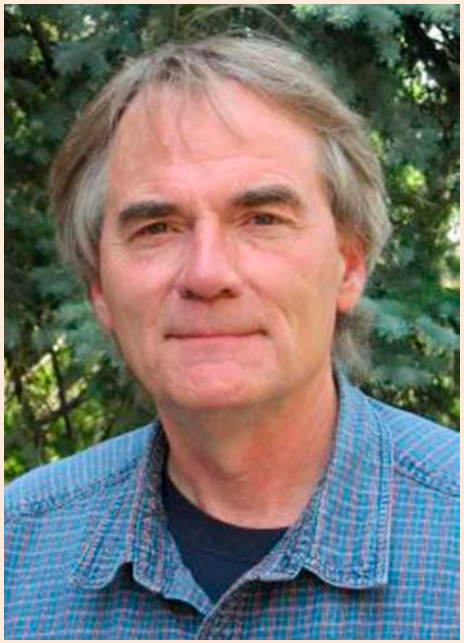
Design Triangulation as a Path for Biological Inquiry

Proposal for a Method of
Design-Enabled Biological Research

12 September 2020

Paul A. Nelson
Senior Fellow, Discovery Institute, www.discovery.org/csc
Adjunct Faculty, MA Program in Science & Religion
Biola University, www.biola.edu



Professor Scott Minnich
Microbiology
University of Idaho
USA

When I first met Scott Minnich in the mid-1990s, when I was still a graduate student, he told he something I have never forgotten:

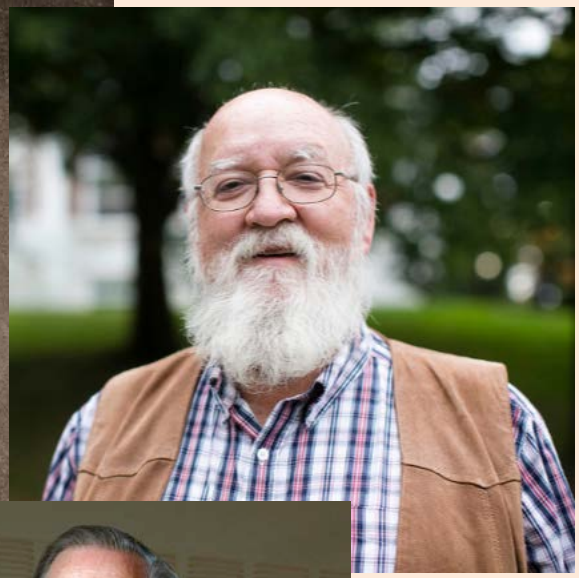
"Paul, none of this intelligent design debate should be all that controversial. **The reality, whether anyone is conscious of it or not, is this: most molecular biologists are *de facto* design theorists already."**

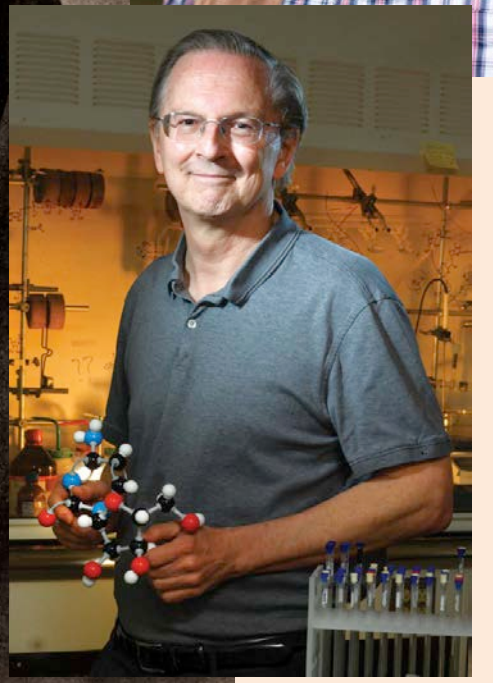
And, he added, they have been doing that for a very long time indeed.

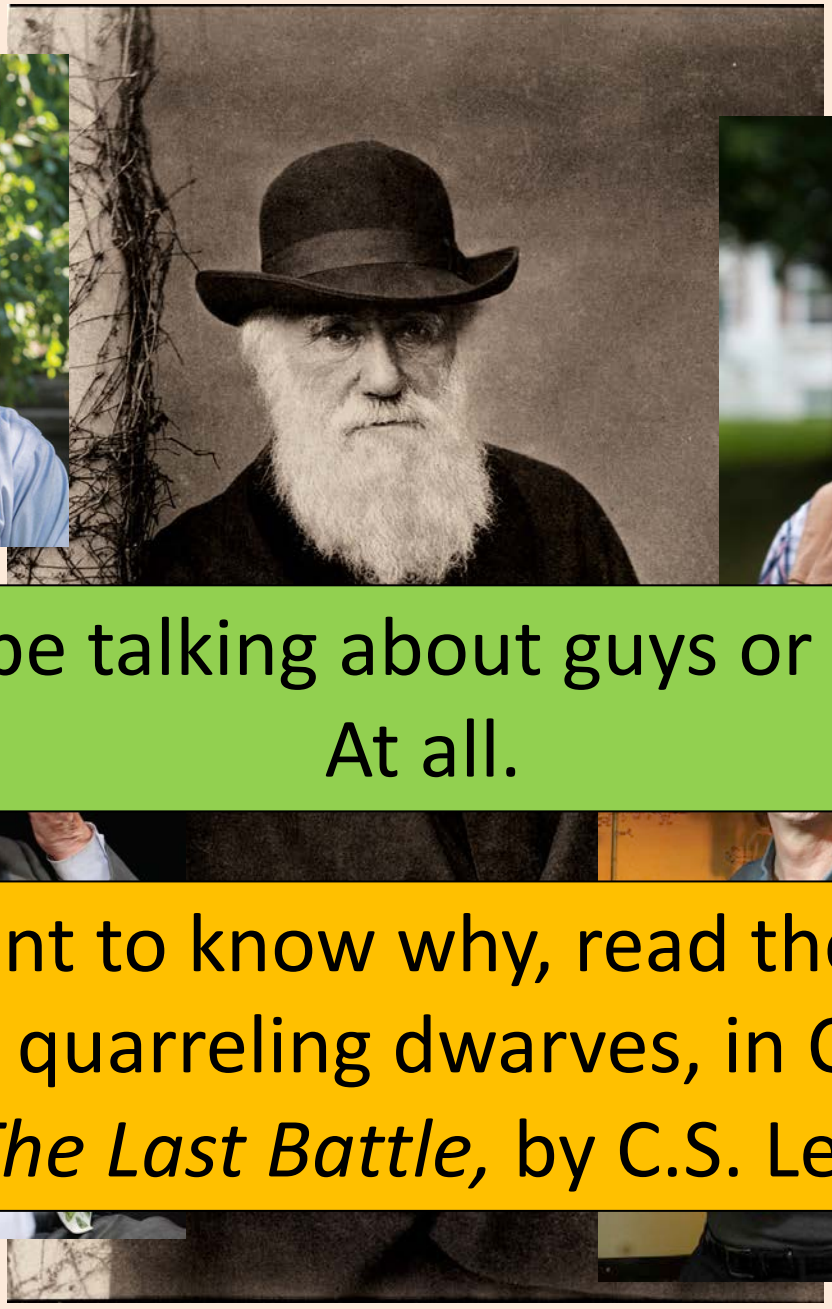
But how could that be possible – even as a *de facto* practice – when design was widely seen as intrinsically unscientific, employing a cause (i.e., a transcendent mind) understood to be unobservable in principle?

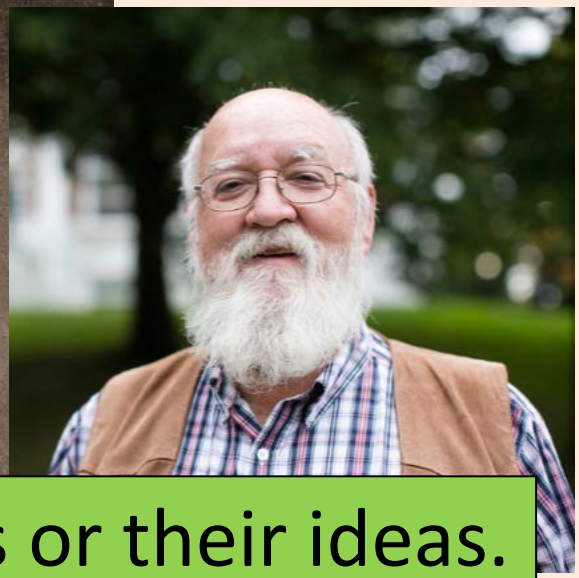
In this talk, I explain what Scott meant, and why it matters, using an idea called "design triangulation."











We won't be talking about guys or their ideas. At all.

If you want to know why, read the episode about the quarreling dwarves, in Chapter 13 of *The Last Battle,* by C.S. Lewis.

And that's pretty much
all you are going to hear
about evolution in this
presentation.

There's a bit more (for
explanatory purposes only)
in the big pdf coming next week.

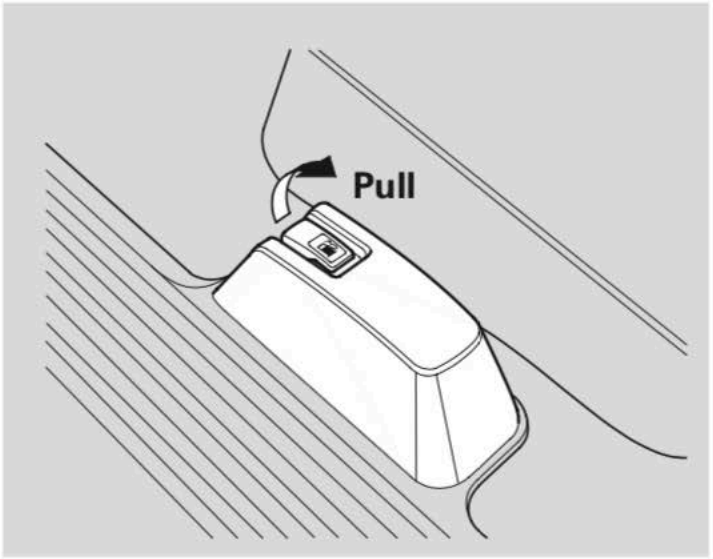
I.

Why won't the sliding door of my Honda Odyssey open?

The basic logic of functional triangulation

Let's start with my 2003 Honda Odyssey.

Filling the Fuel Tank



1. Because the fuel fill cap is on the driver's side of the vehicle, park with that side closest to the service station pumps.

2. Open the fuel fill door by pulling on the handle to the left of the driver's seat.

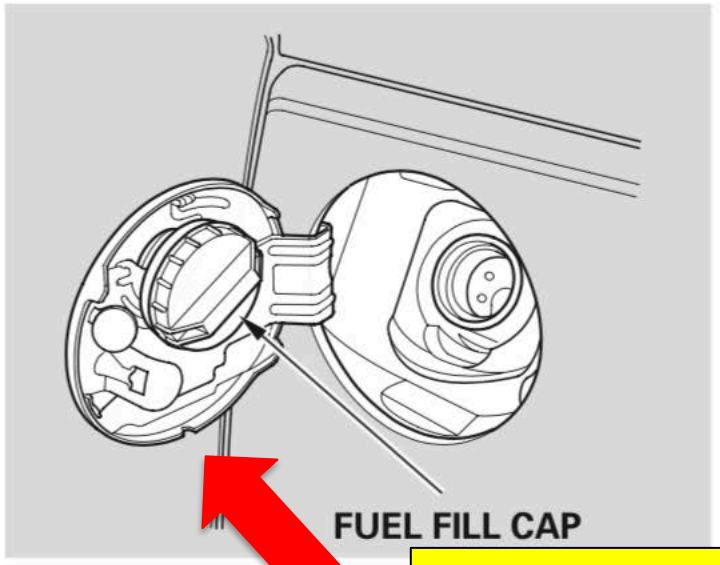
Before refueling, make sure the rear sliding door on the driver's side is closed.

BUT WHY?

⚠ WARNING

Gasoline is highly flammable and explosive. You can be burned or seriously injured when handling fuel.

- Stop the engine and keep heat, sparks, and flame away.
- Handle fuel only outdoors.
- Wipe up spills immediately.



(here's a clue)

Why do I need to make sure, Honda Corporation?

Because two macroscopic objects with mass (like doors) cannot occupy the same location in space at the same time.

And if you're stupid enough to try (Paul), *we've designed the car to deal with that.*

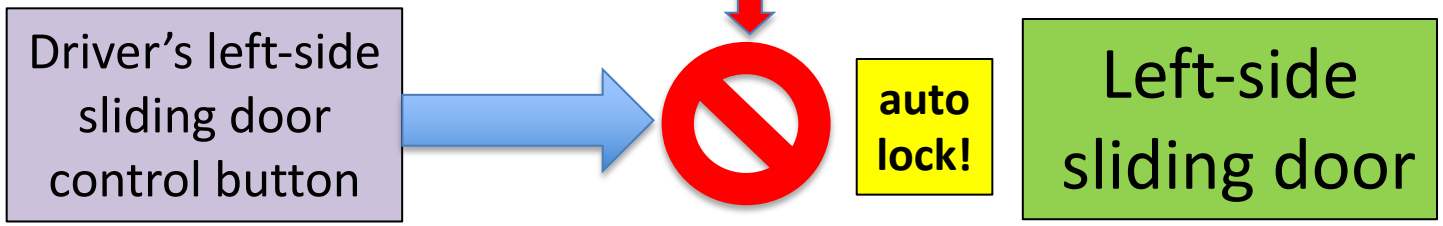
A Note About Refueling

Before refueling, make sure the driver's side sliding door is fully closed. When you release the fuel fill door, the driver's side sliding door automatically locks so it cannot open and interfere with the fuel door.

During the design ("foresight") stage of Odyssey development:

Honda engineers: hey, some drivers are clueless. They'll need help so they won't jam the driver's side sliding door.

Wait a minute, ***is the fuel door open?***



Now suppose you are Paul's younger brother Peter Nelson MD – who is a situationally aware and clued-in person, not a badly distracted philosopher of science.

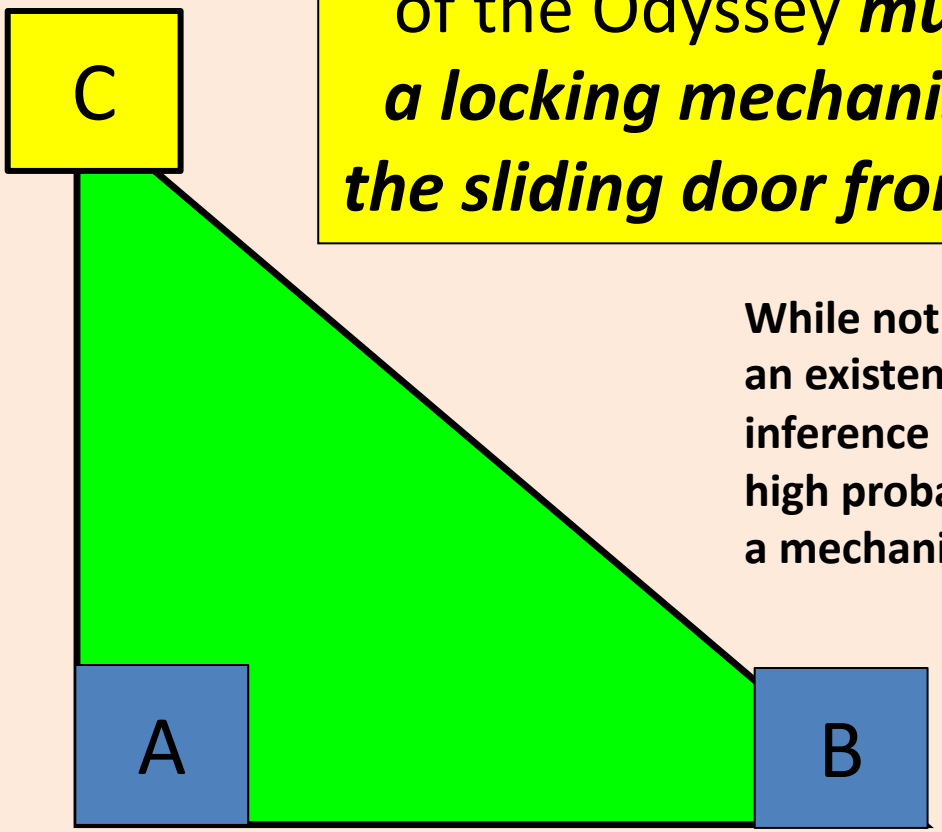
Peter, who is not an Odyssey owner, and has no experience with this vehicle, studies the minivan doors (sliding and fuel) while waiting for me in the driveway. He says to himself – hmm:

Honda must have a foolsafe mechanism to prevent door jamming.

Peter triangulates to an unobserved, but necessary, function of the system, as follows.

Functional triangulation (everyday version):

Therefore, the global design of the Odyssey ***must include a locking mechanism to stop the sliding door from jamming***.

