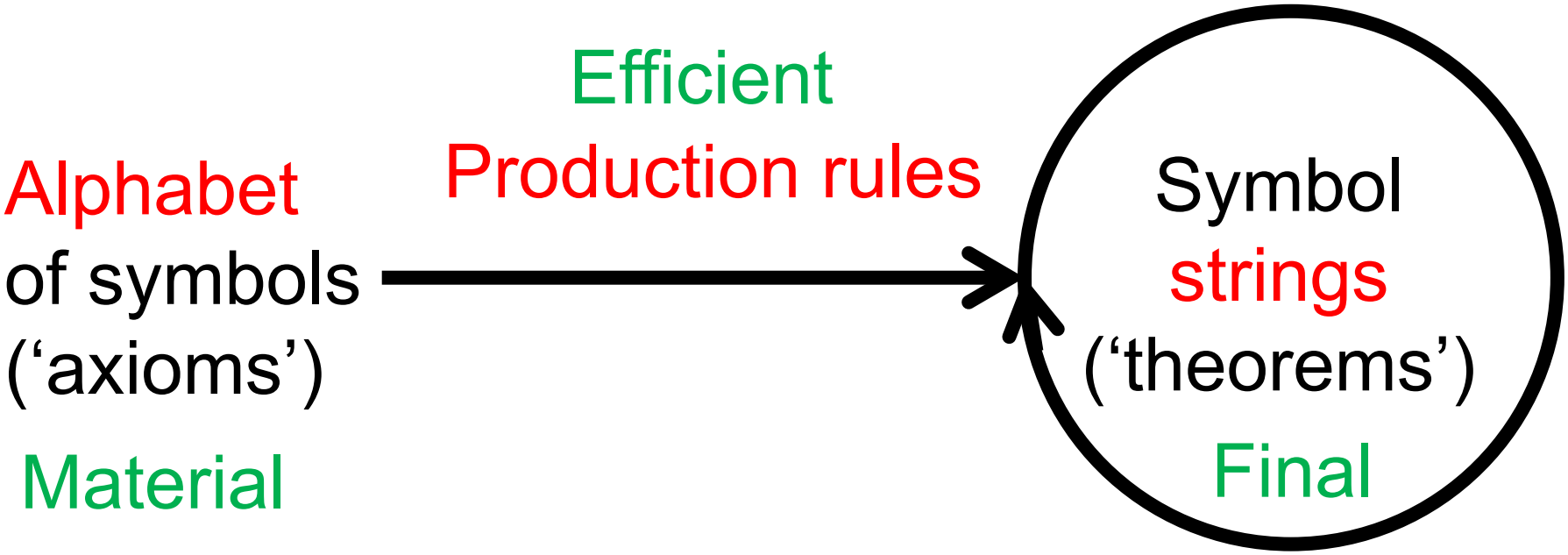
Ask: to which degree does it satisfy Robert Rosen's modelling criterion of *congruence* between inferential entailment in the formal system and causal entailment in the living cell that is modelled?

A self-producing formal system (or language)

that

I can explain to my non-scientific, non-mathematical, but literate grandmother

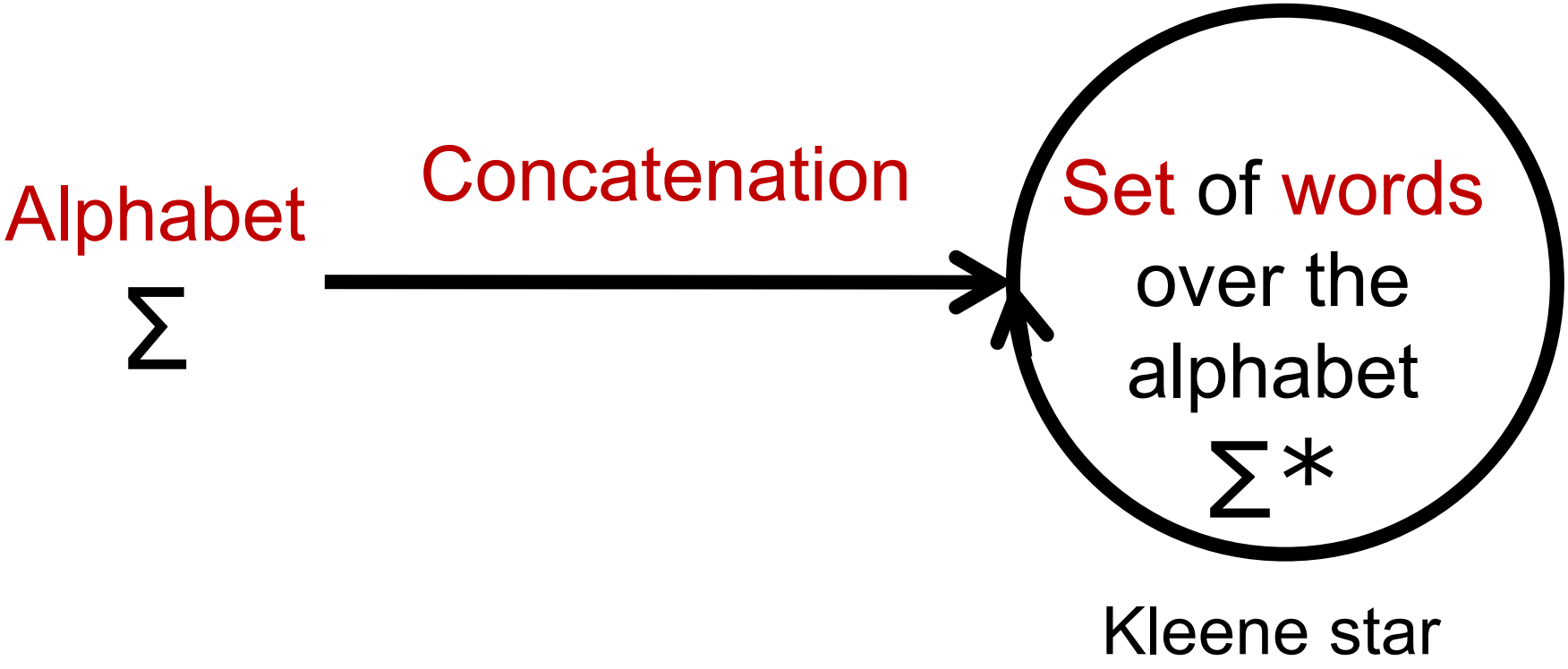
Formal system



Algorithm = Rule application sequence

Formal

Formal language (subset of Σ^*)



Mathematical object:

Free monoid generated from an alphabet

Constraint 1

Everything in the system

--- also the production rules ---

must be made from the same ‘stuff’

{a,b,c,d,e,f,g,h,I,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z}

Constraint 2

No general metarule such as ‘concatenation’ that sits outside the system is allowed

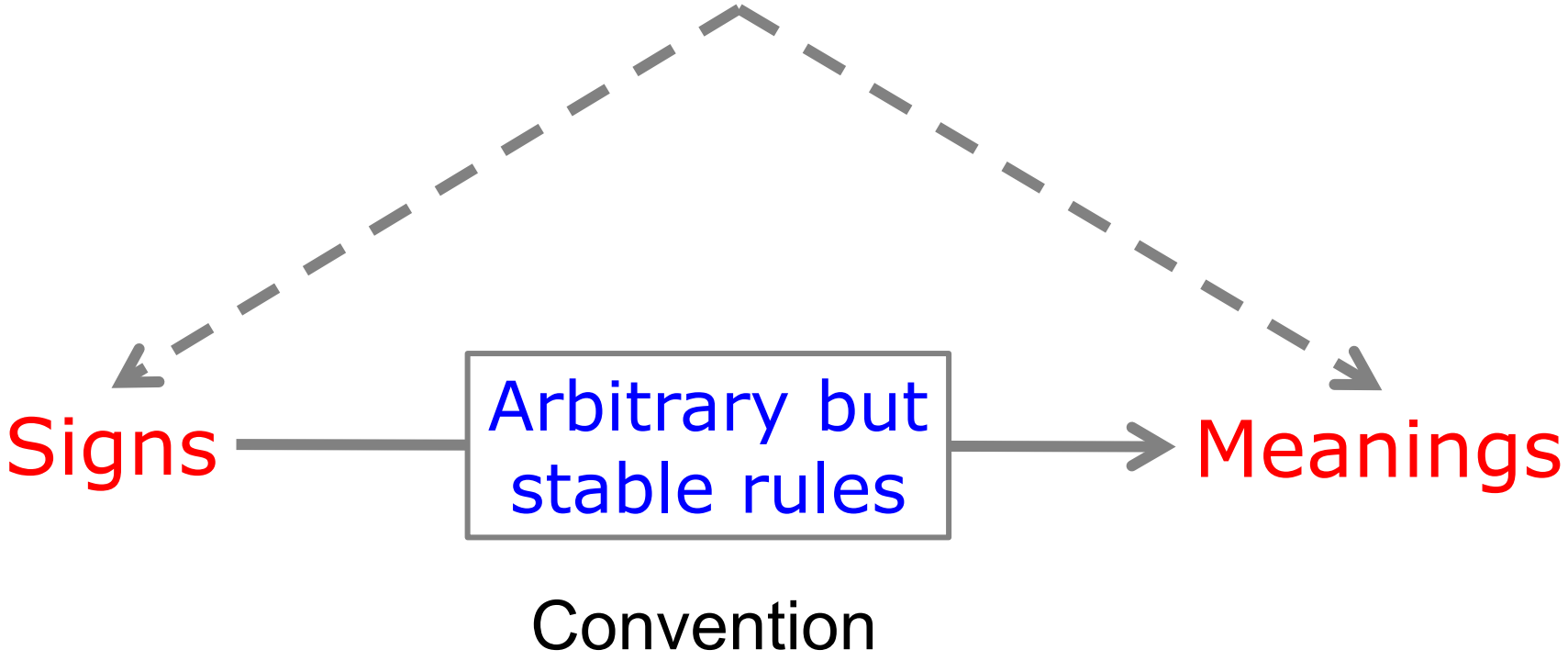
Internal, specific rules
(sentences such as ‘join a and b’)
must be constructed within the system

Sequence construction logically necessitates:

Formal system	Living system
Alphabet (letters)	Nutrients
Production rules	Specific catalysts (enzymes)
Sentence activation	Polypeptide folding (environment)
Description of sequences	DNA (mRNA)
A code/translation system	Genetic code/ribotype
A tag-matching-and-splitting algorithm for self-assembly	Environment that pushes protein subunits together (hydrophobic effect) and specific binding interactions