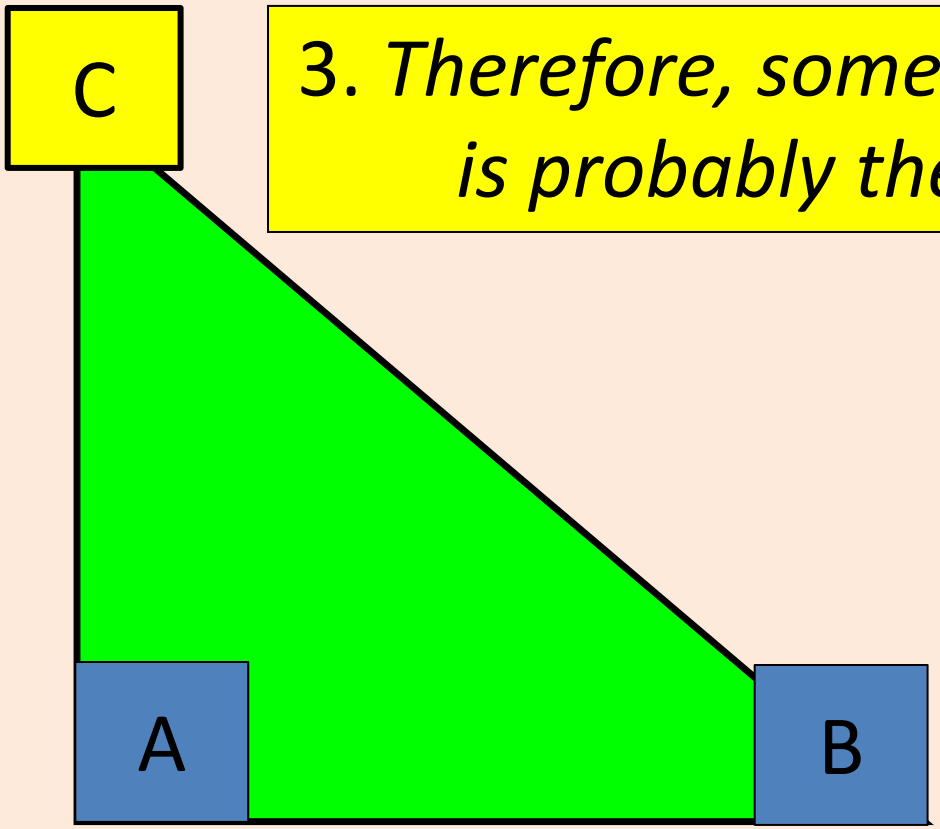


While not strictly speaking an existence proof, this inference confers a very high probability that such a mechanism operates.

The sliding door moves to left rear, over the fuel door.

If the fuel door is open, the sliding door would jam.

Triangulation as a metaphor for scientific inquiry:

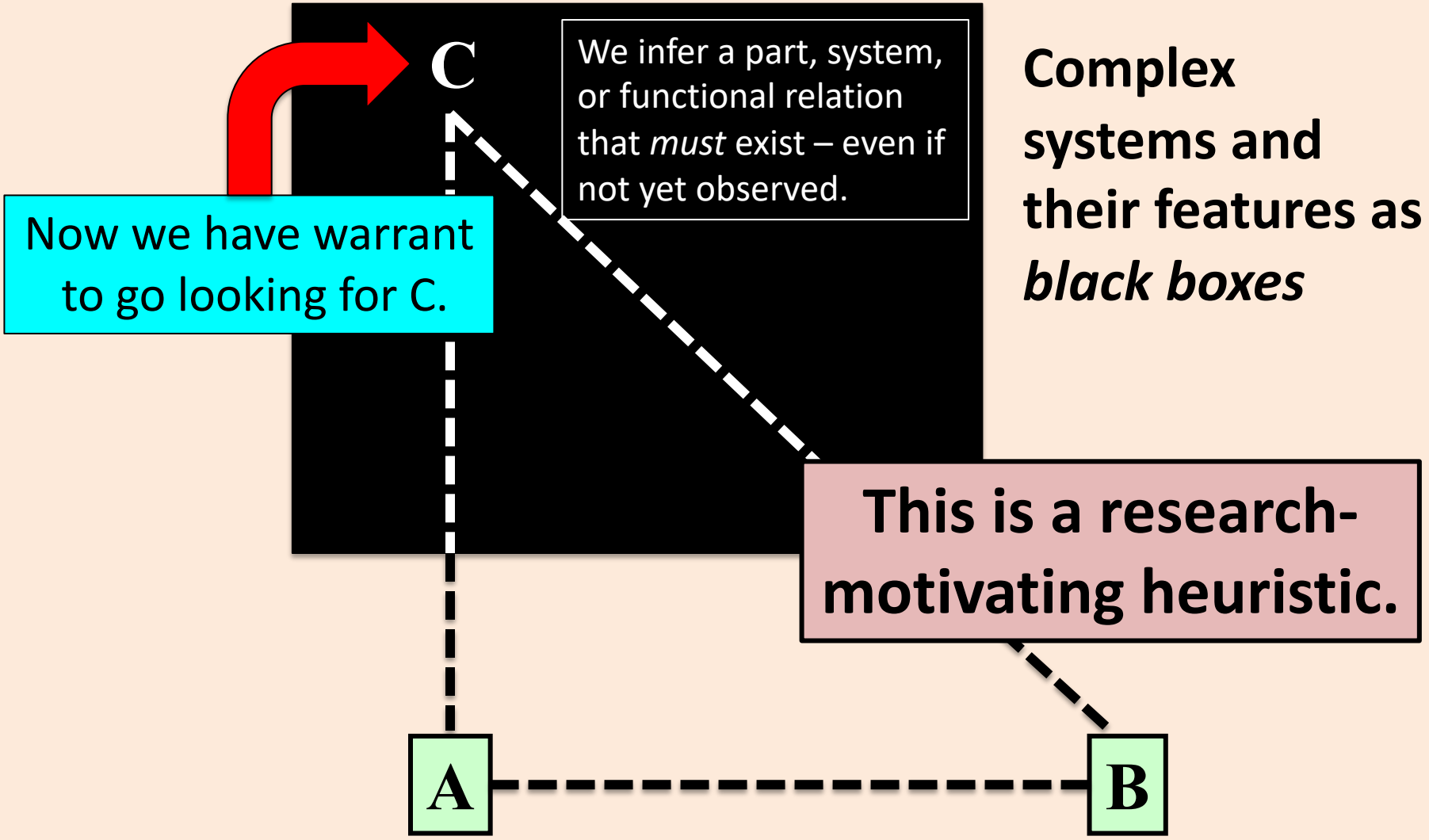


3. *Therefore, something like C is probably the case.*

1. We know A.

2. We know B.

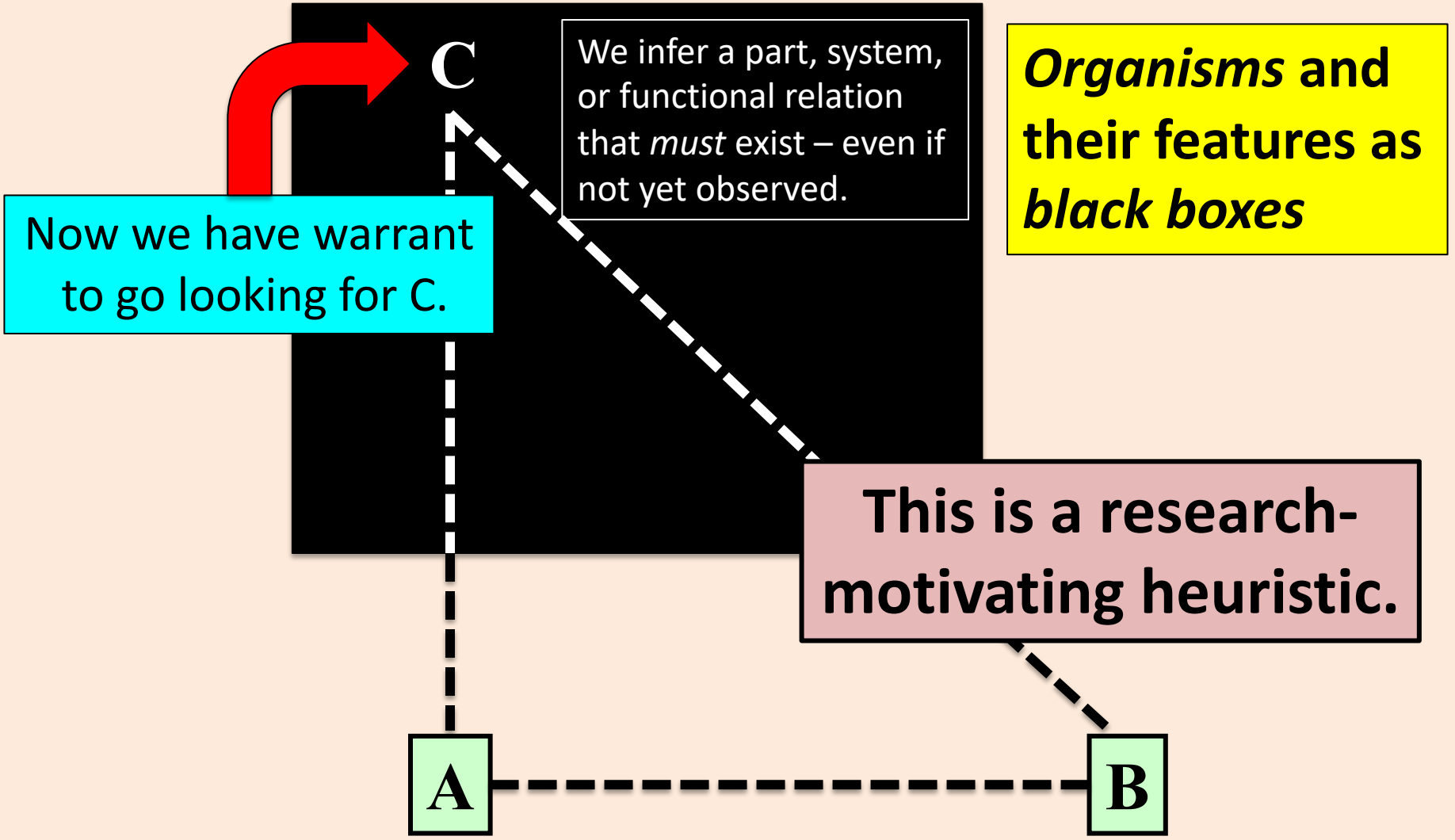
Functional triangulation



Complex systems and their features as *black boxes*

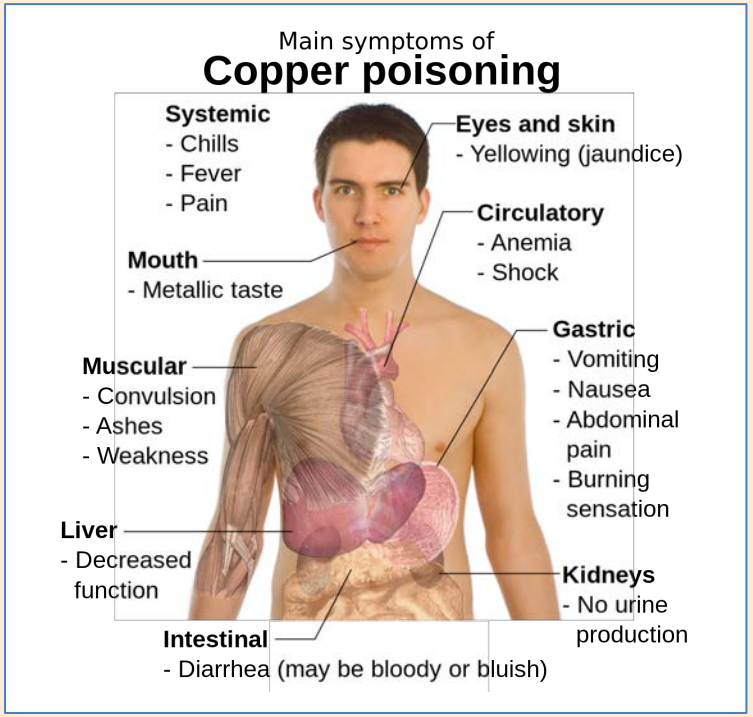
II.
Functional triangulation in biology

Let's try a thought experiment.



Consider a pair of biologically related facts:

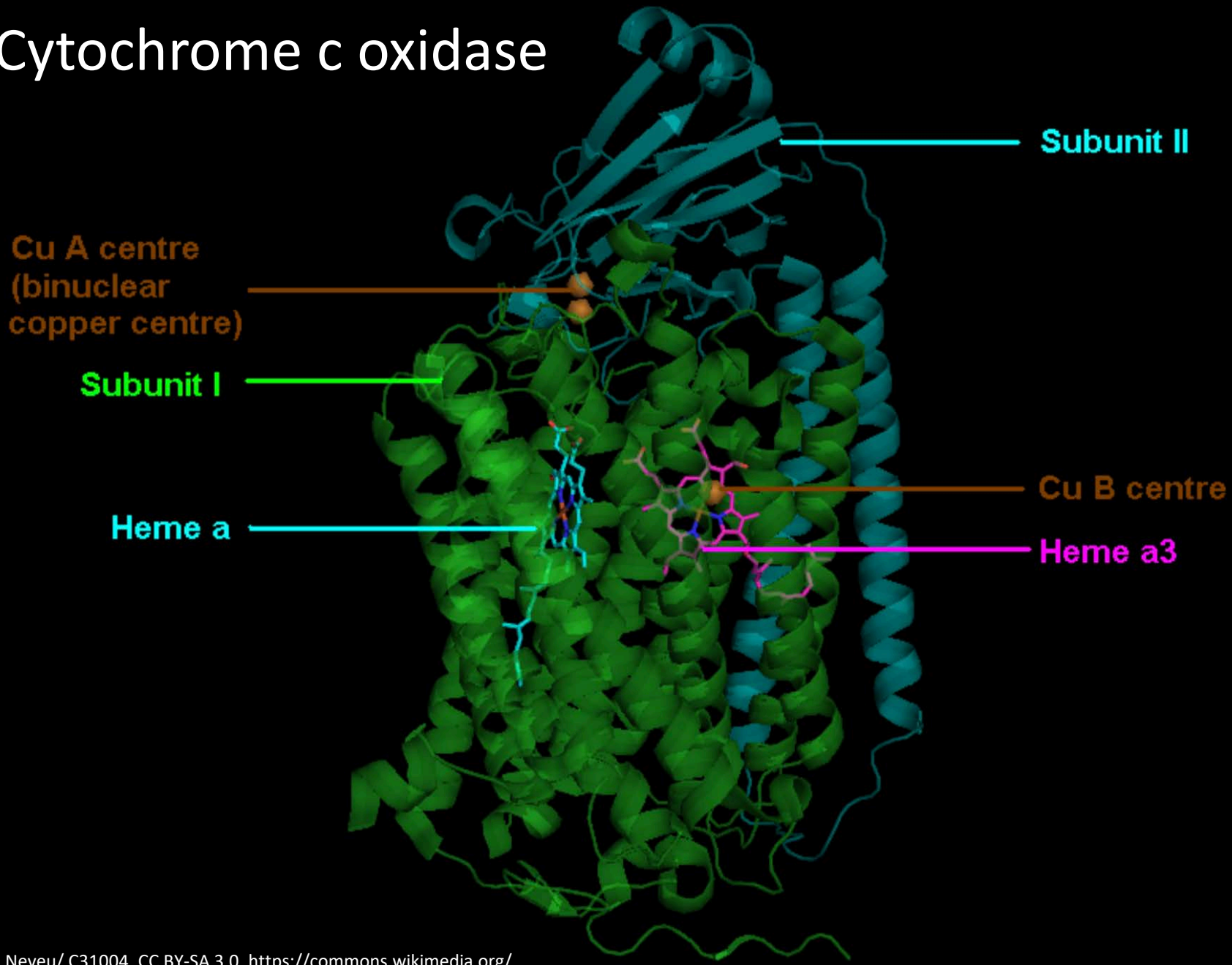
1. **Copper is a poison.**



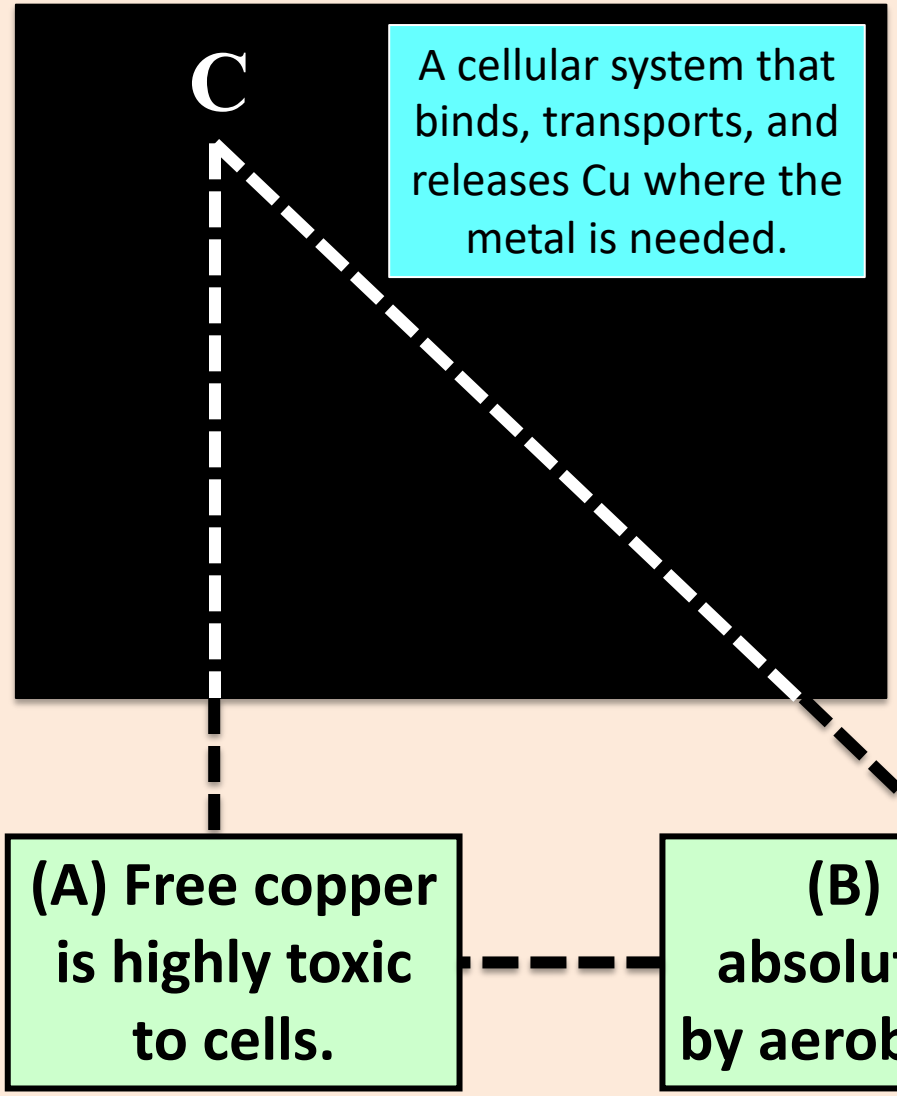
2. **Copper is absolutely required by aerobically respiring organisms (like you), as an essential co-factor in several enzymes.**

(illustration credit Mikael Häggström, Wikicommons)

Cytochrome c oxidase



Trust the logic: triangulate.



What MUST exist in aerobically respiring organisms?

“But wait a second, Paul…how can you be *certain* the copper-binding system is really there in the cell?”

Don’t worry: it’s there.

Organisms make use of mechanisms for their very viability; whether we see the mechanisms or not, *they exist*.

Call this the ***no magic principle***.

This "existence proof" for a Cu binding & transport system was acknowledged *long before* the system was actually observed:

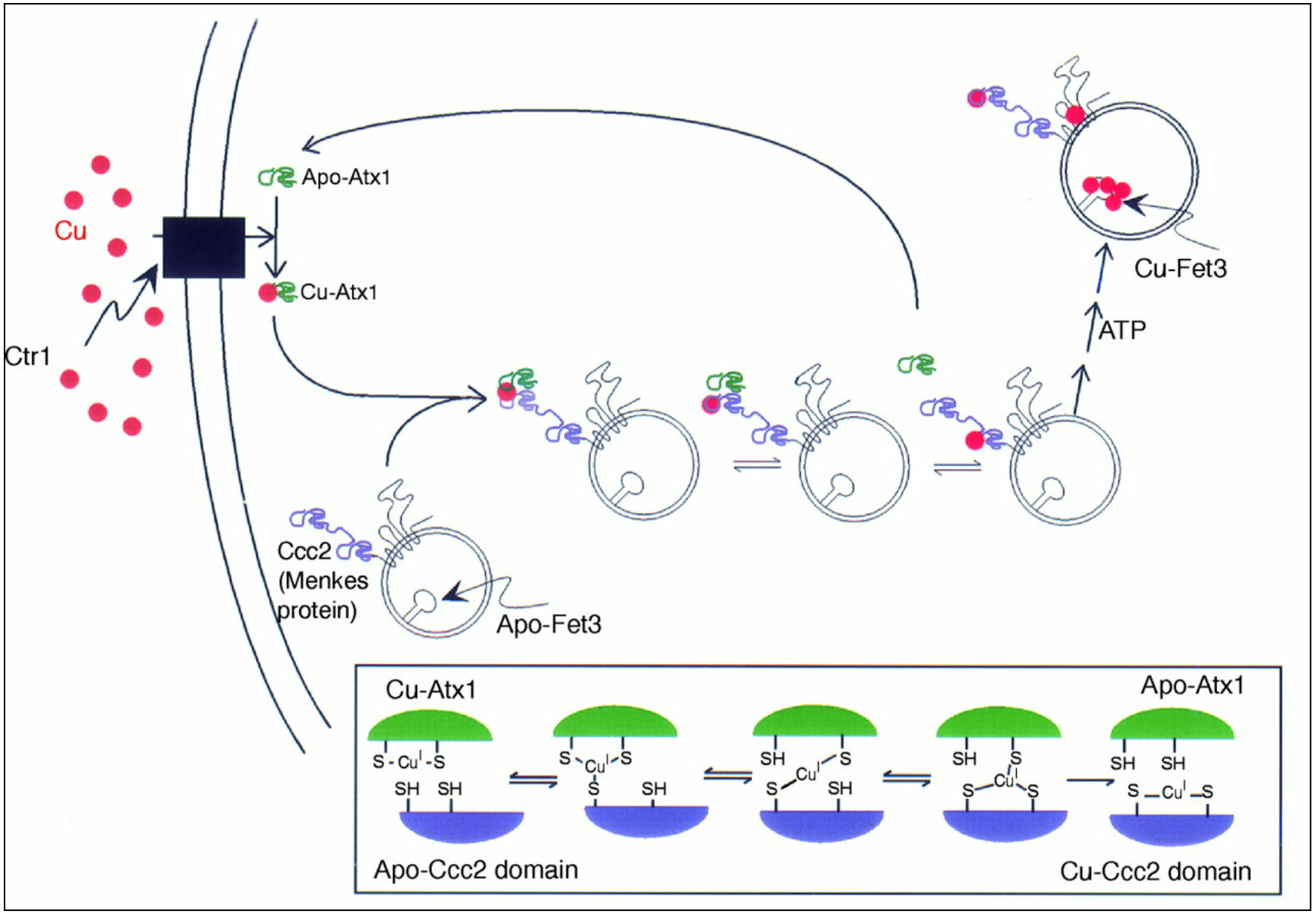
"Copper is absolutely required for aerobic life, and yet, paradoxically, is highly toxic. … This apparent contradiction has been rationalized by assuming that Cu, like other redox-active metals, is sequestered in nonactive forms as it is transported through cellular compartments."

Valentine and Gralla, *Science* 278 (1997):817.

This "existence proof" for a Cu binding & transport system was acknowledged *long before* the system was actually observed:

"However, the agents of such trafficking and the mechanisms of delivery of Cu to its final destinations have, until recently, remained largely unknown."

Valentine and Gralla, *Science* 278 (1997):817.



Valentine and Gralla, *Science* 278 (1997):817.

and F. Whitby for assistance with data collection; V. Ramakrishnan and members of the Sundquist and Hill laboratories for critical comments on the manuscript; M. Martin for plasmid pNL4-3; D. Trono for plasmid R9; and J. Cassatt for support and encouragement. Supported by NIH grants RO1 AI40333 and RO1 AI43036 (W.I.S. and C.P.H.), the Lucille P. Markey Charitable Trust, and a postdoctoral fellowship from the Cancer Research Institute (T.R.G.). Coordinates (1am3) and diffraction data (r1am3sf) for GA151–231) have been deposited in the Protein Data Bank (Brookhaven National Laboratory).

17 March 1997; accepted 23 September 1997

Metal Ion Chaperone Function of the Soluble Cu(I) Receptor Atx1

R. A. Pufahl, C. P. Singer, K. L. Peariso, S.-J. Lin, P. J. Schmidt, C. J. Fahrni, V. Cizewski Culotta, J. E. Penner-Hahn, T. V. O'Halloran*

Reactive and potentially toxic cofactors such as copper ions are imported into eukaryotic cells and incorporated into target proteins by unknown mechanisms. Atx1, a prototypical copper chaperone protein from yeast, has now been shown to act as a soluble cytoplasmic copper(I) receptor that can adopt either a two- or three-coordinate metal center in the active site. Atx1 also associated directly with the Atx1-like cytosolic domains of Ccc2, a vesicular protein defined in genetic studies as a member of the copper-trafficking pathway. The unusual structure and dynamics of Atx1 suggest a copper exchange function for this protein and related domains in the Menkes and Wilson disease proteins.

Although Cu is an essential cofactor for mitochondrial, cytosolic, and vesicular oxygen-processing enzymes (*1*), it can be toxic even at low concentrations. Cu(I) and Cu(II) ions can bind with high affinity to adventitious sites in partially folded proteins and catalyze auto-oxidation of lipids, proteins, and nucleic acids. To investigate the mechanisms by which cells overcome the dilemma of maintaining Cu availability while controlling deleterious reactivity of the free ions, we have determined the Cu chemistry and Cu-handling function of Atx1, an intracellular eukaryotic protein implicated in Cu trafficking. Our results indicate that Atx1 functions as a metal ion chaperone, a protein that protects and guides Cu(I) ions to activate target enzymes.

ATX1 is one of several genes in the Cu-dependent, high-affinity iron uptake pathway in yeast. These genes encode Ctr1, a Cu uptake protein in the plasma membrane; Ccc2, an intracellular membrane protein; and the multicopper oxidase Fet3 (*2*–*6*). Although the role of Fet3 in iron uptake is unclear, Cu loading into this enzyme occurs in a post-Golgi vesicle and is mediated by Ccc2 (*4*). The *ATX1* gene and its human homolog, *HAH1*, encode cytosolic proteins implicated in Cu trafficking to these Ccc2-containing vesicles (*7*–*9*). The Ccc2 protein in yeast and its human homologs, the Menkes disease protein (*10*) and Wilson disease protein (*11*), are members of the P-type adenosine triphosphatase (ATPase) cation transporter family (*12*) and are present in the membranes of secretory vesicles (*13*). Each transporter contains two or more Atx1-like cytoplasmic domains (Fig. 1), the functions of which are not known. The conserved MTCXXC sequence (X, any residue), a motif observed in several bacterial Hg(II) transport proteins (*14*), is thought to be a Cu binding site, although it does not correspond to known Cu(I) or Cu(II) sites in structurally characterized proteins. Establishing the Cu oxidation state and coordination environment in Atx1 should provide insight into its function and that of the homologous domains in its partner proteins.

Atx1 was expressed in *Escherichia coli* and purified as the apoprotein (*15*). Addition of $CuSO_4$ to apo-Atx1 in the presence of excess thiol reproducibly yielded a complex with a copper/protein ratio of 0.6 to 0.8 (*16*). Gel filtration experiments indicated that the predominant form of the protein was a monomer under these conditions, regardless of the presence or absence of Cu (*15*). The Cu oxidation state was investigated by electron paramagnetic resonance (EPR) spectroscopy and x-ray absorption spectroscopy (XAS). No EPR signal was observed at 77 K, suggesting the absence of a mononuclear Cu(II) site. XAS experiments (*17*) indicated that the bulk sample contained Cu(I). The Cu-Atx1 x-ray absorption near edge structure (XANES) spectrum exhibited a weak shoulder at 8984 eV, which is typical of Cu(I) and inconsistent with comparable edge features for Cu(II) compounds, the energies of which are 3 to 4 eV greater (Fig. 2A). The intensity of the 8984-eV transition varies with Cu(I) geometry, ranging from low for tetrahedral sites to high and well resolved for digonal sites (*18*). The observed transition is typical of those of trigonal Cu(I) model compounds (Fig. 2A). Mononuclear Cu(I) coordination complexes, however, are usually not stable in aqueous solution, and typically undergo auto-oxidation or disproportionation reactions to give Cu(II) (aqueous) and Cu(solid). In contrast, millimolar concentrations of Cu(I)-Atx1 are stable in air at neutral pH for at least 30 min (*16*), suggesting that the coordination environment in Atx1 stabilizes the Cu(I) state and suppresses disproportionation.

Proteins that stabilize Cu(I) form either polynuclear metal thiolate clusters, as in metallothionein and the transcription factors Ace1, CUP2, and AMT (*19*), or a constrained His_3Cys coordination environment, as in blue copper proteins (*20*). In contrast, extended x-ray absorption fine structure (EXAFS) measurements indicated that Atx1 stabilization of Cu(I) is achieved in a mononuclear site through an all-sulfur coordination environment (Fig. 2B). The high intensity and relatively high frequency of the EXAFS oscillations were typical of those observed for Cu-S interactions. The data could be modeled with a single three-sulfur (3S) shell with a Cu-S distance of 2.26 Å; however, the Debye-Waller factor, σ^2, was somewhat larger than expected (7×10^{-3} Å^2). This Cu-S distance is typical of three-coordinate Cu (*21*–*23*) and is 0.1 Å longer than that in two-coordinate Cu-thiolate complexes (*24*). Given that Atx1 contains only two conserved cysteine residues, the third ligand can be either a low-Z (atomic number) ligand (O or N), a methionine sulfur, or an exogenous thiol. A 2S + 1S fit with a shell of two sulfurs at 2.25 Å and a shell of one sulfur at 2.40 Å reproduced the data better than did the single-shell fit, and gave more reasonable σ^2 values (3×10^{-3} and 5×10^{-3} to 6×10^{-3} Å^2, respectively). Furthermore, attempts to model the data with a shell of 2S and a low-Z shell (1N) gave chemically unrealistic Debye-Waller factors for the N shell ($\sigma^2 \approx 0$).

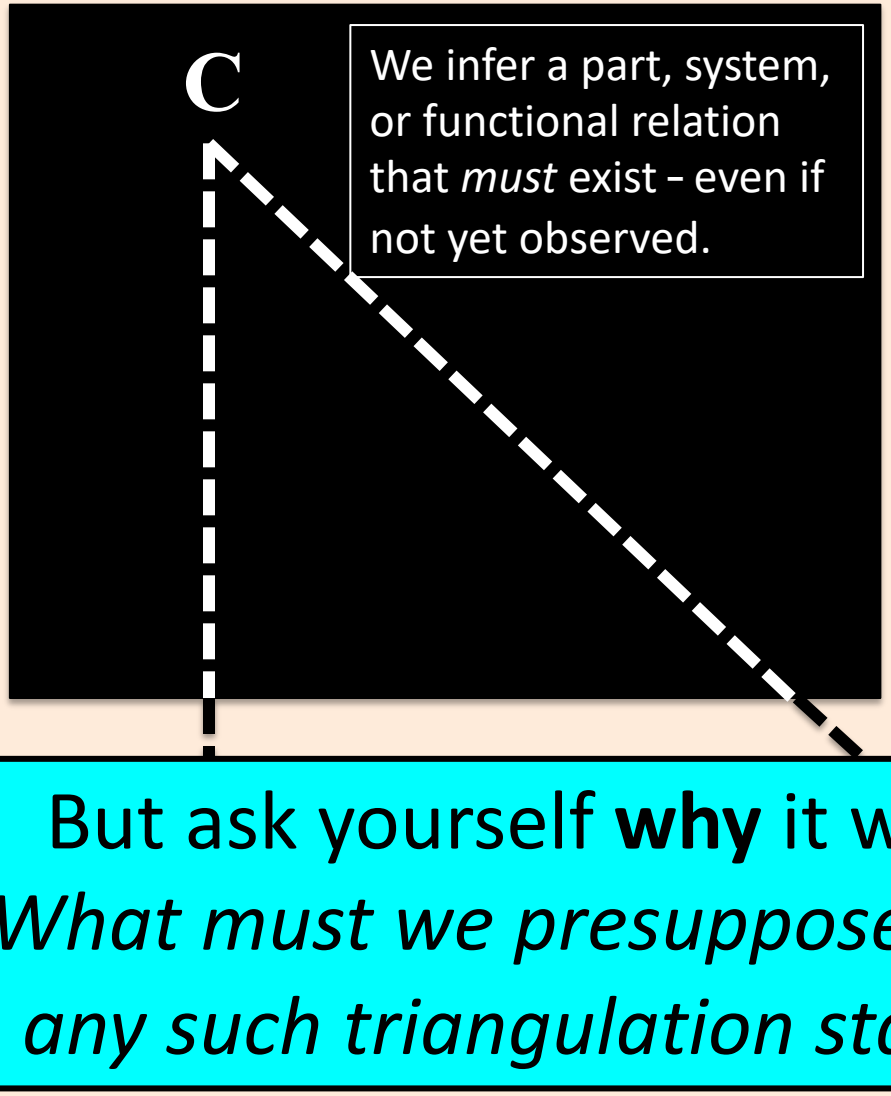
The 2S + 1S result is unexpected for several reasons. The most common coordination environment for Cu(I) complexes is distorted four-coordinate, and there are few precedents among mononuclear Cu(I) complexes for the low-coordination number environ-

R. A. Pufahl, C. P. Singer, C. J. Fahrni, T. V. O'Halloran, Department of Chemistry and Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, IL 60208, USA.
K. L. Peariso and J. E. Penner-Hahn, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA.
S.-J. Lin, P. J. Schmidt, V. Cizewski Culotta, Department of Environmental Health Sciences, Johns Hopkins University, Baltimore, MD 21205, USA.
*To whom correspondence should be addressed. E-mail: t-ohalloran@nwu.edu

R.A. Pufahl *et al*., "Metal Ion Chaperone Function of the Soluble Cu(I) Receptor Atx1," *Science* 278 (1997):853-856

By delivering essential cofactors or substrates to apoenzymes, the emerging class of metal ion chaperones facilitates formation of an active state of a protein. These chaperones guide metal ions to their appropriate biological partners and protect them from being trapped at adventitious sites. They may also protect cellular components by sequestering specific ions or inorganic clusters (34) and preventing adverse reactions. Our results underscore the idea that cells make use of elaborate machinery for recruiting, trafficking, compartmentalizing, and, ultimately, inserting into the appropriate enzyme reactive cofactors such as mononuclear Cu ions. The intracellular activation of apometalloenzymes by binding of the correct metal ion cofactors is unlikely to proceed by spontaneous self-assembly. Rather, metal insertion is emerging as an orchestrated event controlled by metal ion transport and chaperone proteins whose functions are kinetically and thermodynamically coupled.