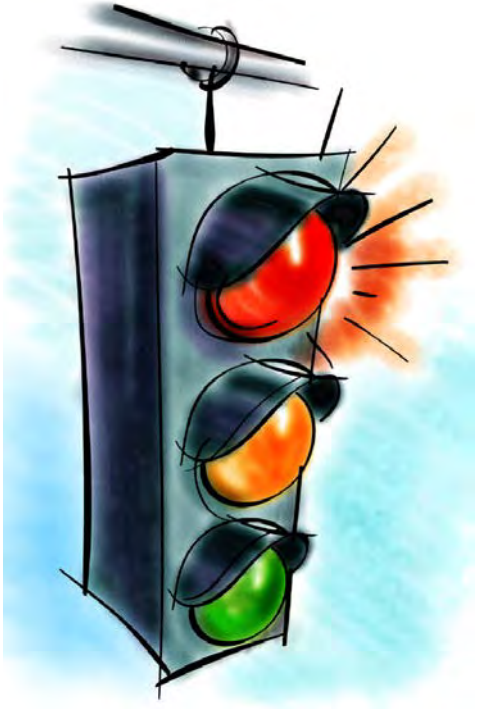
Codes in general

A code establishes a
conventional relationship
between
two independent worlds



World 1 (signs)

World 2 (meanings)