This works. **The big pdf has more examples, from the history of biology and today.**



Organisms and their features as *black boxes*

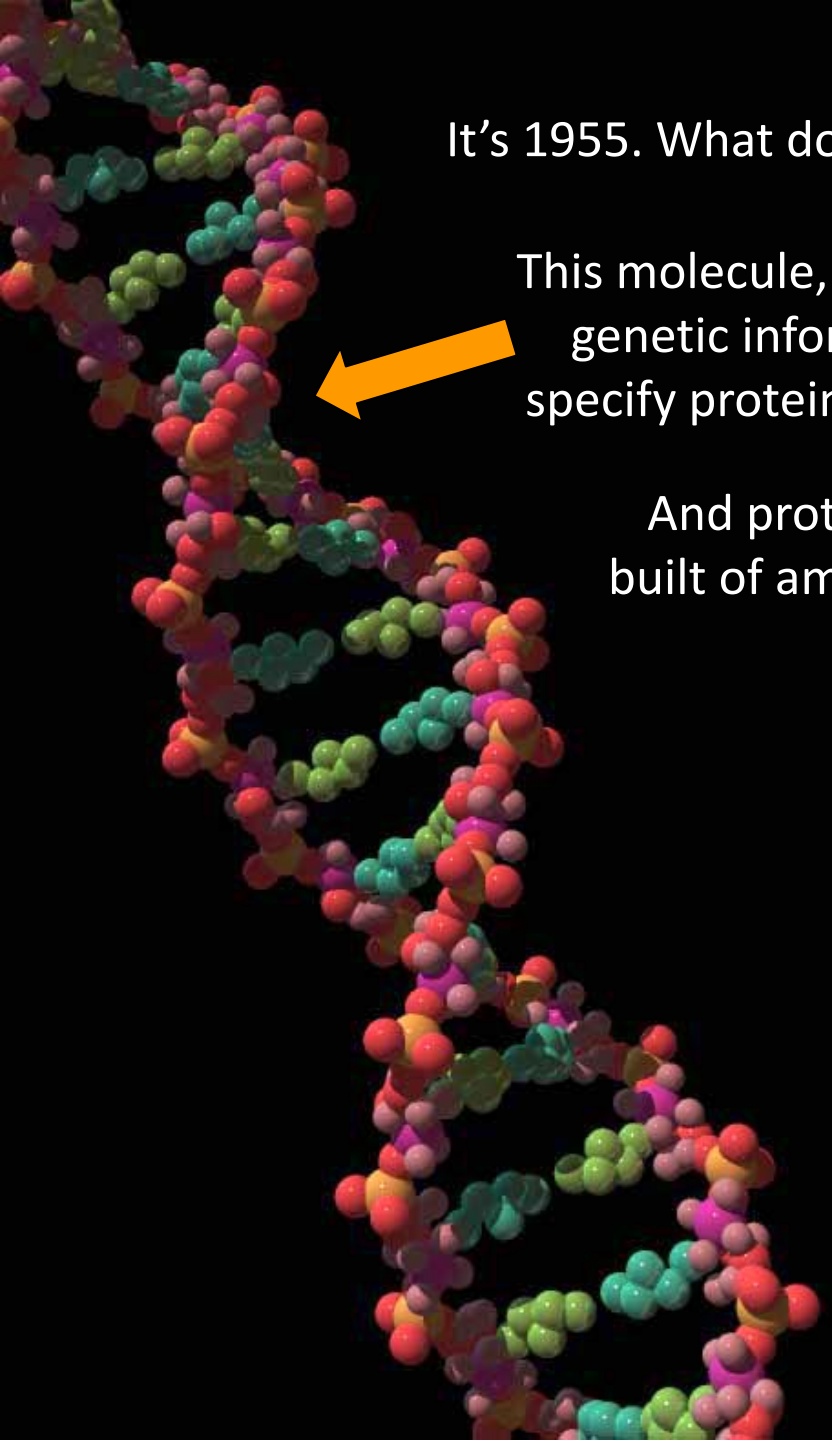
But ask yourself **why** it works.
What must we presuppose to get any such triangulation started?

The next example
is the most telling – and
also carries the answer to
the question “Why does
functional triangulation ***work***
in biology...and why should
we call it ***design*** triangulation?”

III.
The luckiest guess in the history of science – ever.

What I'm going to describe next wasn't a ***guess*** at all, of course.

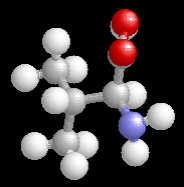
It was a rational inference, using functional triangulation.



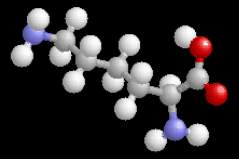
It's 1955. What do biologists know?

This molecule, DNA, carries genetic information to specify protein sequences.

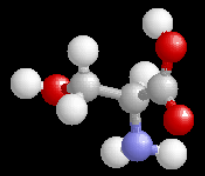
And proteins are built of amino acids.



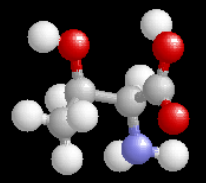
Valine



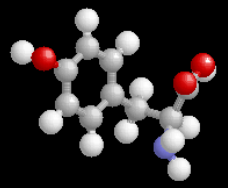
Lysine



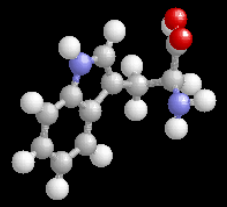
Serine



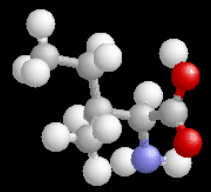
Threonine



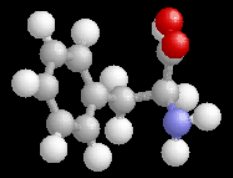
Tyrosine



Tryptophan



Isoleucine



Phenylalanine

But what is mediating information transfer between these two very different chemistries?

(photo credit: Special Collections, Oregon State University)

F.H.C. Crick

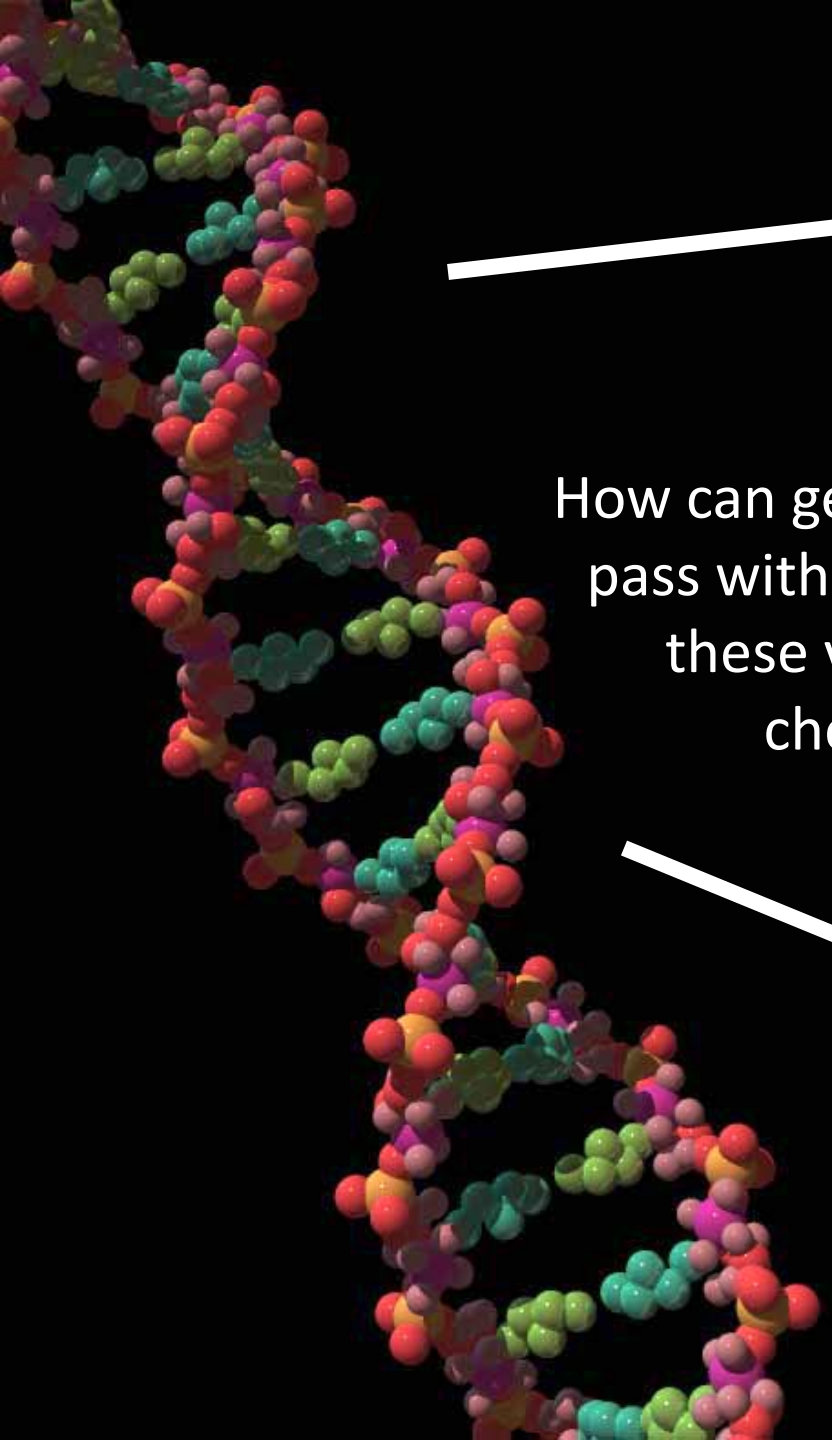
ON DEGENERATE TEMPLATES AND THE ADAPTOR HYPOTHESIS

F.H.C. Crick,

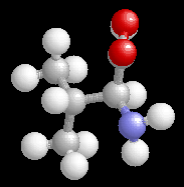
Medical Research Council Unit for the Study of
the Molecular Structure of Biological Systems,

Cavendish Laboratory, Cambridge, England.

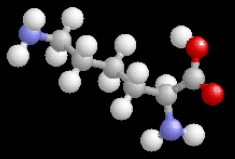
(Crick's original manuscript is available for downloading as a pdf here:
https://profiles.nlm.nih.gov/spotlight/sc/catalog/nlm:nlmuid-101584582X73-doc)



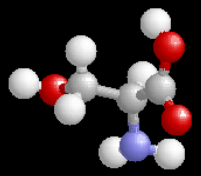
How can genetic information pass with fidelity between these very different chemistries?



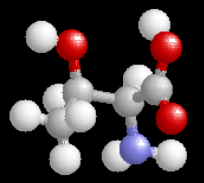
Valine



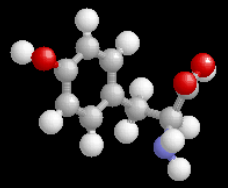
Lysine



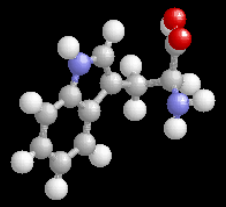
Serine



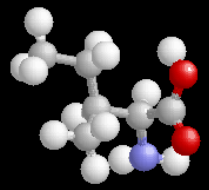
Threonine



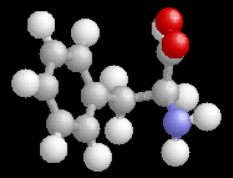
Tryptophan



Tyrosine



Phenylalanine



Isoleucine

Crick (1955) grasps the difficulty:

"Now what I find profoundly disturbing is that I cannot conceive of *any* structure (for either nucleic acid) acting as a direct template for amino acids, or at least as a specific template....I don't think anybody looking at DNA or RNA would think of them as templates for amino acids."

Crick (1955) grasps the difficulty:

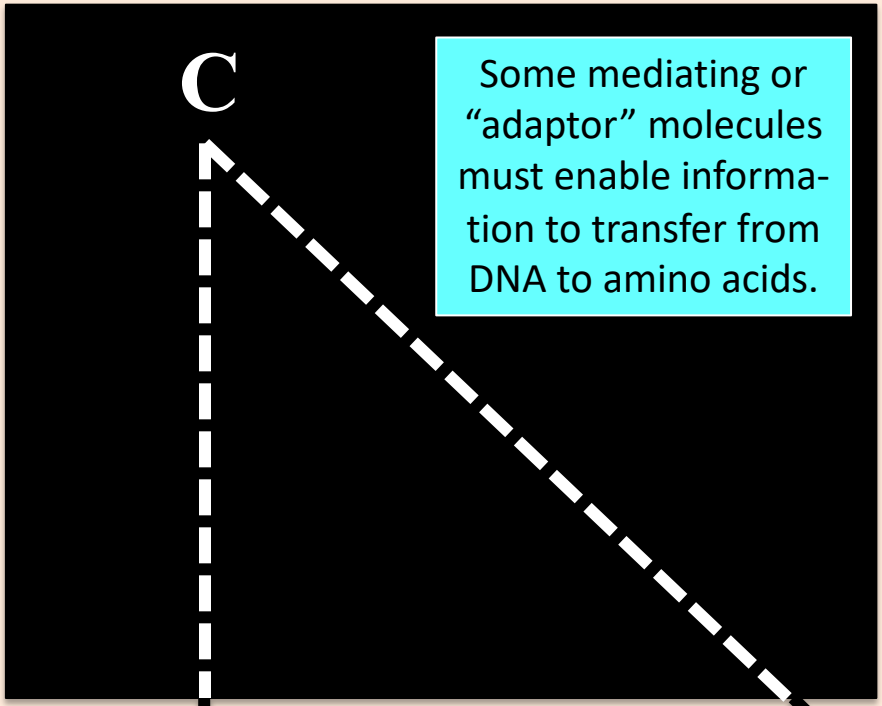
"Where are the knobly hydrophobic surfaces to distinguish valine from leucine and isoleucine? Where are the charged groups, in specific positions, to go with the acidic and basic amino acids?...What the DNA structure *does* show (and probably RNA will do the same) is a specific pattern of *hydrogen bonds*, and very little else."

Functional triangulation, however, underwrites inferences to unobserved entities which ***must*** exist:

When one discovers a complex system performing specialized functions, *assume that a rational logic, and well-matched parts, are enabling the functions.*

No magic. Look for the mechanism: it's there.

Crick triangulates from the cell's information-transfer requirements to its unobserved, but functionally necessary, parts:



Okay, now we can go looking for the system. Here are some of the features it will need to specify each of 20 amino acids in protein assembly...

(A) DNA carries information, but is chemically non-specific ("hydrogen bonds, little else")

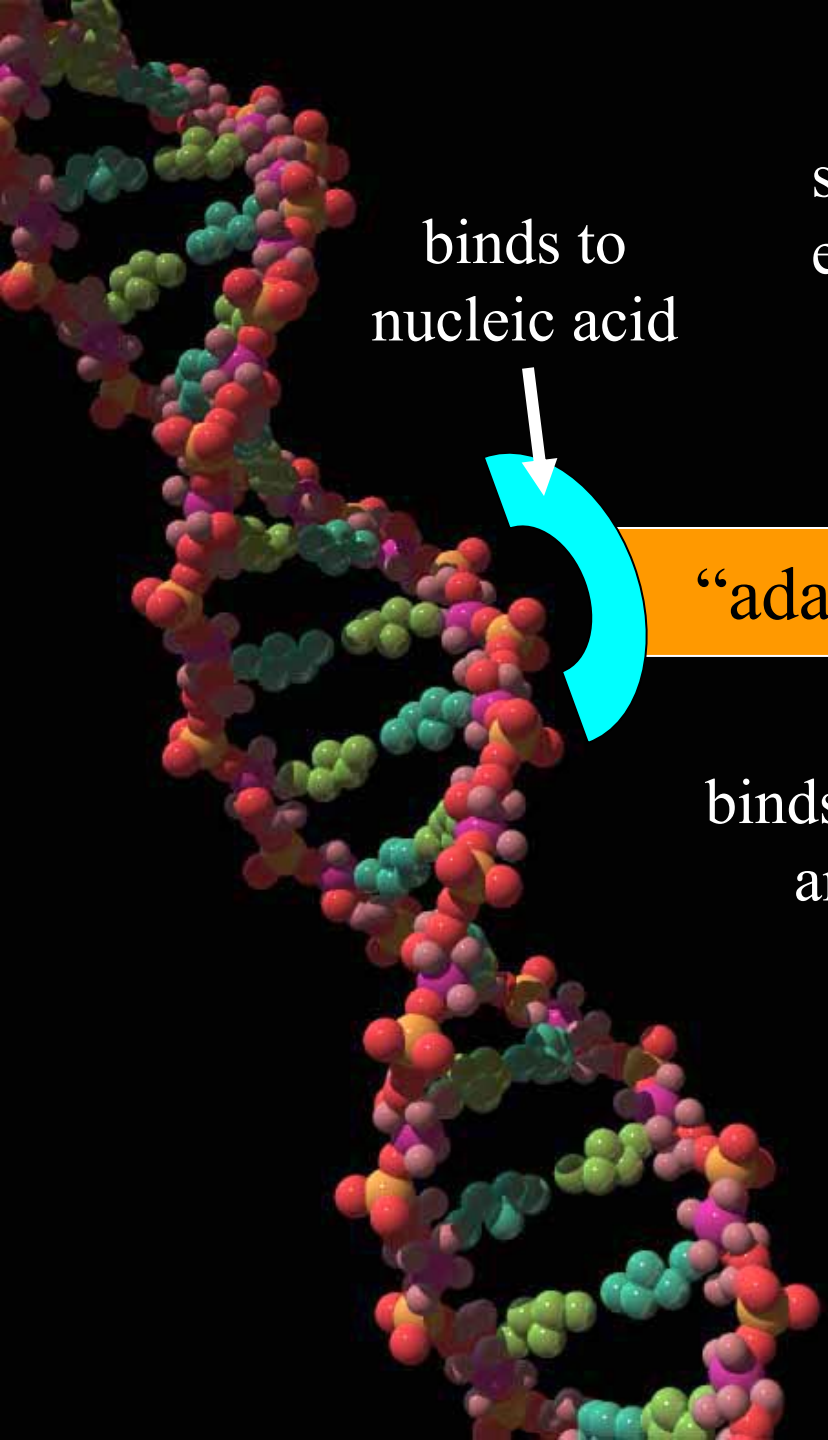
(B) Yet amino acids have specific geometries, which require a system to recognize them.

Crick boldly triangulates:

"...each amino acid would combine chemically, at a special enzyme, with a small molecule which, having a specific hydrogen-bonding surface, would combine specifically with the nucleic acid template.... In its simplest form there would be 20 different kinds of adaptor molecule...."

Crick boldly triangulates:

"...one for each amino acid, and 20 different enzymes to join the amino acid to their adaptors. Sydney Brenner, with whom I have discussed this idea, calls this 'the adaptor hypothesis' since each amino acid is fitted with an adaptor to go on to the template."

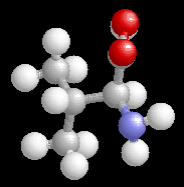


binds to nucleic acid

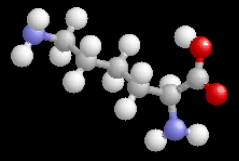
works with a special dedicated enzyme (protein)

"adaptor"

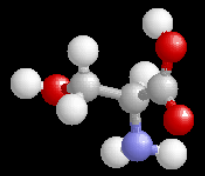
binds to a *specific* amino acid



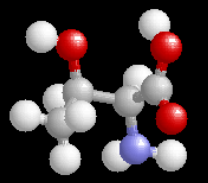
Valine



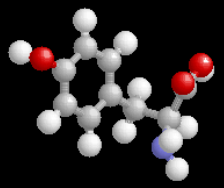
Lysine



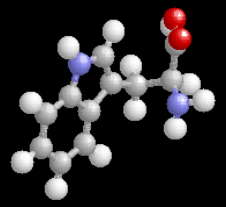
Serine



Threonine

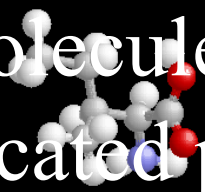


Tyrosine



Tryptophan

Phenylalanine



Isoleucine

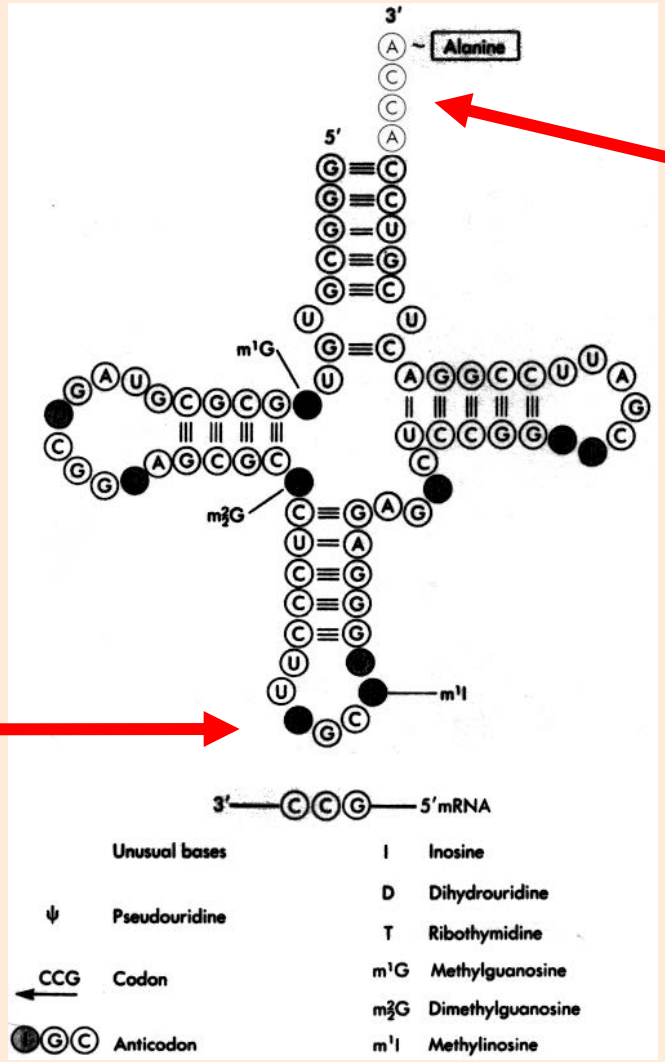
And, by the way – we need 20 of these molecules, with 20 specially dedicated proteins, one for each amino acid.

Crick: there is no evidence yet, but some such mediating molecules ***must*** exist.

"The usual argument presented against this latter scheme is that no such small molecules have been found, but this objection cannot stand."

The inference from *systems-level functional necessity* is very strong.

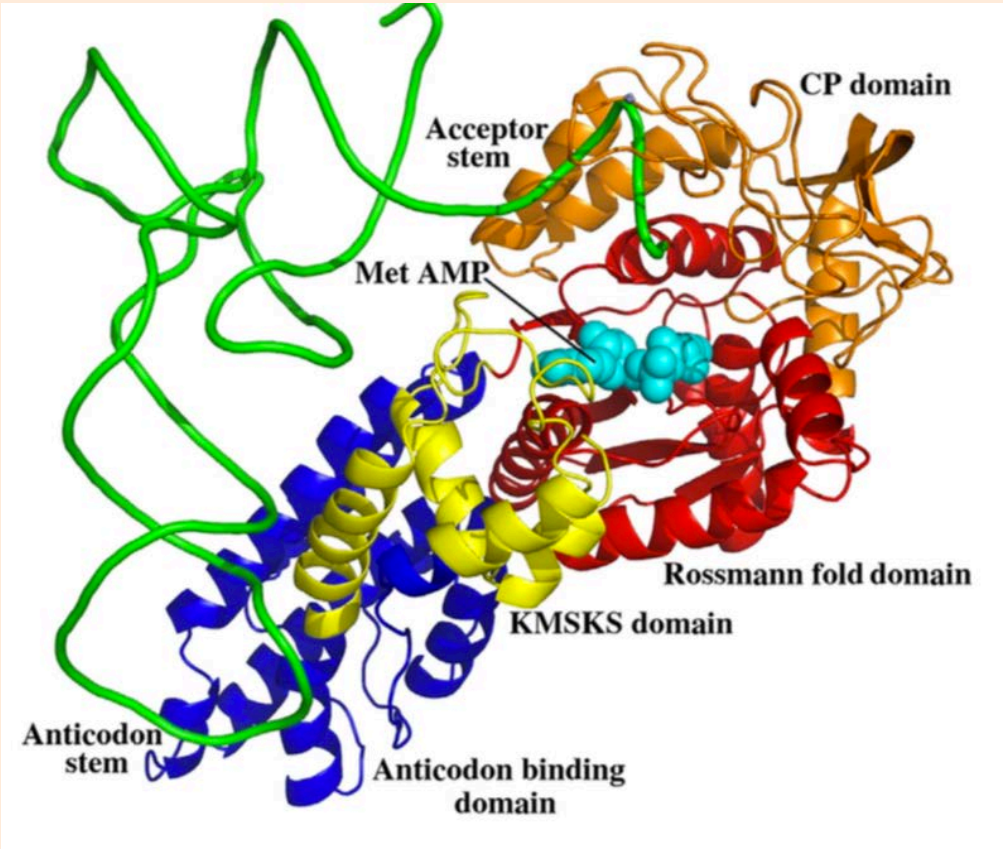
What came to be known as **transfer RNA** – Crick's "adaptor" – was discovered 3 years after Crick's 1955 prediction of its existence. When characterized, it possessed the features Crick said the molecule would need (e.g., a hydrogen-bonding surface [the anticodon]).



CCA stem (the amino acid attachment site)

anticodon

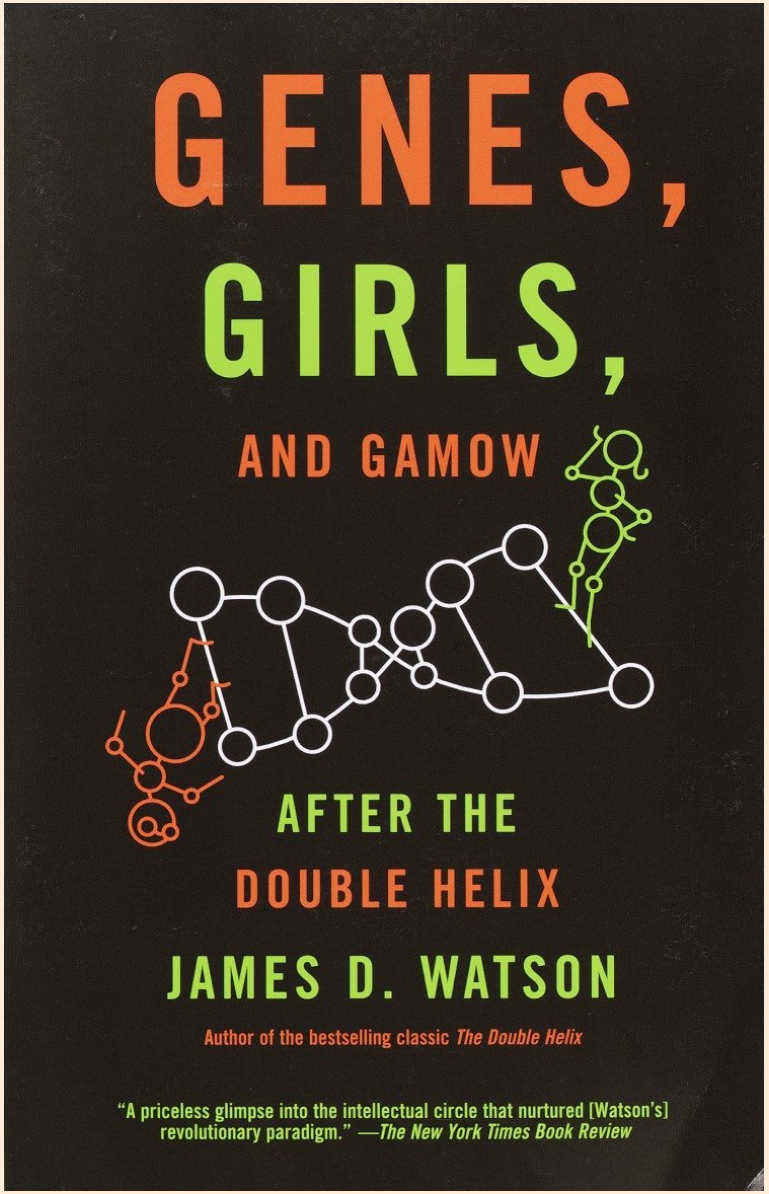
Crick also predicted 20 dedicated enzymes, to attach specific amino acids to their ”adaptors” – again, without having any direct evidence that such enzymes existed: **aminoacyl-tRNA synthetases.**



Docked structure of *E. coli* MetRS–tRNAfMet complex along with Met-AMP.
(figure from Rajendren *et al*. 2018, 402)

Now – on to the brutally honest antagonist of the story.

IV.
Why functional triangulation is really design triangulation: the causal primacy of the organism



In his 2002 autobiography, Watson describes the period in the mid-1950s, when he, Crick, and others in the "RNA Tie Club" were working out the functional implications of DNA for biological information transfer.

The pre-eminent puzzle was the nature of the genetic code.

But the code needed "hardware" – namely, molecular actors (whether proteins or nucleic acids or both) to carry out the role of actually *transferring* information from DNA to amino acids in protein assembly.

F.H.C. Crick

As a founding member of the RNA Tie Club, and one of Crick's closest collaborators, Watson was a recipient of this unpublished manuscript.

ON DEGENERATE TEMPLATES AND THE ADAPTOR HYPOTHESIS

F.H.C. Crick,

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,

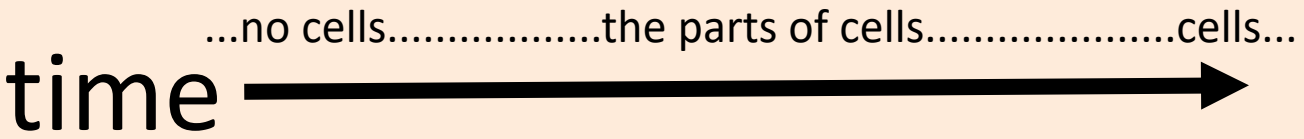
Cavendish Laboratory, Cambridge, England.

But Watson didn't buy the adaptor hypothesis. Why not?

Watson (2002, 139) explains why he didn't like the adaptor hypothesis:

"I did not like the idea at all... More to the point, the adaptor mechanism seemed to me too complicated to have ever evolved at the origin of life."

Watson's biological intuition was bound to an implicit time axis.

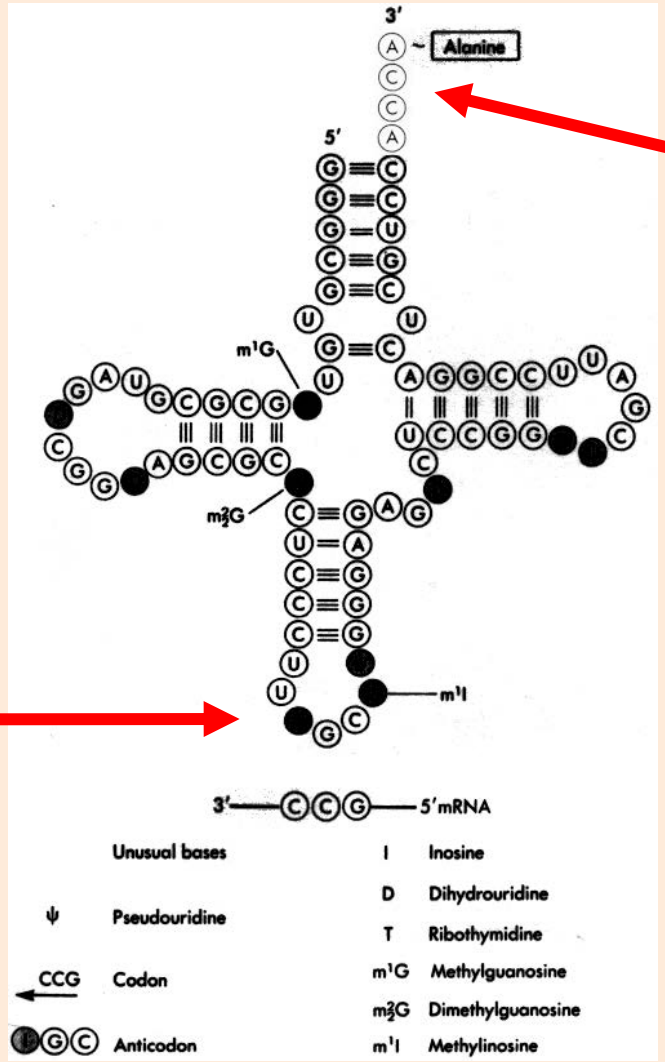


Within the naturalistic picture of life,
complex biological systems cannot arise all at once.
Organisms are fundamentally historical entities,
and “history is just one damn thing after another.”