The highway code

establishes a relation between

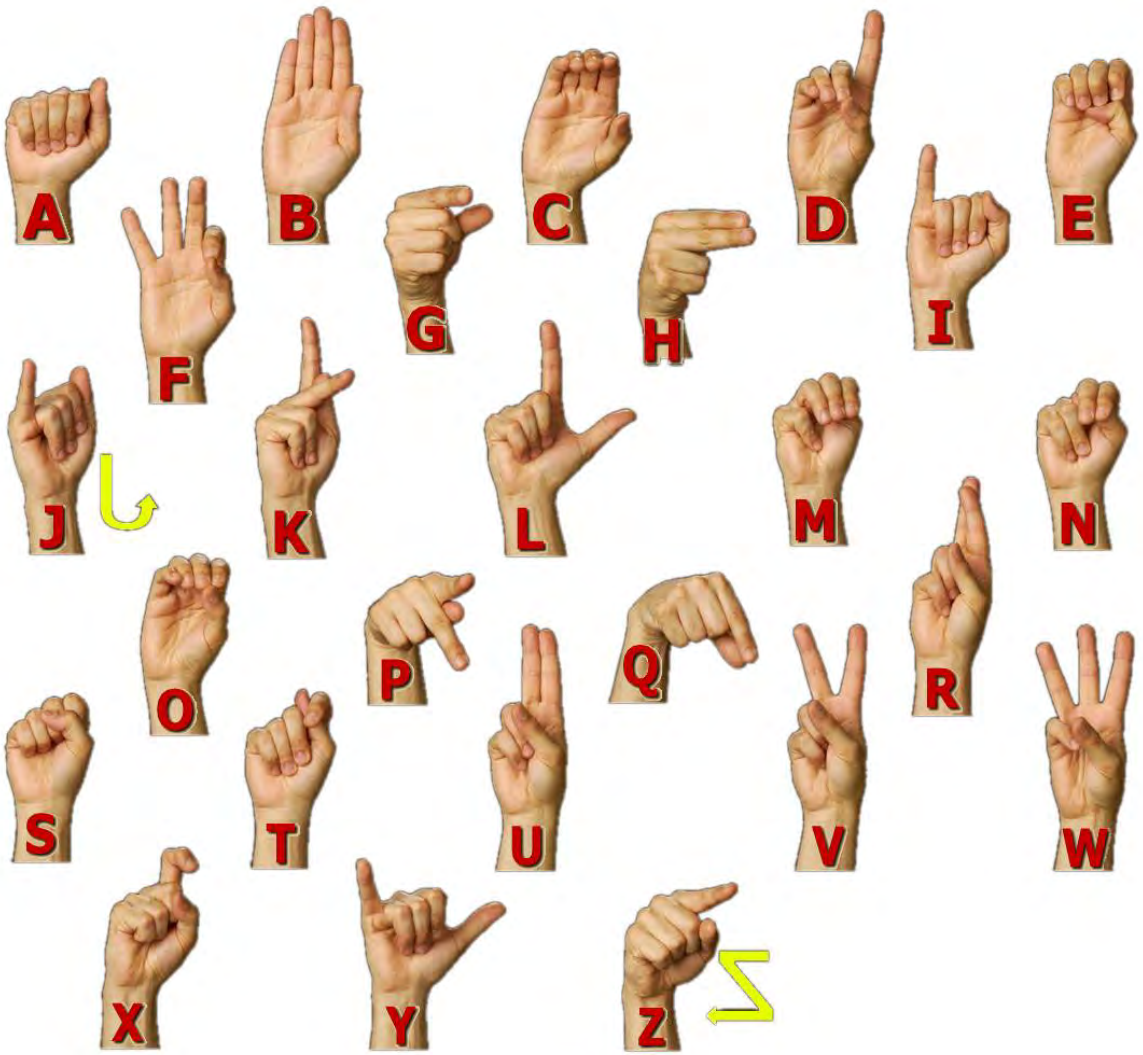
Traffic signs

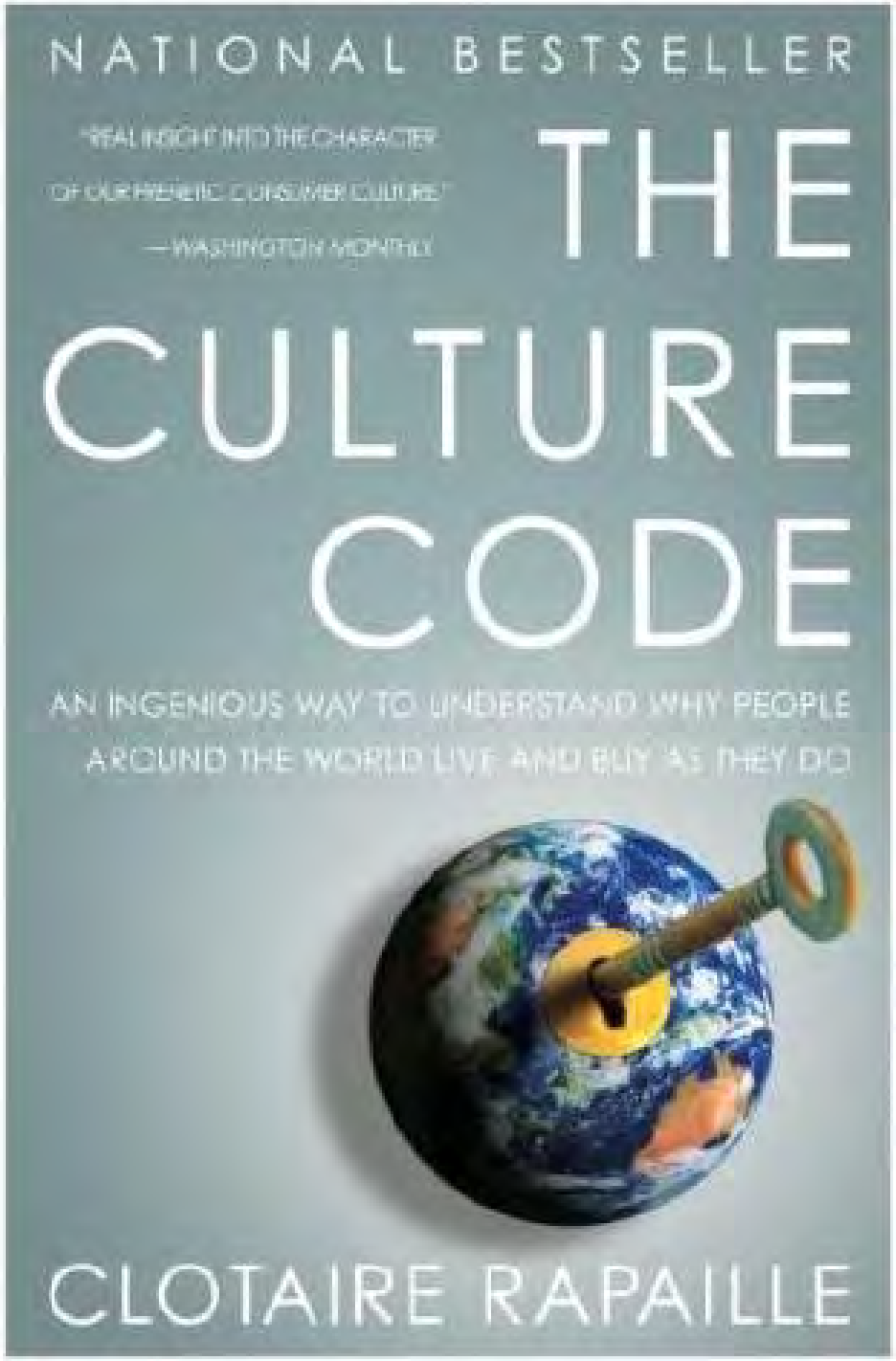
Driving behaviours

German legal code from the 15th century

The codes of languages

Sign language code

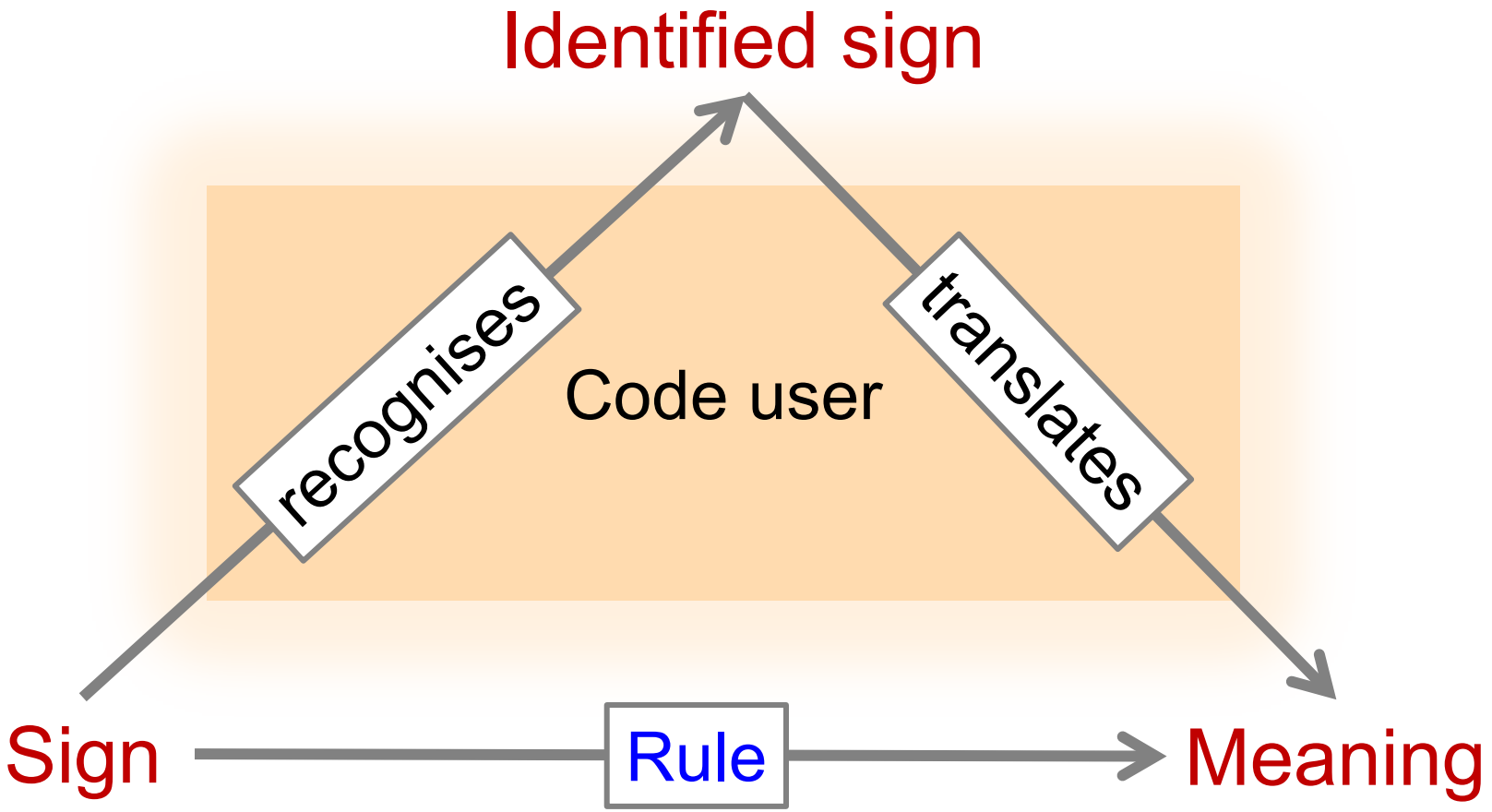




The Consumer Code

"An ingenious way to understand why people around the world live and buy as they do."

The Morse Code


Identified sign
recognises
Code user
translates
Sign
Rule
Meaning

Formal language definition of a code

Entities	Example: **Morse code**
S (source alphabet)	$\{a, b, c, \ldots, z, 0, 1, 2, \ldots, 9\}$
T (target alphabet)	$\{\bullet, -\}$
*T**, set of words over *T*	$\{\bullet-, -\bullet\bullet, -\bullet-\bullet, \ldots, ---\}$
Code $C : S \longrightarrow T^*$	$a \mapsto \bullet-$ $b \mapsto -\bullet\bullet$ $c \mapsto -\bullet-\bullet, \ldots$
*S**, set of words over *S*	$\{the, morse, code, aaa, aba, aca, \ldots\}$
Extension $E : S^* \longrightarrow T^*$	$the \mapsto -, \bullet\bullet\bullet\bullet, \bullet$ $morse \mapsto --, ---, \bullet-, \bullet\bullet\bullet, \bullet$ $code \mapsto \bullet-\bullet-, ---, -\bullet\bullet, \bullet, \ldots$