The same is true, of course, for any naturalistic theory of abiogenesis (origin of life).

Biological complexity can only accrete over time.

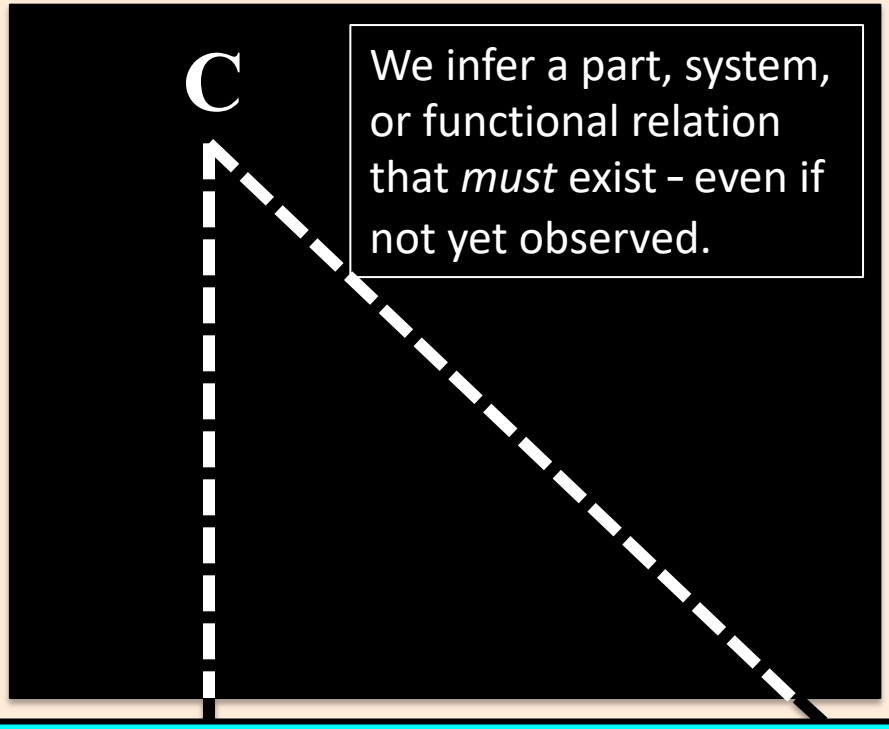
But transfer RNA, or something very much like it, ***had to be there*** – once the facts about DNA, amino acids, and protein assembly were in place.



CCA stem (the amino acid attachment site)

anticodon

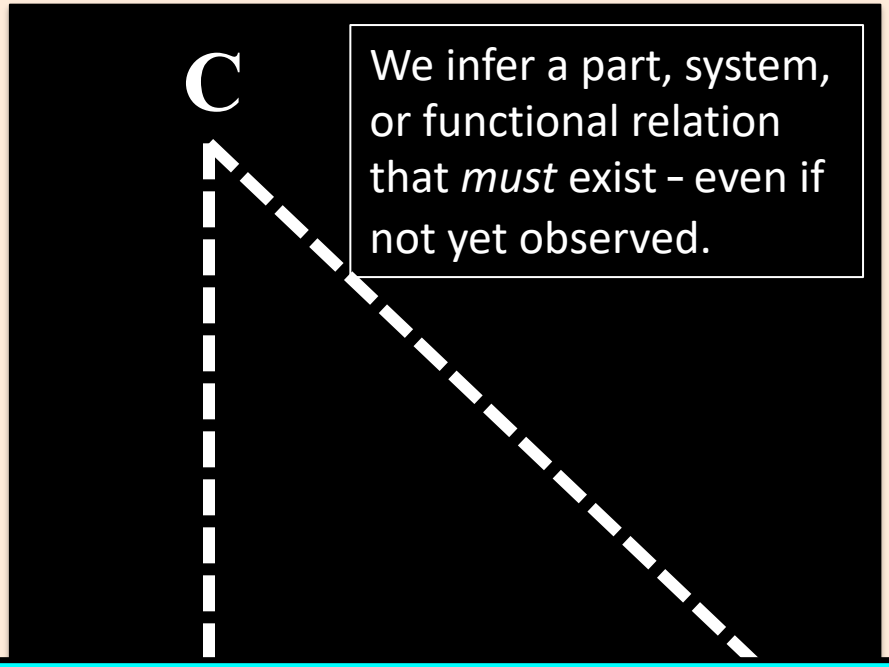
But the time axis, and the complexity-can-only-accrete assumption, aren't there *in the biology itself*.



Organisms and their features as *black boxes*

"Simplest at the start and only then more complicated," and "one must begin with the parts," flow from a *naturalistic metaphysics of explanation*. From philosophy. **Not from organisms themselves. Not from the evidence.**

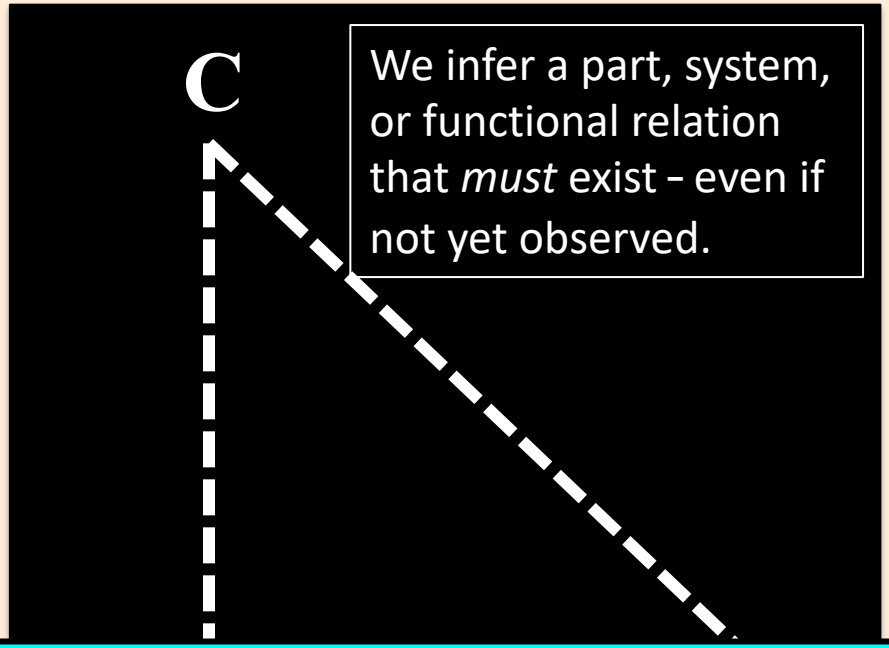
But the time axis, and the complexity-can-only-accrete assumption, aren't there *in the biology itself*.



Organisms and their features as *black boxes*

For biological inquiry – for actually finding things out – triangulation is demonstrably successful. However, the method does make a demand on the investigator.

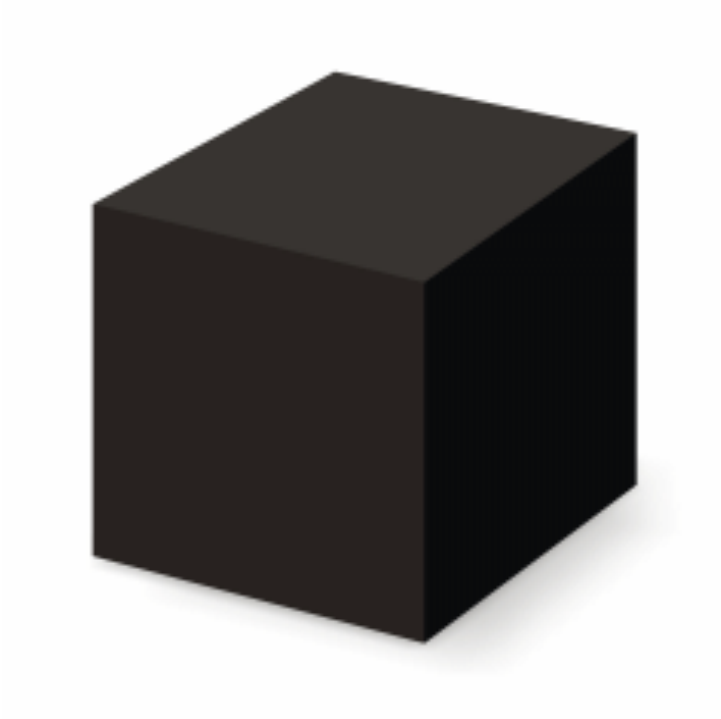
But the time axis, and the complexity-can-only-accrete assumption, aren't there *in the biology itself*.



Organisms and their features as *black boxes*

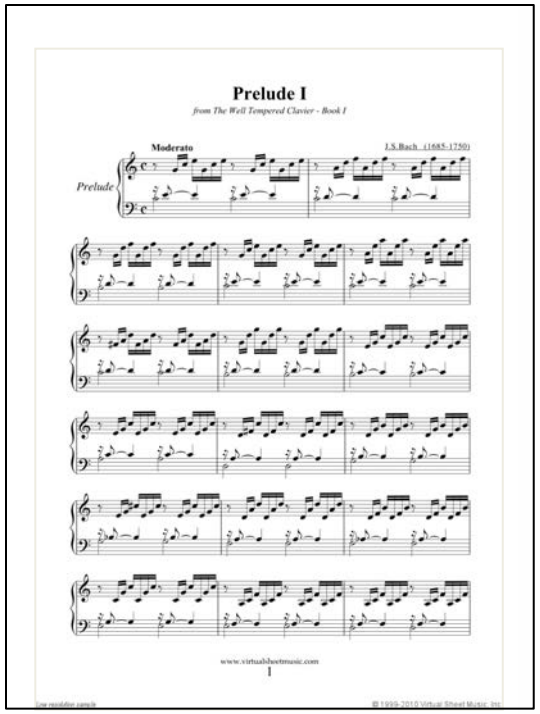
One must presuppose the prior existence of the ***system as a whole***, to obtain the functional necessity relations ***that warrant any triangulation to its unobserved parts***.

One last thought experiment
will help to illuminate
this point.

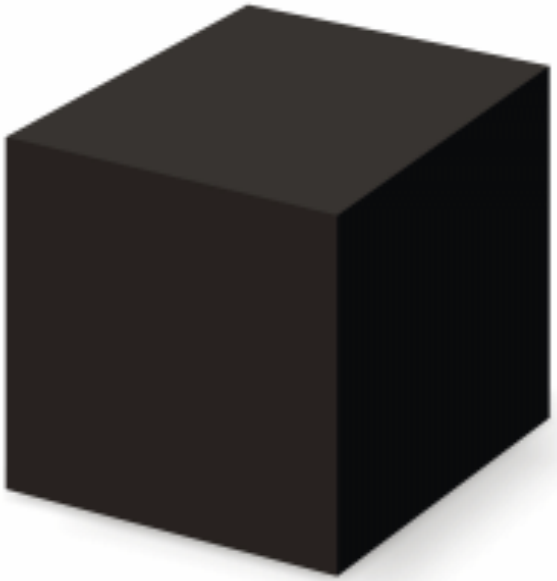


Here is a black box that a friend gave you.

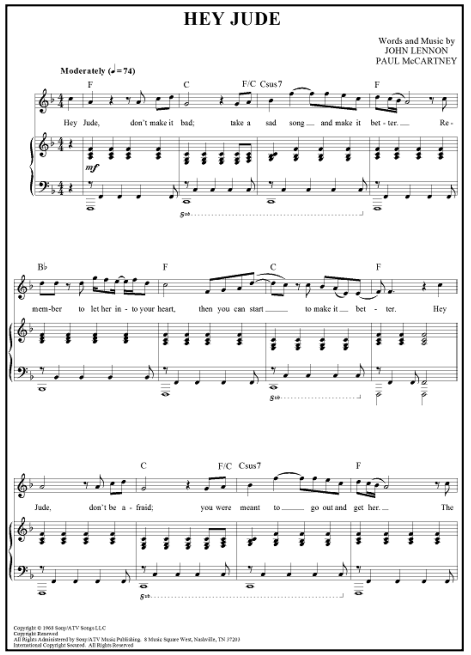
Even-numbered days



8:00 AM



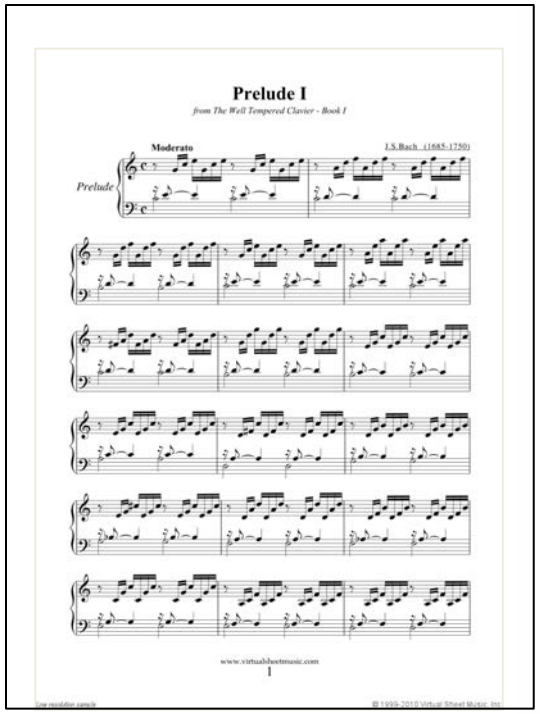
Odd-numbered days



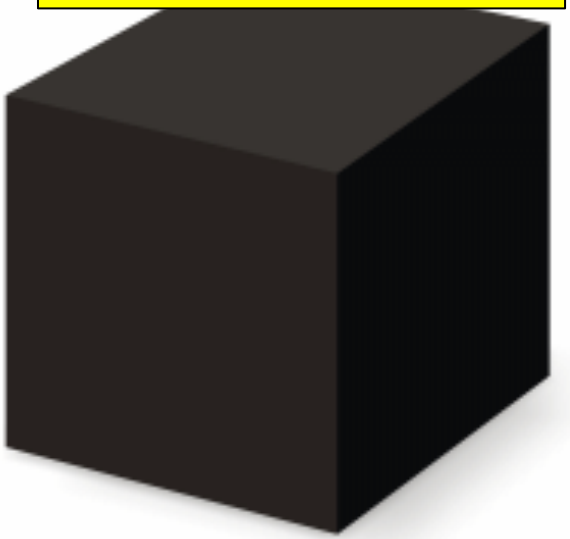
Suppose you observe the following...

On even-numbered days, at 8:00 AM, the black box plays the music of J. S. Bach. On odd-numbered days, the same box plays Beatles music. The box infallibly keeps track of even versus odd (according to the calendar), and always plays Bach, or the Beatles, depending on the value of the date.

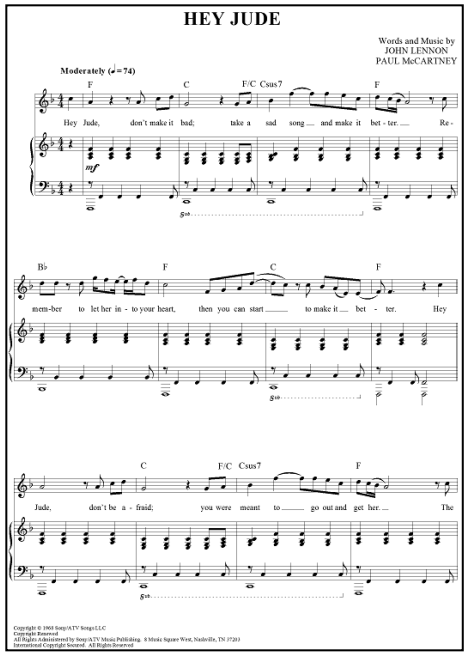
Even-numbered days



What can we then infer must exist?



Odd-numbered days



1. **An accurate clock and perpetual calendar.**
2. **Something storing the music of J.S. Bach and the Beatles.**
3. **Something scheduling the "right" music per the day & time.**
4. **A power supply, amplifier, speaker, *et cetera*.**

But pay close attention to the logical structure of the causal inference here.

What is inferentially *basic?*

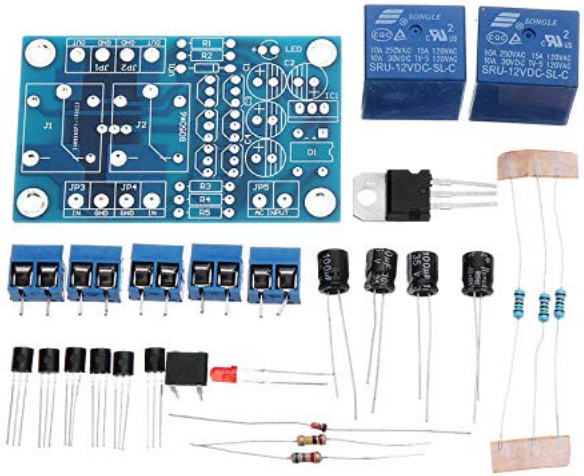
Hint: it is not the *parts* of the black box.