The Organic Codes

Marcello Barbieri

The Genetic Code

3.8 billion years

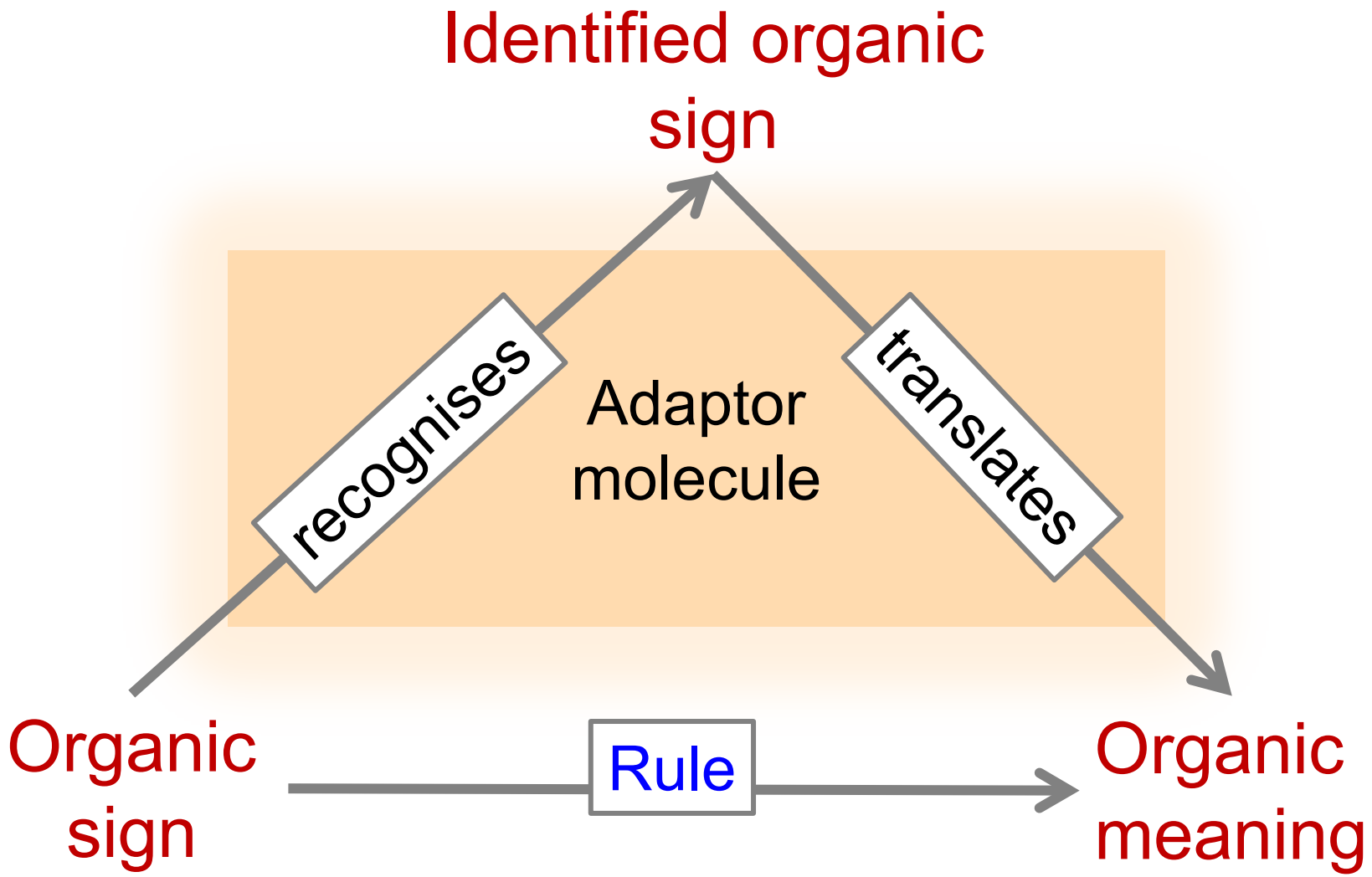
Cultural Codes

Nothing?
No more codes?

Yes, lots of them!

Code Biology

Organic codes

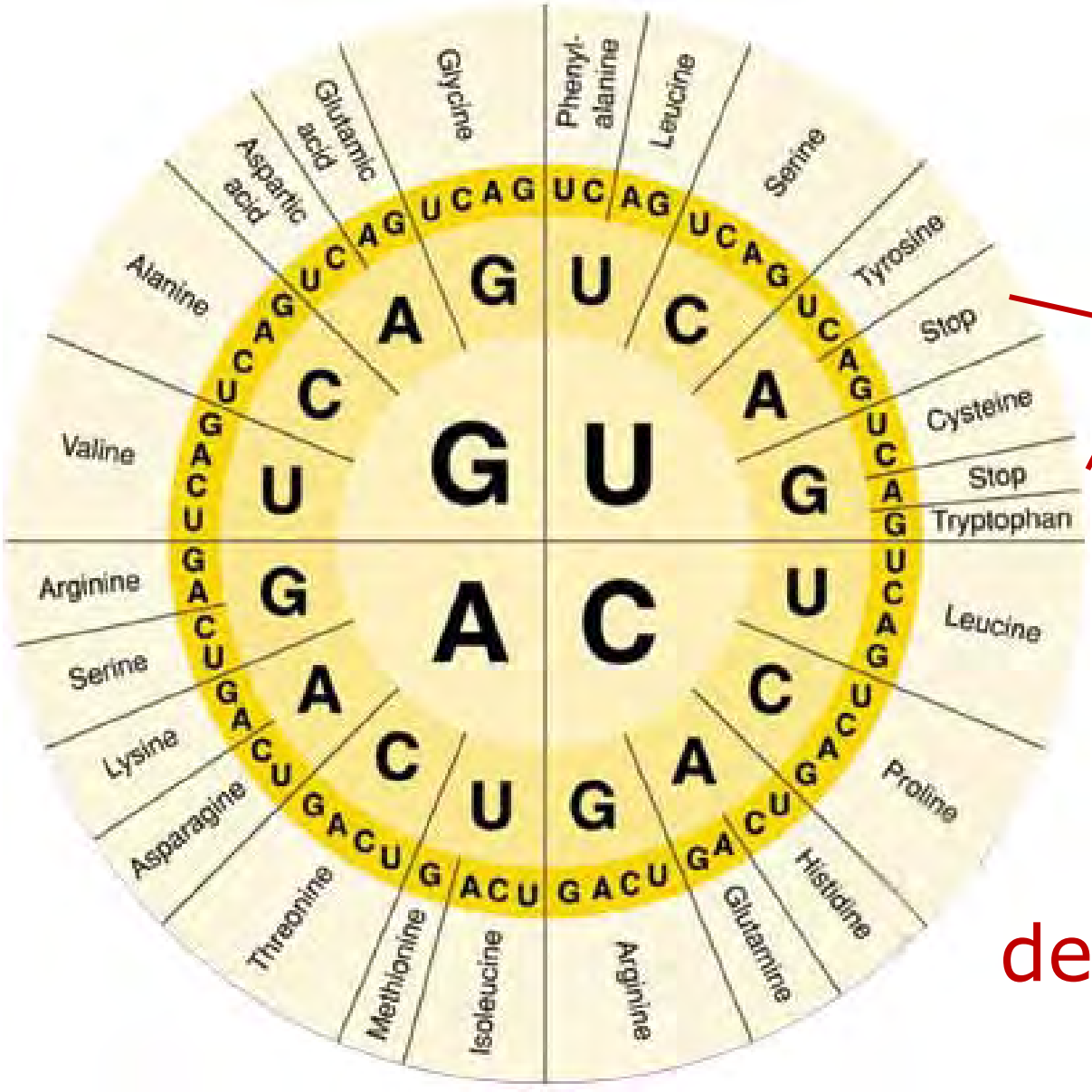


Genetic code

Formal definition of the genetic code

Entities	Example: **Genetic code**
S (ribonucleotides)	$\{\mathrm{A}, \mathrm{G}, \mathrm{C}, \mathrm{T}\}$
*S**, words over *S* (triplets, etc)	$\{\mathrm{A},\mathrm{G},\mathrm{C},\mathrm{T},\ldots,\mathrm{AAA}, \mathrm{AAG}, \mathrm{AGG}, \mathrm{AAC},\ldots\}$
T (amino acids)	$\{\mathrm{Gly}, \mathrm{Ala}, \mathrm{Ser}, \mathrm{Val},..., \mathrm{Try}, \mathrm{stop}\}$
Code $C_p : S^* \longrightarrow T$	$\mathrm{GGU} \mapsto \mathrm{Gly}$ $\mathrm{GGC} \mapsto \mathrm{Gly}$ $\mathrm{GCU} \mapsto \mathrm{Ala}, \ldots$
*S***, words over *S** (triplet seqs)	$\mathrm{GUU\ UUA}, \mathrm{UCU\ UAU\ GCU},...$
*T**, words over *T* (peptide seqs)	$\{\mathrm{Val\text{-}Leu}, \mathrm{Ser\text{-}Tyr\text{-}Ala}, \ldots\}$
Extension $E : S^{**} \longrightarrow T^*$	$\mathrm{GUU\ UUA} \mapsto \mathrm{Val\text{-}Leu}$ $\mathrm{UCU\ UAU\ GCU} \mapsto \mathrm{Ser\text{-}Tyr\text{-}Ala}, \ldots$

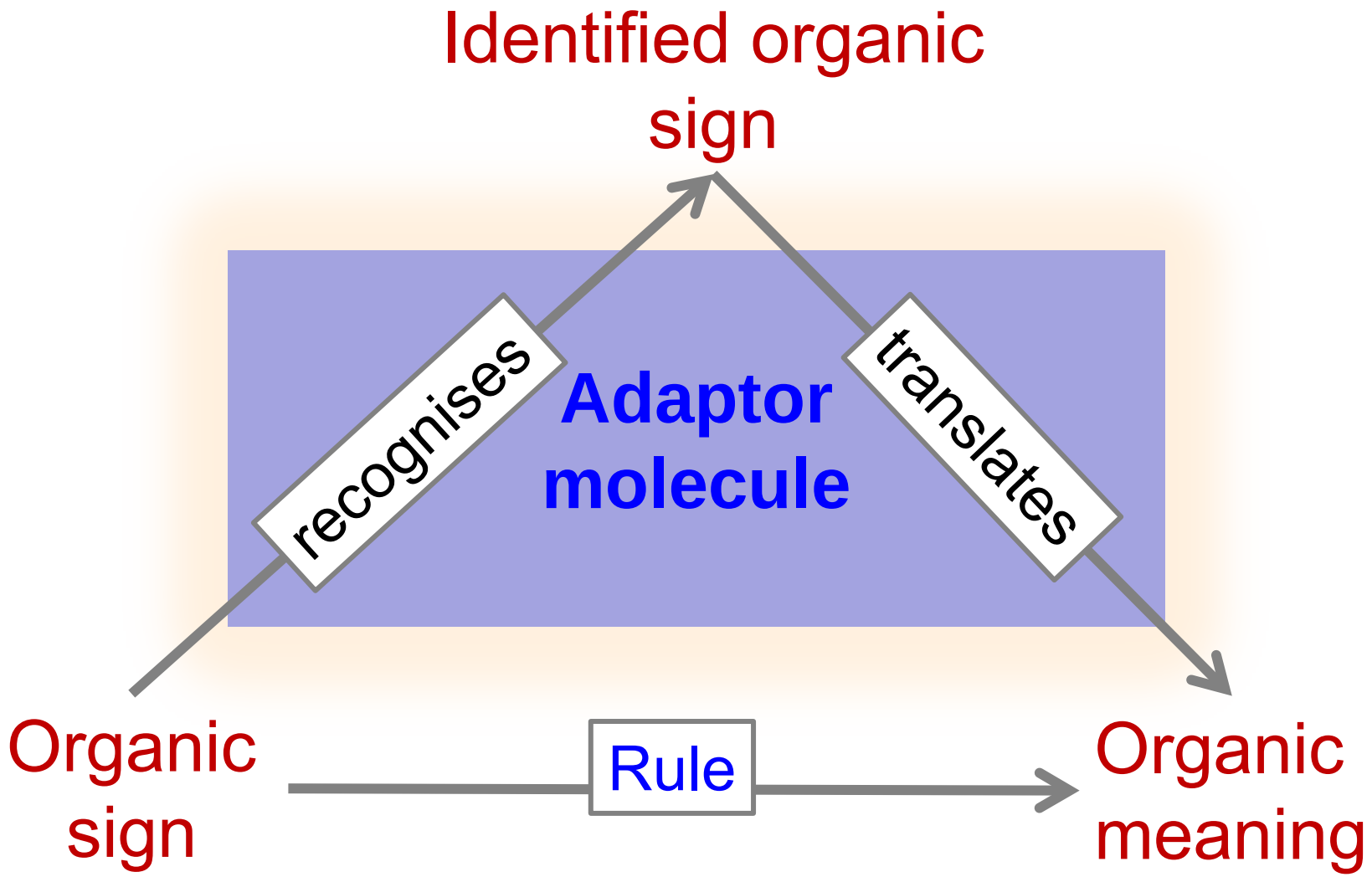
The genetic code

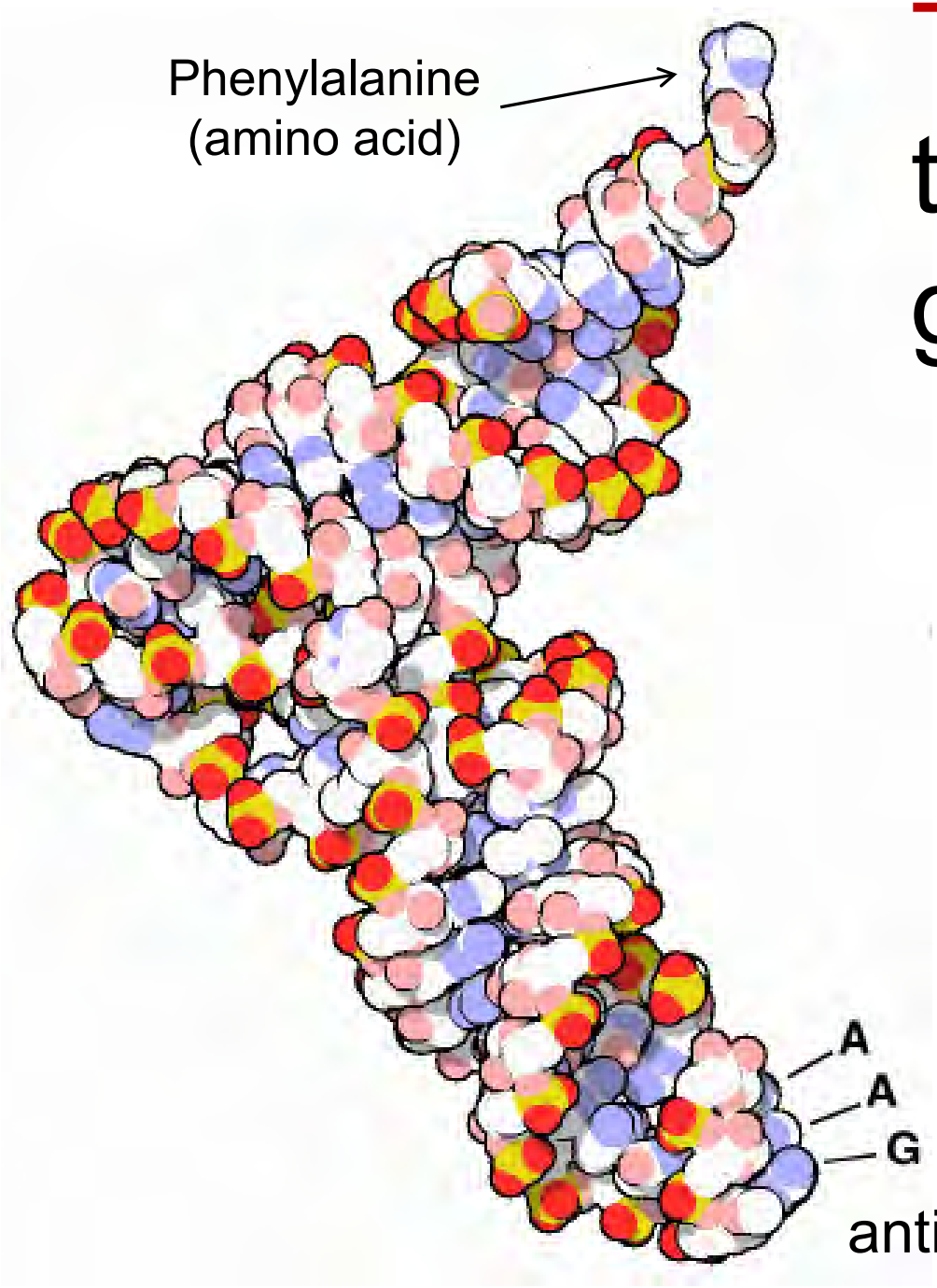


Punctuation marks

Code is degenerative

Organic codes

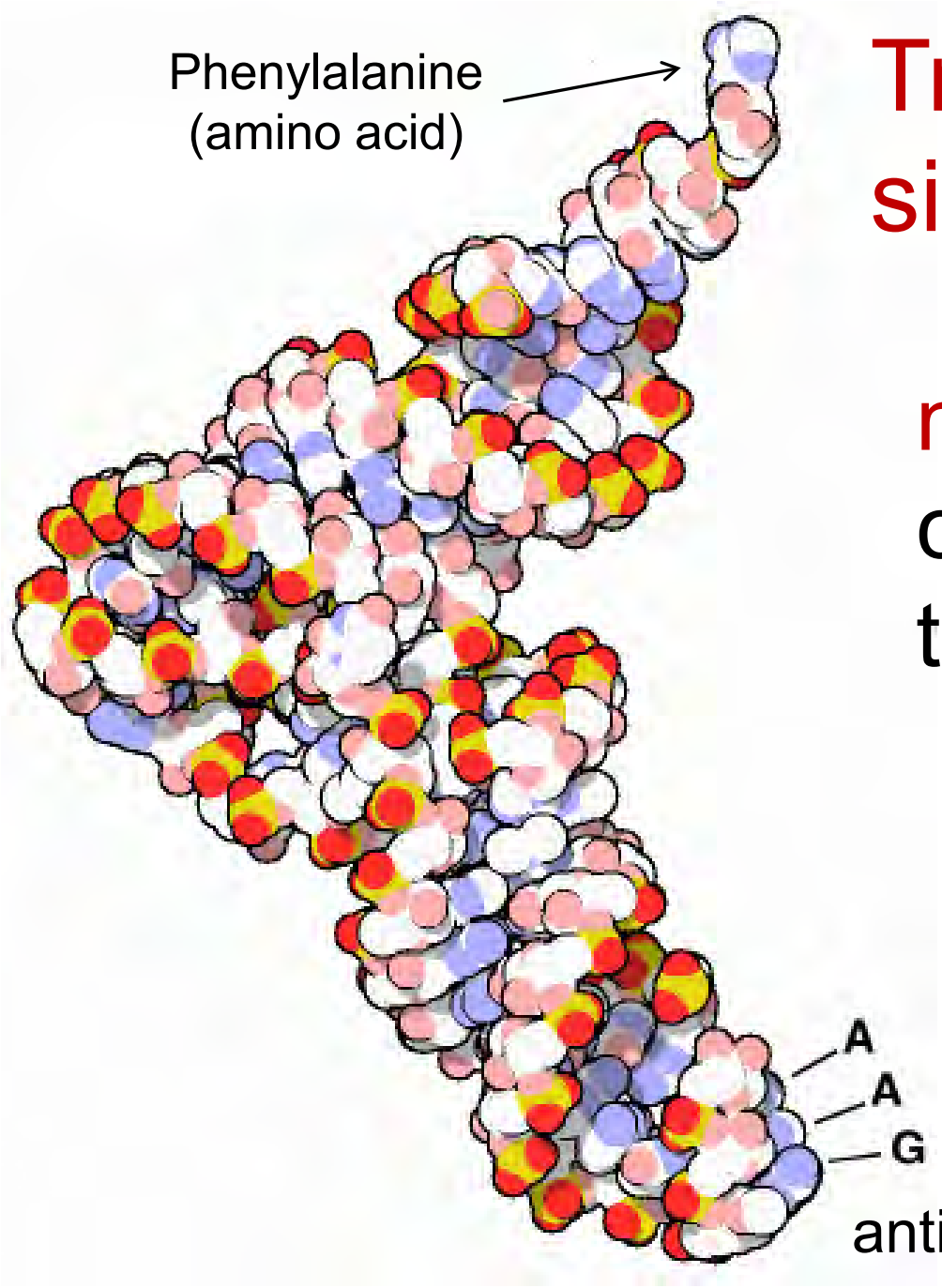




Transfer-RNA

the adaptor of the genetic code

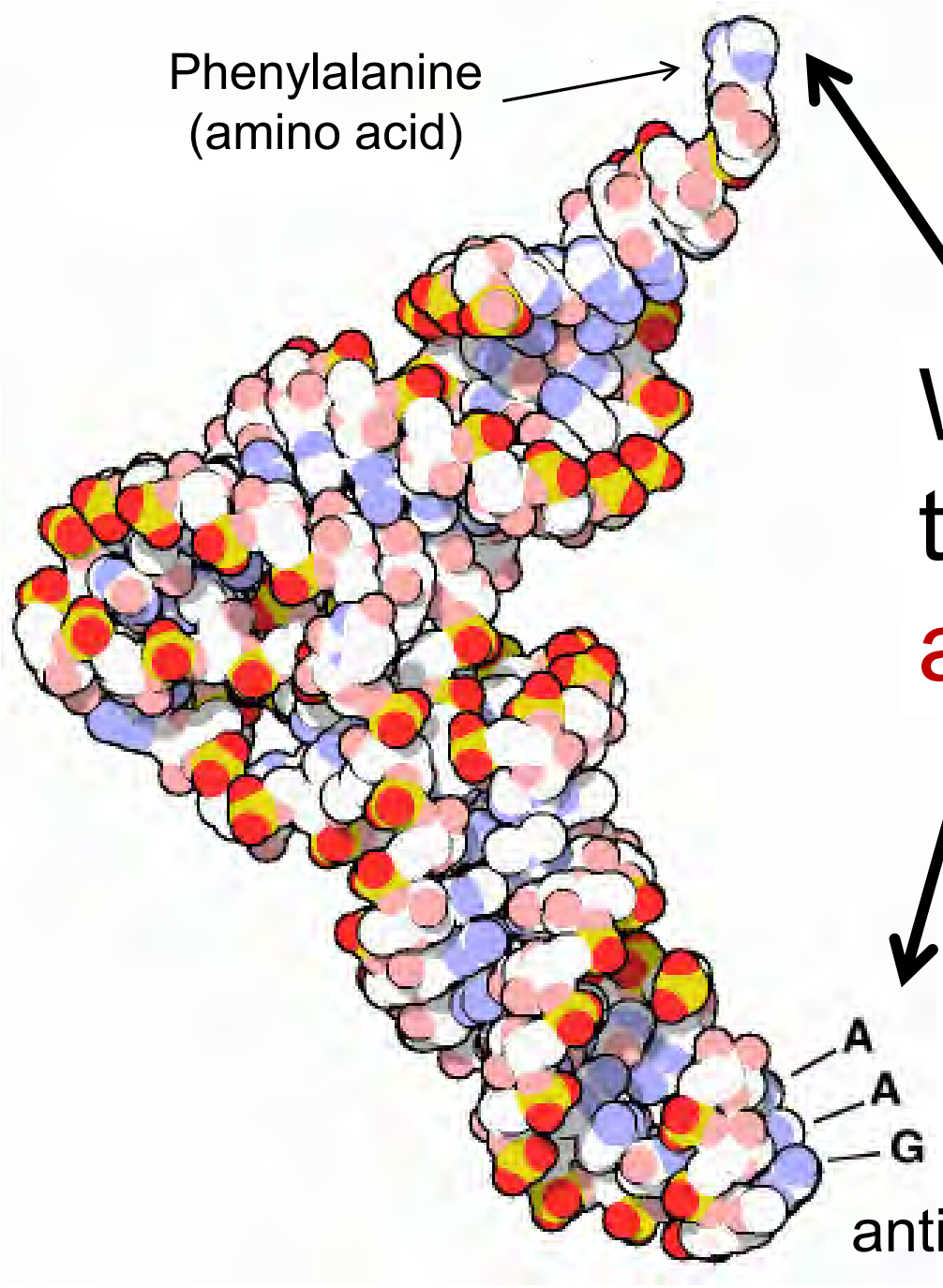
Adaptors are the "fingerprints" of organic codes