If we do not know the target state and its functions already, we cannot select – from the universe of all possible parts – the correct set.

This logical asymmetry decisively favors the higher level in any functional or causal analysis. **Multiple realizability** and **many-to-one relations** render it intractable to start at the lower level and infer the functions of the whole. ***The space of possibilities is too large.***

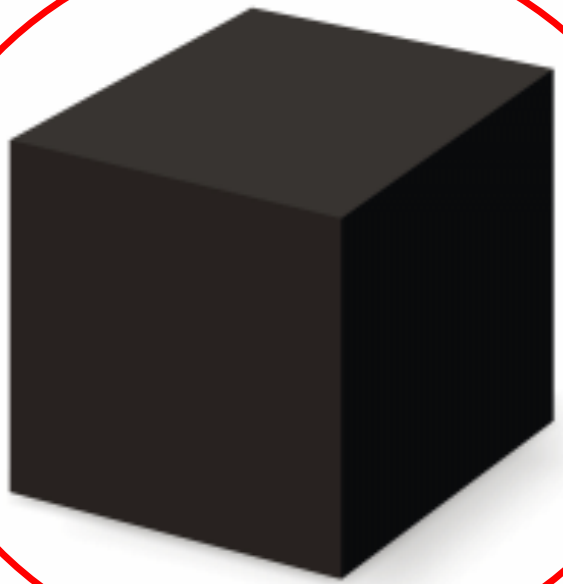
What higher-level system, with its unique functions, is entailed by these parts?

How would you know? The parts represented here (a set already drawn for the sake of illustration from the infinitude of possible entities) are consistent with an indefinitely large number of ***different*** higher-level systems.



The whole – i.e., the ***highest-level system*** – underwrites causal inferences ***to the existence and functions of its parts.***

But we cannot go in the other direction: without the functional whole to guide us, we cannot pick out (from all possibilities) the right parts or relations.



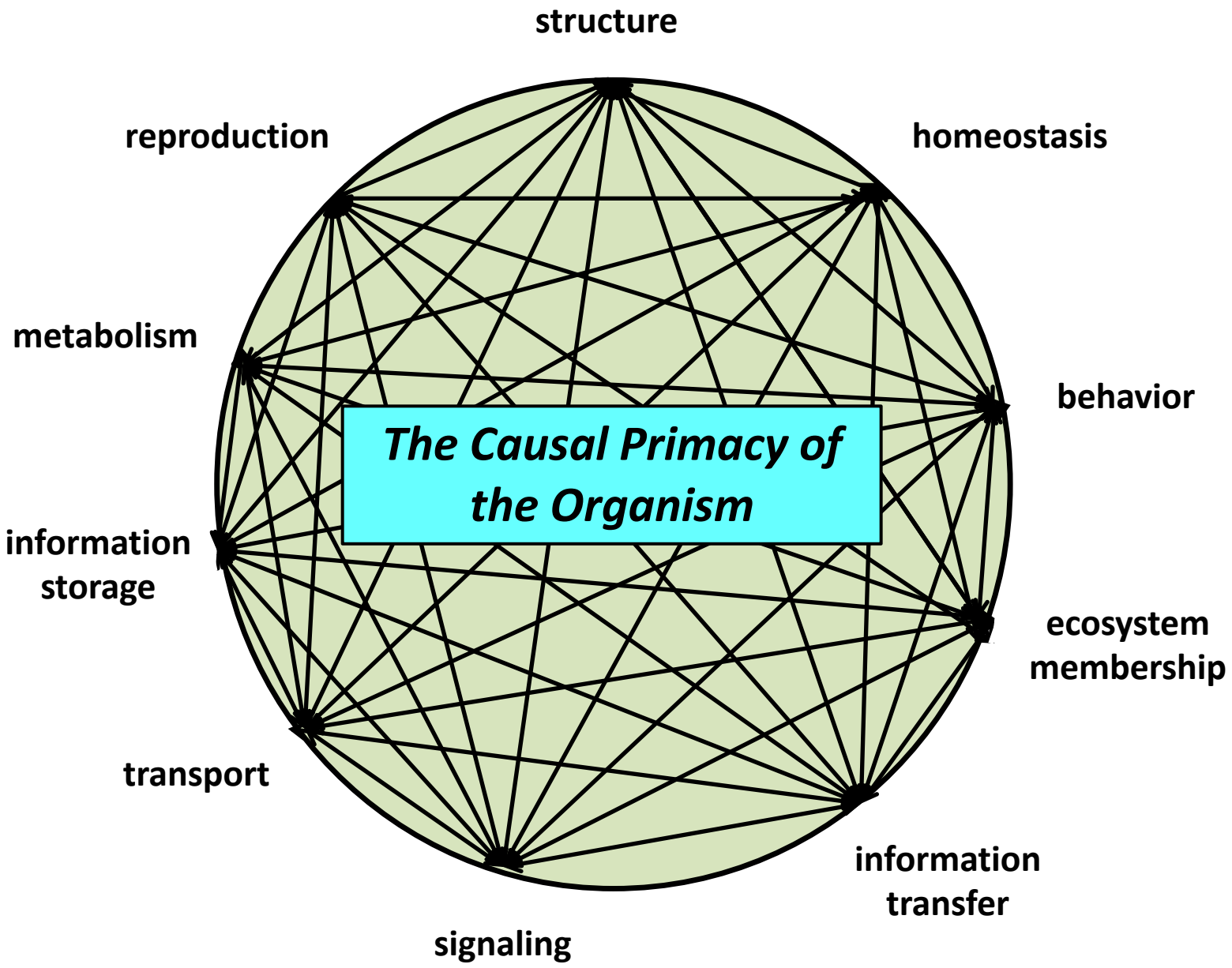
1. **An accurate clock and perpetual calendar.**
2. **Something storing the music of J.S. Bach and the Beatles.**
3. **Something playing the "right" music per the day & time.**
4. **A power supply, amplifier, speaker, *et cetera*.**

Here's that key point again:

The whole – i.e., the ***highest-level system*** – enables or underwrites causal inferences ***to the existence and functions of its parts.***

But we cannot go in the other direction**:** the parts must be selected from the universe of all possibilities – which cannot be accomplished without the prior existence of the functional whole as the target state.

Since we are talking about organisms, and not strange boxes which play Bach and the Beatles, let's call this thesis ***the causal primacy of the organism.***


structure
reproduction
homeostasis
metabolism
behavior
The Causal Primacy of the Organism
information storage
ecosystem membership
transport
information transfer
signaling

Multiple lines of evidence,
accumulated over the past 40 years,
point strongly to the causal
irreducibility of the cell,
and thus its analytical primacy.

(For details, see the big pdf next week.)

Peter Tompa
VIB, Belgium

George Rose
Johns Hopkins

The Levinthal paradox of the interactome

Peter Tompa[1]* and George D. Rose[2]

[1]VIB Department of Structural Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[2]Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland MD 21218

Received 6 September 2011; Revised 22 September 2011; Accepted 23 September 2011
DOI: 10.1002/pro.747
Published online 10 October 2011 proteinscience.org

Abstract: The central biological question of the 21st century is: how does a viable cell emerge from the bewildering combinatorial complexity of its molecular components? Here, we estimate the combinatorics of self-assembling the protein constituents of a yeast cell, a number so vast that the functional interactome could only have emerged by iterative hierarchic assembly of its component sub-assemblies. A protein can undergo both reversible denaturation and hierarchic self-assembly spontaneously, but a functioning interactome must expend energy to achieve viability. Consequently, it is implausible that a completely "denatured" cell could be reversibly renatured spontaneously, like a protein. Instead, new cells are generated by the division of pre-existing cells, an unbroken chain of renewal tracking back through contingent conditions and evolving responses to the origin of life on the prebiotic earth. We surmise that this non-deterministic temporal continuum could not be reconstructed *de novo* under present conditions.

PROTEIN SCIENCE 2011 VOL 20:2074—2079

This paper represents the shadow of a science yet to be born.

The minimal gene complement ("parts list") of *Mycoplasma genitalium* (Fraser *et al.*, 1995)

MGI#	Identification	%ID
	Amino acid biosynthesis	
	Serine family	
*394	serine hydroxymethyltransferase (glyA)	55
	Biosynthesis of cofactors, prosthetic groups, and carriers	
	Folic acid	
*013	5,10-methylene-tetrahydrofolate DHase (folD)	33
*228	dihydrofolate RDase (dhfr)	33
	Heme and porphyrin	
*255	protoporphyrinogen oxidase (hemK)	31
	Thioredoxin, glutaredoxin, and glutathione	
*124	thioredoxin (trx)	36
*109	thioredoxin RDase (trxB)	39
	Cell envelope	
	Membranes, lipoproteins, and porins	
318	fibronectin-BP (fnbA)	25
040	membrane lipoprotein (lmp0)	31
*095	prolipoprotein diacylglyceryl Tase (lgt)	29
	Surface polysaccharides, lipopolysaccharides and antigens	
137	dTDP-4-dehydrorhamnose RDase (rfbD)	32
*356	ic-1 operon pt (licA) motif	28
*060	LPS biosyn pt (rfbV) motif	36
*269	surface pt antigen precursor (pag) motif	26
025	TraB	28
	Surface structures	
192	114 kDa pt, MgPa operon (mgp)	100
191	attachment pt, MgPa operon (mgp)	100
315	cytadherence accessory pt (hmw1)	42
312	cytadherence-accessory pt (hmw1)	39
*386	cytadherence-accessory pt (hmw1)	34
313	cytadherence-accessory pt (hmw1)	53
*317	cytadherence-accessory pt (hmw3)	41
*459	surface exclusion pt (prg4) (Plasmid pCF10)	26
	Cellular processes	
	Cell division	
*467	cell division pt (ftsH)	50
*297	cell division pt (ftsY)	36
*224	cell division pt (ftsZ)	31
*434	mukB suppressor pt (smbA)	41
	Cell killing	
*146	hemolysin (tlyC)	26
220	pre-procytotoxin (vacA)	36
	Chaperones	
*019	heat shock prt (dnaJ)	34
002	heat shock prt (dnaJ) motif	40
*260	heat shock prt (dnaJ) motif	34
*392	heat shock prt (grpEL)	52
*261	heat shock prt (grpE)	32
*393	heat shock prt 60-like pt (PggroES)	40
*305	heat shock prt 70 (hsp70)	57
	Detoxification	
*066	thiophene and furan oxidizer (tdnF)	32
	Protein and peptide secretion	
*138	GTP-binding membrane prt (lepA)	48
*179	haemolysin secretion ATP-BP (hlyB) motif	35
*072	preprotein translocase (secA)	44
*172	preprotein translocase secY sub (secY)	39
210	prolipoprotein signal peptidase (lsp)	32
*048	signal recognition particle prt (ffh)	43
	Transformation	
316	competence locus E (comE3) motif	39
	Central intermediary metabolism	
	Degradation of polysaccharides	
*217	bifunctional endo-1,4-beta-xylanase xyla precursor (xynA) motif	38
	Other	
*357	acetate kinase (ackA)	43
*038	glycerol kinase (glpK)	47
293	glycerophosphoryl diester phosphodiesterase (glpQ)	39
*295	phosphotransacetylase (pta)	45
	Phosphorus compounds	
*351	inorganic pyrophosphatase (ppa)	39
	Energy metabolism	
	Aerobic	
*039	glycerol-3-phospate DHase (GUT2)	43
*460	L-lactate DHase (ldh)	50
275	NADH oxidase (nox)	39
	ATP-proton motive force interconversion	
*405	adenosinetriphosphatase (atpB)	36
*401	ATP Sase alpha chain (atpA)	63
*403	ATP Sase B chain (atpF)	37
*399	ATP Sase beta chain (atpD)	81
*404	ATP Sase C chain (atpE)	50
*402	ATP Sase delta chain (atpH)	34
*398	ATP Sase epsilon chain (atpC)	37
*406	ATP Sase gamma chain (atpG)	38
	Glycolysis	
*063	1-phosphofructokinase (fruK)	26
*215	6-phosphofructokinase (pfk)	39
*407	enolase (eno)	54
*023	fructose-bisphosphate aldolase (fba)	45
*301	GAPD (gap)	56
*111	phosphoglucose isomerase B (pgiB)	35
*306	phosphoglycerate kinase (pgk)	51
430	phosphoglycerate mutase (pgm)	45
*216	pyruvate kinase (pyk)	35
*431	triosephosphate isomerase (tim)	40
	Pentose phosphate pathway	
*264	6-phosphogluconate DHase (gnd)	38
*068	transketolase 1 (TK 1) (tktA)	33
	Pyruvate DHase	
*272	dihydrolipoamide acetyltransferase (pdhC)	45
*271	dihydrolipoamide DHase (pdhD)	39
*274	pyruvate DHase E1-alpha sub (pdhA)	43
*273	pyruvate DHase E1-beta sub (pdhB)	55
	Sugars	
*112	D-ribulose-5-phosphate 3 epimerase (cfxE)	33
*060	deoxyribose-phosphate aldolase (deoC)	83
396	galactosidase acetyltransferase (lacA)	40
*063	phosphomannomutase (cpsG)	39
	Fatty acid and phospholipid metabolism	
*212	1-acyl-sn-glycerol-3-phosphate acetyltransferase (plsC)	32
*437	CDP-diglyceride Sase (cdsA)	38
388	fatty acid/phospholipid synthesis prt (plsX)	29
085	hydroxymethylglutaryl-CoA RDase (NADPH)	23
344	lipase-esterase (lip1)	27
*116	phosphatidylglycerophosphate Sase (pgsA)	29
	Purines, pyrimidines, nucleosides, and nucleotides	
	2'-Deoxyribonucleotide metabolism	
*231	ribonucleoside-diphosphate RDase (nrdE)	54
*229	ribonucleotide RDase 2 (nrdF)	50
*227	thymidylate Sase (thyA)	37
	Nucleotide and nucleoside interconversions	
*382	uridine kinase (udk)	34
	Purine ribonucleotide biosynthesis	
*107	5'-guanylate kinase (gmk)	43
*171	adenylate kinase (adk)	32
*058	phosphoribosylpyrophosphate Sase (prs)	44
	Salvage of nucleosides and nucleotides	
*276	adenine PRTase (apt)	34
*052	cytidine deaminase (cdd)	38
*330	cytidylate kinase (cmk)	40
268	deoxyguanosine-deoxyadenosine kinase(II) sub 2	30
*458	hypoxanthine-guanine PRTase (hpt)	38
*049	purine-nucleoside phosphorylase (deoD)	44
*034	thymidine kinase (tdk)	48
051	thymidine phosphorylase (deoA)	53
*095	thymidylate kinase (CDC8)	28
030	uracil PRTase (upp)	45
	Sugar-nucleotide biosynthesis and conversions	
*118	UDP-glucose 4-epimerase (galE)	34
*453	UDP-glucose pyrophosphorylase (galU)	48
	Regulatory functions	
*024	GTP-BP (gtp1)	47
*384	GTP-BP (obg)	40
*387	GTP-BP (era)	35
246	major sigma factor (rpoD)	29
*448	pilin repressor (pilB)	53
*408	pilin repressor (pilB) motif	49
*104	virulence-associated prt homolog (vacB)	29
	Replication	
	Degradation of DNA	
032	ATP-dependent nuclease (addA)	27
	DNA replication, restriction, modification, recombination, and repair	
*463	chromosomal replication initiator prt (dnaA)	31
*004	DNA pyrase sub A (gyrA)	100
*003	DNA pyrase sub B (gyrB)	99
*244	DNA helicase II (rnaB#)	36
*254	DNA ligase (lig)	38
262	DNA polymerase I (polI) motif	39
*031	DNA polymerase III (polC)	38
*261	DNA polymerase III alpha sub (dnaE)	32
*001	DNA polymerase III beta sub (dnaN)	100
*429	DNA polymerase III sub (dnaH)	49
007	DNA polymerase III sub (dnaH) motif	23
*250	DNA primase (dnaE)	27
*010	DNA primase (dnaE) motif	26
*122	DNA topoisomerase I (topA)	39
235	endonuclease IV (nfo)	39
*421	excinuclease ABC sub A (uvrA)	48
*073	excinuclease ABC sub B (uvrB)	48
*206	excinuclease ABC sub C (uvrC)	28
*379	glucose-inhibited division prt (gidA)	49
*380	glucose-inhibited division prt (gidB)	25
*358	Holliday junction DNA helicase (ruvA)	26
*359	Holliday junction DNA helicase (ruvB)	35
164	MTase (hsdM)	43
*339	recombination prt (recA)	47
*094	replicative DNA helicase (dnaB)	33
438	restriction-modification enzyme EcoD specificity sub (hsdS)	25
*047	S-adenosylmethionine Sase 2 (metX)	44
*091	single-stranded DNA BP (ssb)	22
*204	DNA topoisomerase IV sub A (parC)	100
*203	DNA topoisomerase IV sub B (parE)	100
*097	uracil DNA glycosylase (ung)	33
	Transcription	
	Degradation of RNA	
*367	ribonuclease III (rnc)	30
*465	RNase P C5 sub (rnpA)	40
	RNA synthesis, modification, and DNA transcription	
*308	ATP-dependent RNA helicase (deaD)	23
*425	ATP-dependent RNA helicase (deaD)	32
*018	helicase (rnaT) motif	44
*141	N-utilization substance prt A (nusA)	36
*177	RNA polymerase alpha core sub (rpoA)	31
*341	RNA polymerase beta sub (rpoB)	47
*340	RNA polymerase beta' chain (rpoC)	47
022	RNA polymerase delta sub (rpoE)	29
*249	RNA polymerase sigma A factor (sigA)	44
*054	transcription antitermination factor (nusG)	31
	Translation	
	Amino acyl tRNA synthetases and tRNA modification	
*292	Ala-tRNA Sase (alaS)	34
*378	Arg-tRNA Sase (argS)	34
*113	Asn-tRNA Sase (asnS)	41
*036	Asp-tRNA Sase (aspS)	41
*253	Cys-tRNA Sase (cysS)	34
*462	Glu-tRNA Sase (gltX)	43
251	Gly-tRNA Sase	36
*035	His-tRNA Sase (hisS)	31
*345	Ile-tRNA Sase (ileS)	33
*266	Leu-tRNA Sase (leuS)	43
*136	Lys-tRNA Sase (lysS)	46
*365	Met-tRNA formyltransferase (fmt)	24
*021	Met-tRNA Sase (metS)	38
*083	peptidyl-tRNA hydrolase homolog (pth)	38
*195	Phe-tRNA Sase alpha chain (pheT)	26
*194	Phe-tRNA Sase beta chain (pheS)	35
*283	Pro-tRNA Sase (proS)	23
*182	pseudouridylate Sase I (hisT)	27
*005	Ser-tRNA Sase (serS)	43
*375	Thr-tRNA Sase (thrS)	39
*445	tRNA (guanine-N1)-MTase (trmD)	41
*126	Trp-tRNA Sase (trpS)	41
*455	Tyr-tRNA Sase (tyrS)	39
*334	Val-tRNA Sase (valS)	39
	Degradation of proteins, peptides, and glycopeptides	
*301	aminopeptidase	45
*324	aminopeptidase P (pepP)	31
*339	ATP-dependent protease (lon)	44
*365	ATP-dependent protease binding sub (clpB)	49
067	glutamic acid specific protease (SPase)	28
219	IgA1 protease	32
183	oligopeptidase F (pepF)	39
020	proline iminopeptidase (pip)	35
310	proline iminopeptidase (pip)	29
246	serine/glycine protease (gcp)	36
*238	trigger factor (tig)	25
	Protein modification and translation factors	
*089	elongation factor G (fusA)	59
026	elongation factor P (efp)	26
*433	elongation factor Tu (tsf)	59
*435	elongation factor TU (tuf)	100
*106	formylmethionine deformylase (def) motif	37
*179	initiation factor 1 (infA)	40
*172	methionine aminopeptidase (map)	36
*256	peptide chain release factor 1 (RF-1)	43
106	prt phosphatase 2C homolog (ptc1) motif	28
109	prt serine-threonine kinase motif	34
*142	prt synthesis initiation factor 2 (infB)	46
*435	ribosome releasing factor (frr)	35
*282	transcription elongation factor (greA)	40
*196	translation initiation factor IF3 (infC)	31
	Ribosomal proteins: synthesis and modification	
*082	ribosomal prt L1	48
*361	ribosomal prt L10	30
*081	ribosomal prt L11	52
*418	ribosomal prt L13	40
*161	ribosomal prt L14	62
*160	ribosomal prt L15	43
*158	ribosomal prt L16	64
*178	ribosomal prt L17	35
*187	ribosomal prt L18	43
*444	ribosomal prt L19	49
*154	ribosomal prt L2	58
*198	ribosomal prt L20	58
*232	ribosomal prt L21	38
293	ribosomal prt L21 homolog	100
*156	ribosomal prt L22	49
153	ribosomal prt L23	39
*162	ribosomal prt L24	45
*234	ribosomal prt L27	64
*426	ribosomal prt L28	36
*155	ribosomal prt L29	42
159	ribosomal prt L3	43
*257	ribosomal prt L31	37
*353	ribosomal prt L32	48
*325	ribosomal prt L33	58
*466	ribosomal prt L34	67
197	ribosomal prt L35	60
174	ribosomal prt L36	78
*152	ribosomal prt L4	39
*163	ribosomal prt L5	58
*166	ribosomal prt L6	48
*362	ribosomal prt L7/L12 ('A' type)	48
*093	ribosomal prt L9	33
*150	ribosomal prt S10	49
*176	ribosomal prt S11	48
*087	ribosomal prt S12	75
*175	ribosomal prt S13	63
*164	ribosomal prt S14	70
*424	ribosomal prt S15	48
*446	ribosomal prt S16	49
*160	ribosomal prt S17	61
*092	ribosomal prt S18	45
*155	ribosomal prt S19	59
*070	ribosomal prt S2	36
*157	ribosomal prt S3	47
*311	ribosomal prt S4	43
*169	ribosomal prt S5	56
*090	ribosomal prt S6	24
*012	ribosomal prt S6 modification prt (hisK) motif	31
*088	ribosomal prt S7	65
*165	ribosomal prt S8	47
*457	ribosomal prt S9	52
*252	rRNA methylase	39
	Transport and binding proteins	
	Amino acids, peptides, and amines	
226	aromatic amino acid transport prt (aroP)	25
*180	membrane transport prt (glnQ)	37
*303	membrane transport prt (glnQ)	32
*079	oligopeptide transport ATP-BP (amiE)	48
*080	oligopeptide transport ATP-BP (amiF)	47
*076	oligopeptide transport permease prt (dciAC)	33
*077	oligopeptide transport permease prt (oppB)	28
*042	spermidine-putrescine transport ATP-BP (potA)	42
*043	spermidine-putrescine transport permease prt (potB)	27
*044	spermidine-putrescine transport permease prt (potC)	29
	Anions	
*410	peripheral membrane prt B (pstB)	51
409	peripheral membrane prt U (phoU)	27
*451	periplasmic phosphate permease homolog (AGE81)	31
	Carbohydrates, organic alcohols, and acids	
*187	ATP-BP (msmK)	41
*062	fructose-permease IIBC component (fruA)	43
033	glycerol uptake facilitator (glpF)	36
061	hexosephosphate transport prt (uhpT)	31
*188	membrane prt (msmF)	22
189	membrane prt (msmG)	27
*119	methylgalactoside permease ATP-BP (mglA)	33
*429	PEP-dependent HPr prt kinase phosphorylatransferase (ptsI)	46
*041	phosphohistidinoprotein-hexose phosphotransferase (ptsH)	49
*089	phosphotransferase enzyme II, ABC component (ptsG)	43
129	PTS glucose-specific permease	25
*120	ribose transport permease prt (rbsC)	27
	Cations	
071	cation-transporting ATPase (pacL)	34
	Other	
*290	ATP-BP P29	32
289	high affinity transport prt P37 (P37)	36
*390	lactococcin transport ATP-BP (lcnDR3)	22
322	Na^+ ATPase sub J (ntpJ)	31
*014	transport ATP-BP (msbA)	28
*015	transport ATP-BP (msbA)	32
291	transport permease pt P69 (P69)	28
*406	transport permease pt P69 (P69) motif	40
	Other categories	
	Adaptations and atypical conditions	
454	osmotically inducible pt (osmC)	28
	phosphate limitation prt (phlX)	31
470	SpoOJ regulator motif	27
*277	spore germination apparatus pn (gerBB) motif	31
383	sporulation prt (cotB) motif	36
	Drug and analog sensitivity	
*462	high-level kasgamycin resistance (ksgA)	36
	Other	
*298	116 kD prt (p116)	33
199	29 kDa prt, MgPa operon (mgp)	62
*065	heterocyst maturation prt (devA)	38
*467	heterocyst maturation prt (devA)	40
099	hydrolase (sur2)	42
131	hypothetical prt (GB:MS1161_3)	22
*216	macrogolgin	25
*307	magnesium chelatase 30 kD sub (bchO)	27
*304	membrane-associated ATPase (cbiO)	30
364	mobilization prt (mob13) motif	31
*336	nitrogen fixation prt (nifS)	26
267	nodulation prt F (nodF)	35
098	p48 eggshell prt (p48)	23
193	PET112 prt	31
037	pre-B cell enhancing factor (PBEF)	34
298	prt L	31
*329	prt V (terV)	28
*145	prt X	29
*289	sensory rhodopsin II transducer (htrII) motif	16
*059	small prt (smpB)	33
360	UV protection prt (mucB)	22