Translation site

no necessary connection between the two sites

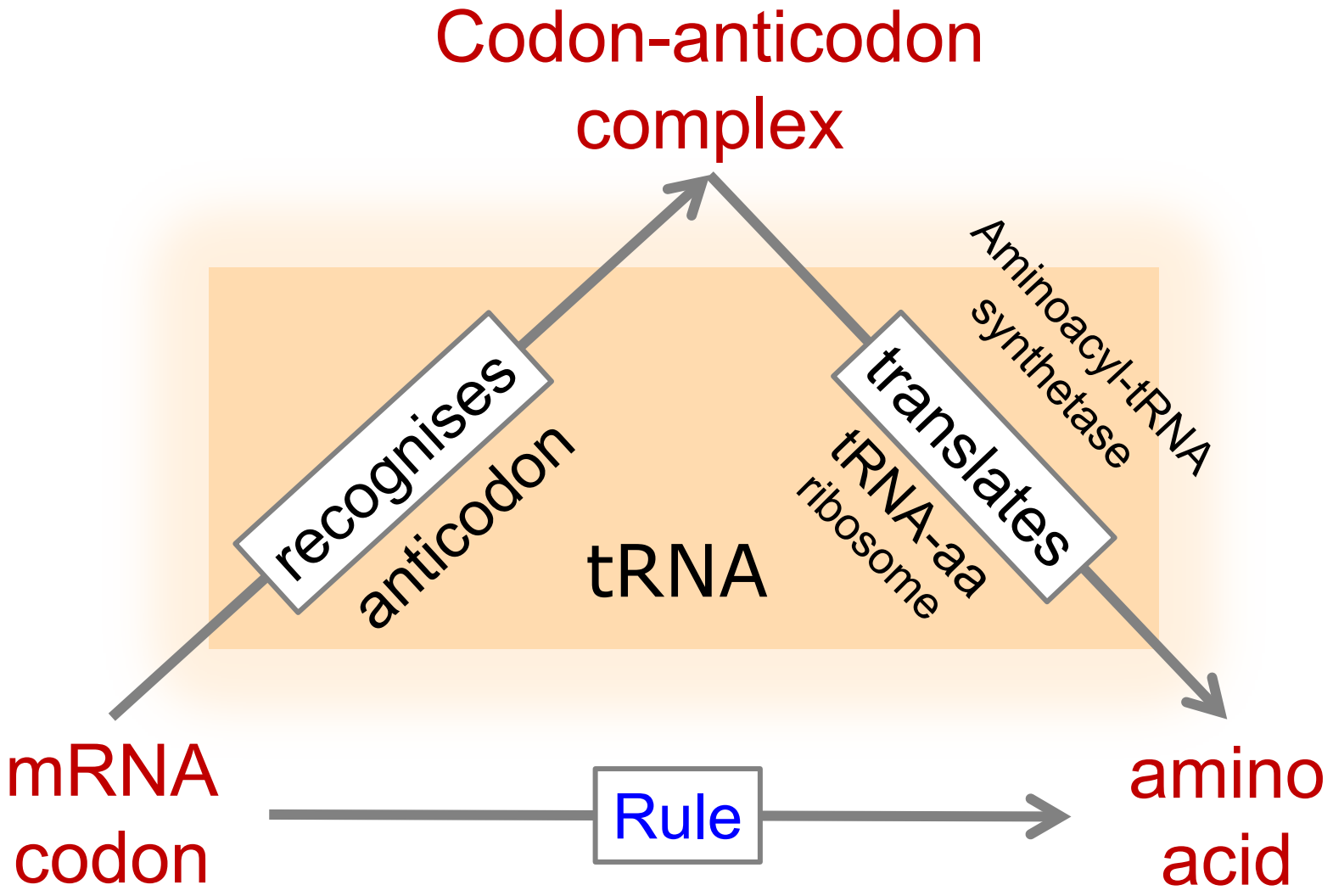
Recognition site



What is connected to what is an arbitrary choice,

that forms a natural convention

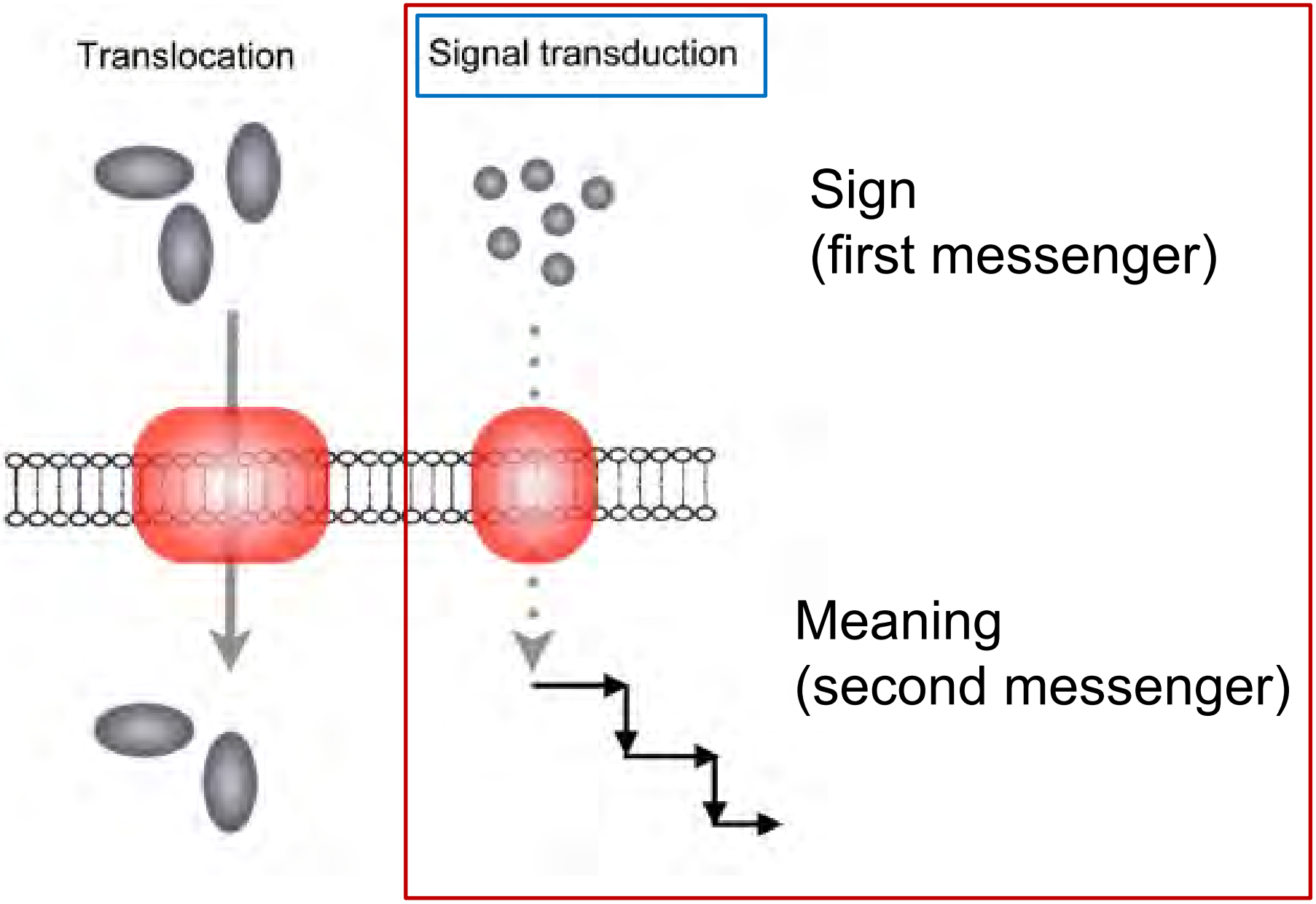
Genetic code



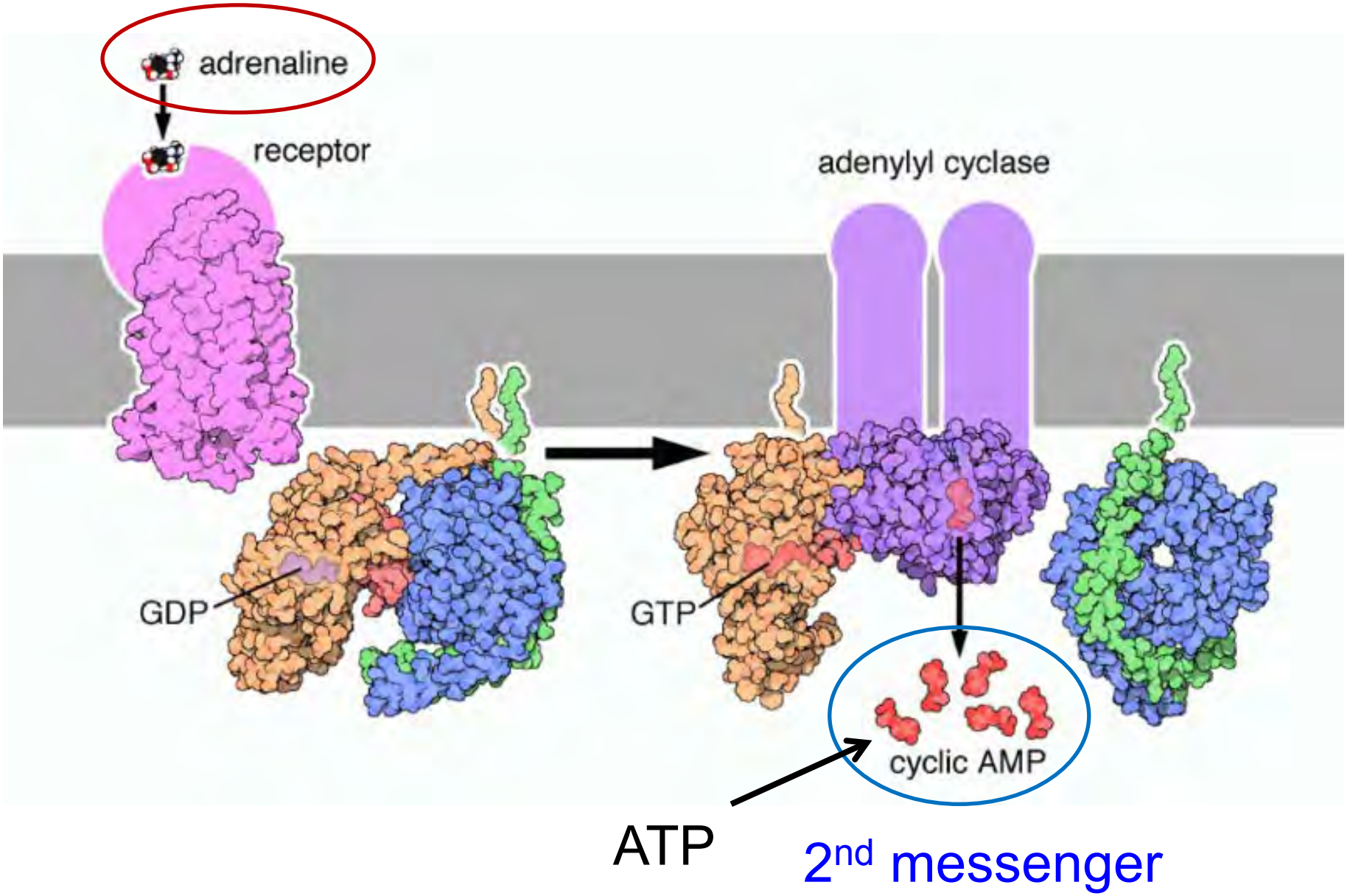
The Organic Codes

Code	Adaptors	World 1	World 2
Genetic	tRNA	DNA	polypeptide
Splicing	snRNA	pre-mRNA	mRNA
Signal transduction	receptor for 1st messenger amplifier for 2nd messenger mediator in between	1st messengers	2nd messengers
Compartment	signal peptides	cytosol	destination compartment
Cytoskeleton	anchoring molecules (accessory proteins)	microtubules	cell shapes
Adhesive (Redies & Takeichi)	cadherins	cell-cell binding	catenins (intracellular)
Histone (Strahl and Allis)	modified histones	transcription factors	gene expression
Sugar (Gabius)	lectins (enzymes, antibodies)	sugar epitopes on glycoconjugates	cellular responses through signalling

Signal transduction code

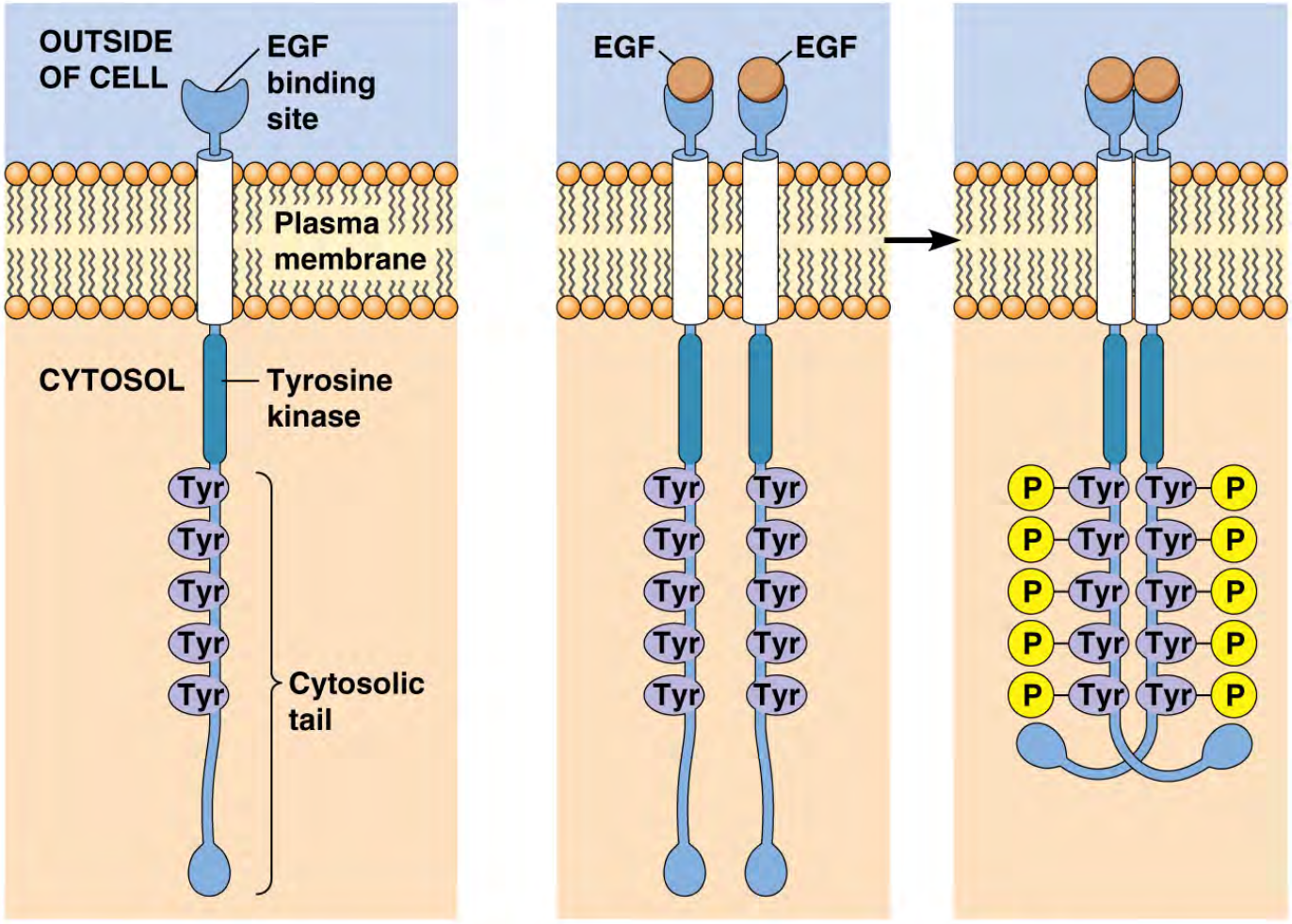

Translocation
Signal transduction
Sign
(first messenger)
Meaning
(second messenger)

1st messenger



2nd messenger

Epidermal growth factor signal transduction

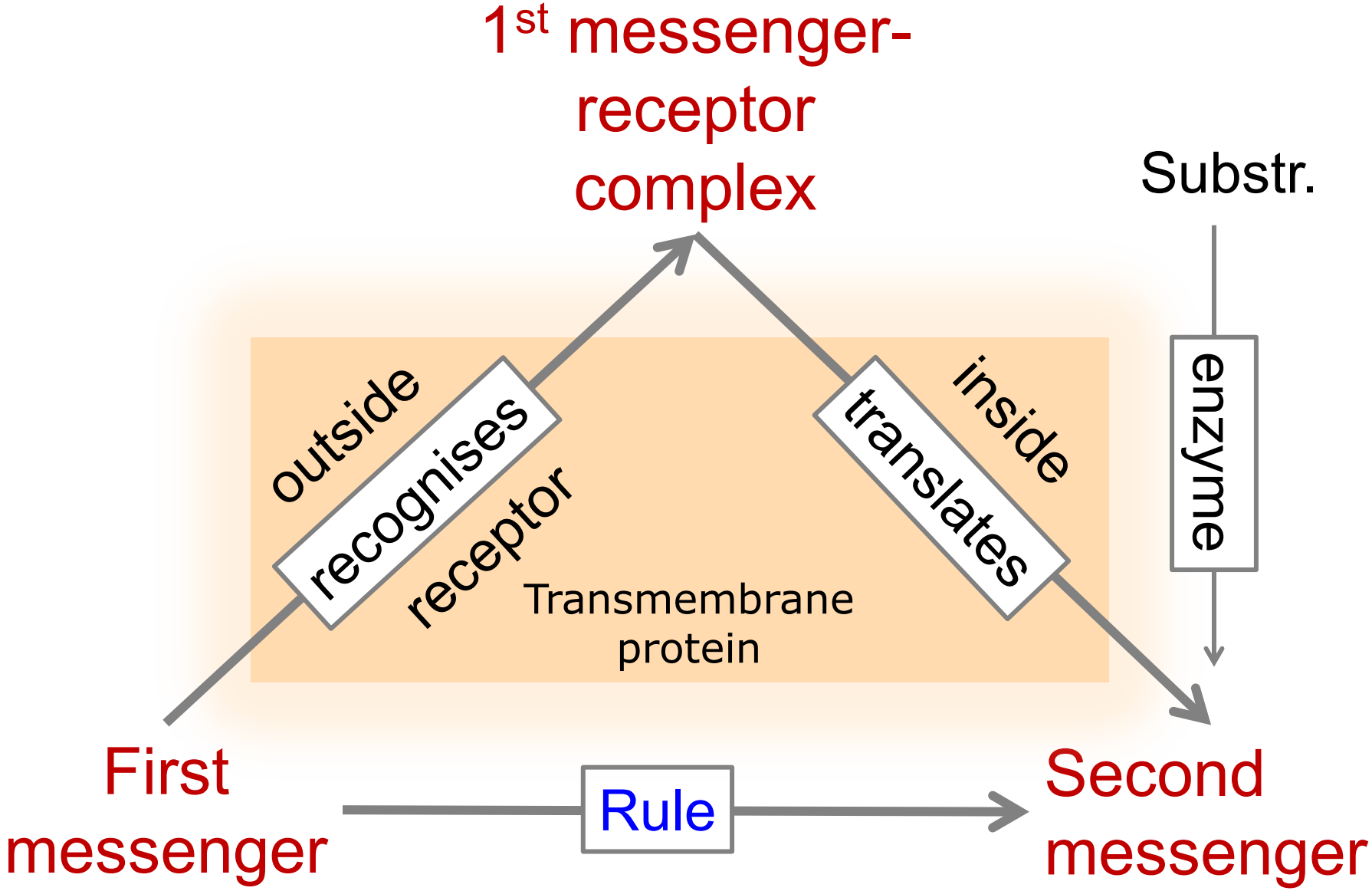


(a) Structure of the epidermal growth factor (EGF) receptor

(b) Activation of the EGF receptor

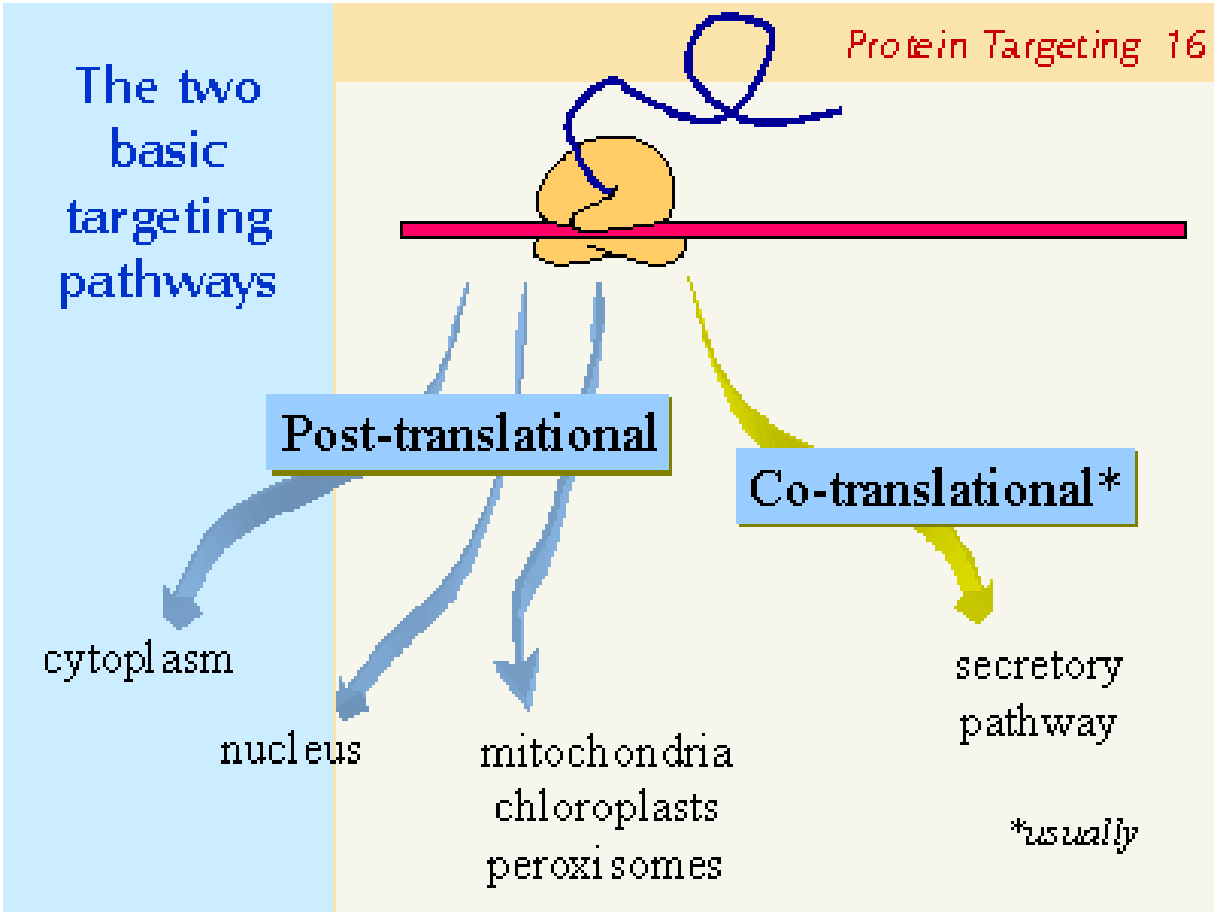
Entities	Example: **Signal transduction code**
S (first messengers, external)	{FM1, FM2, FM3,…}
T (second messengers, internal)	{cAMP, IP3, DAG, Ca^{2+}}
Code $C : S \longrightarrow T$	$\mathrm{FM1} \mapsto \mathrm{cAMP}$ $\mathrm{FM2} \mapsto \mathrm{cAMP}$ $\mathrm{FM2} \mapsto \mathrm{IP3}$ $\mathrm{FM3} \mapsto \mathrm{DAG}$ $\mathrm{FM4} \mapsto \mathrm{Ca}^{2+}, \ldots$

Signal transduction code

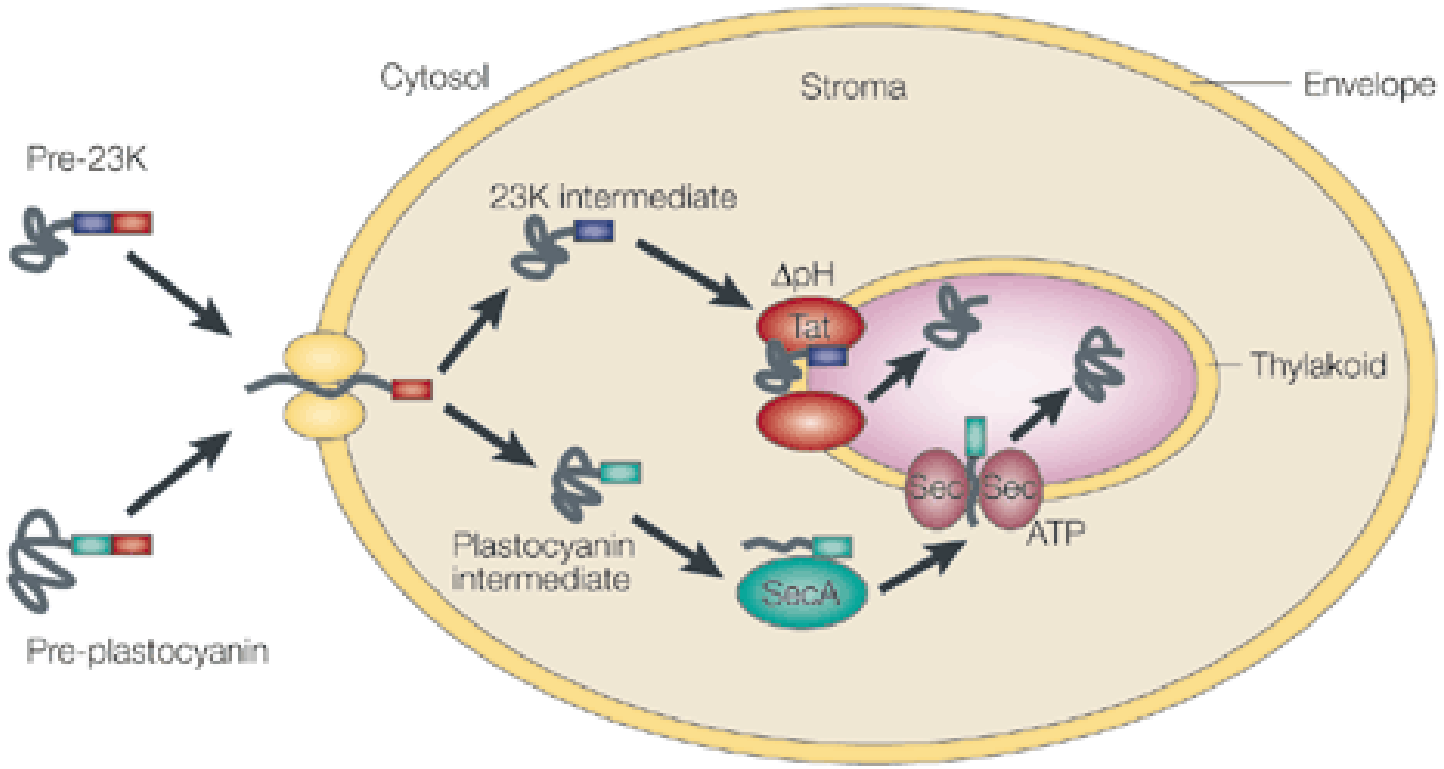


Protein targeting (compartment codes)

Protein targeting

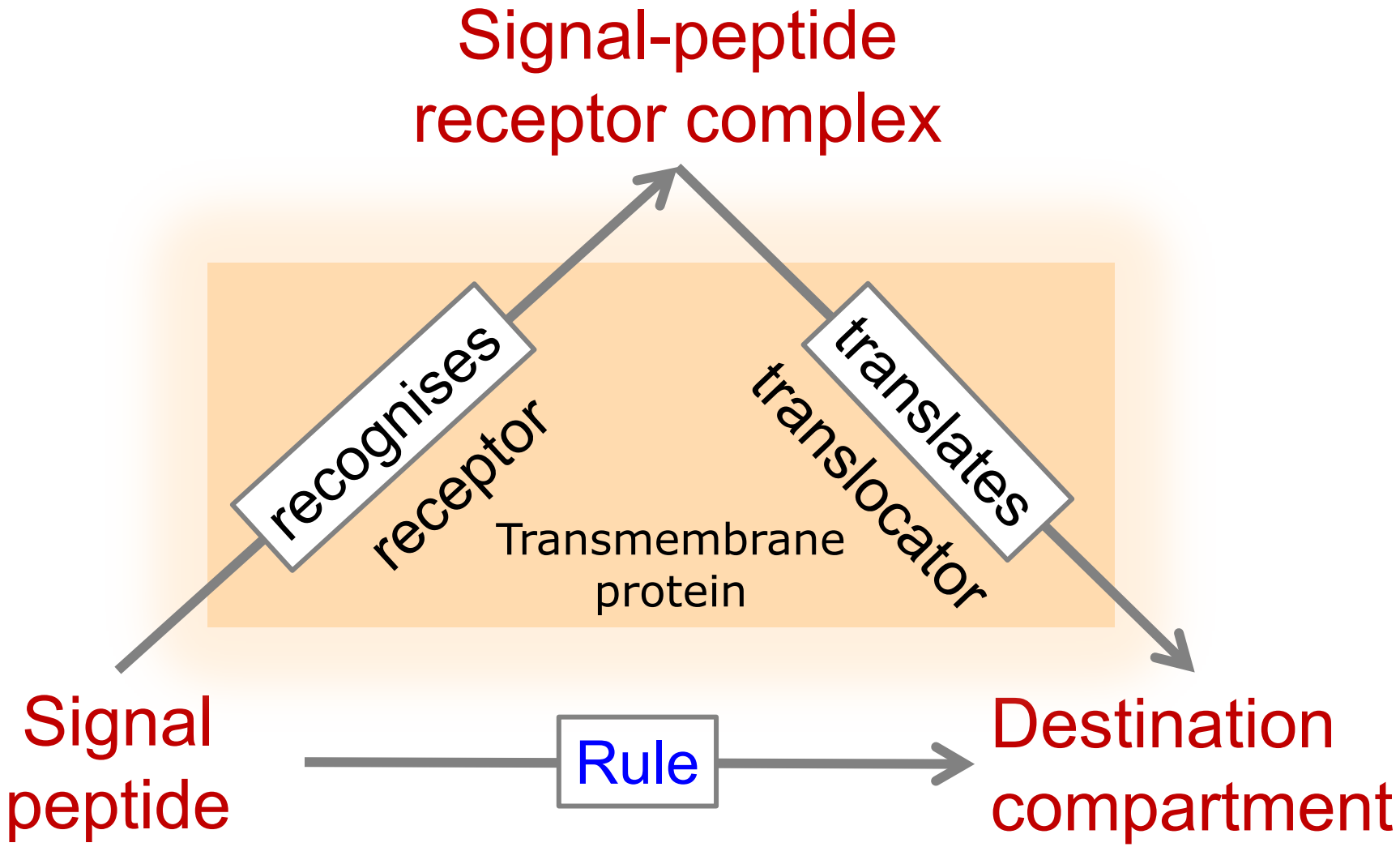


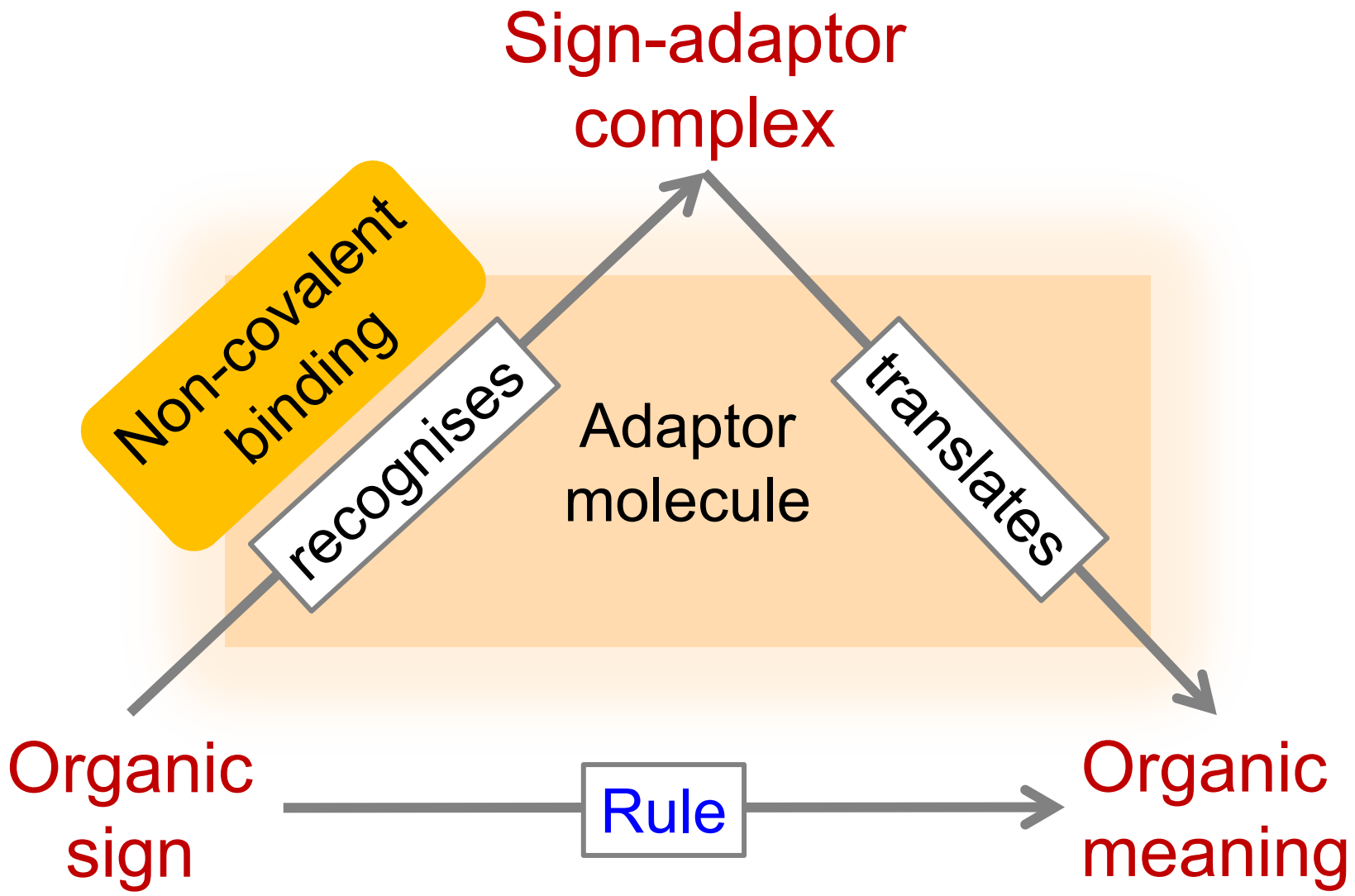
Protein targeting to the thylakoid in the chloroplast



Entities	Example: **Protein targeting code**
S (targeting peptides: presequences or internal sequences)	{TP1, TP2, TP3,…}
T (compartment targets)	{ER, Golgi, endosomes, mitochondria, chloroplasts, endosomes,..}
Code $C : S \longrightarrow T$	$\mathrm{TP1} \mapsto \mathrm{ER}$ $\mathrm{TP2} \mapsto \mathrm{Golgi}$ $\mathrm{TP3} \mapsto \mathrm{Endosomes}$ $\mathrm{TP4} \mapsto \mathrm{Mitochfondria}$ $\mathrm{TP5} \mapsto \mathrm{Chloroplasts}$ $\mathrm{TP6} \mapsto \mathrm{Peroxisomes}, \ldots$

Protein targeting code





No underlying chemical transformation

Dennis Görlich and Peter Dittrich

Computational science perspective
Molecular codes
- contingency in reaction networks

Gérard Battail

Information science perspective
Nested soft codes
- error correcting codes

Thank you for listening!