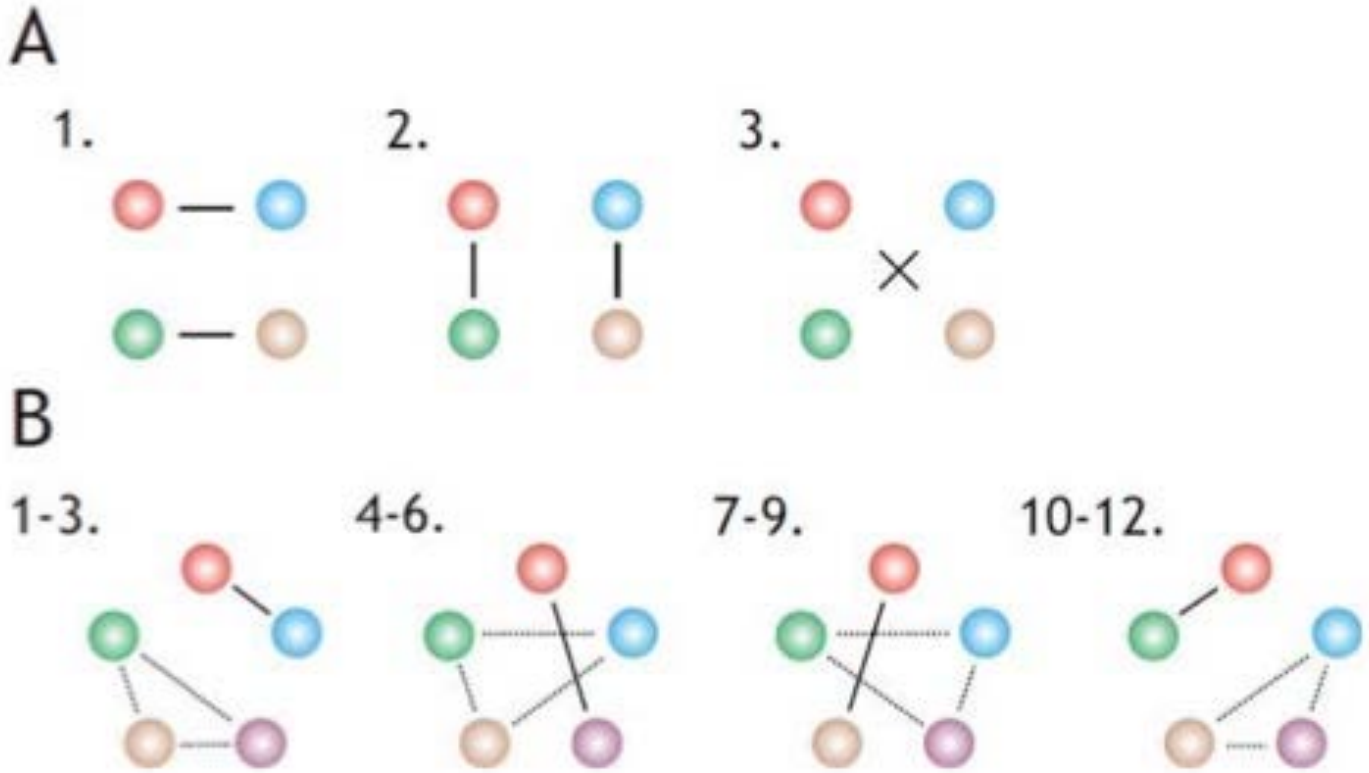


Figure 1. The number of possible interactomes increases exponentially with proteome size. The number of possible different states (patterns of pairwise interactions) of the interactome increases exponentially with the number of its constituent proteins.

Elements	Possible Pairwise Interactions
2	1
3	2
4	3
5	11
6	15
7	74
8	105
9	668
10	945
11	7,350
12	10,395
13	95,555
14	135,135
15	1,433,323
16	2,027,025
17	24,366,493
18	34,459,425

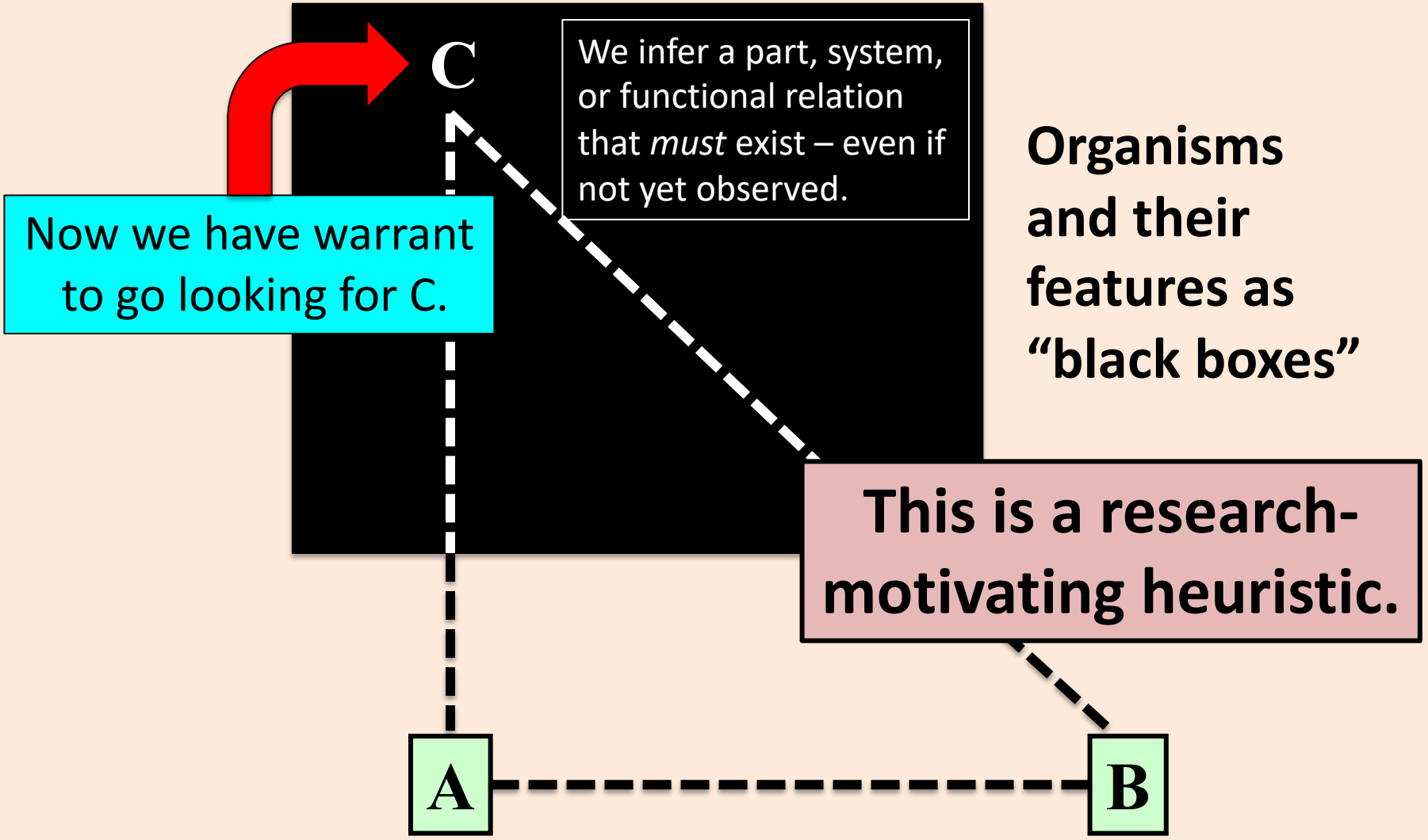
19	462,963,369
20	654,729,075
21	9,722,230,744
22	13,749,310,575
23	223,611,307,117
24	316,234,143,225
25	5,590,282,677,928
26	7,905,853,580,625
27	150,937,632,304,053
28	213,458,046,676,875
29	4,377,191,336,817,530
30	6,190,283,353,629,380
31	135,692,931,441,344,000
32	191,898,783,962,511,000
33	4,477,866,737,564,340,000
34	6,332,659,870,762,850,000
35	156,725,335,814,752,000,000
36	221,643,095,476,700,000,000
37	5,798,837,425,145,820,000,000
38	8,200,794,532,637,890,000,000
39	226,154,659,580,687,000,000,000
40	319,830,986,772,878,000,000,000
41	9,272,341,042,808,170,000,000,000
42	13,113,070,457,688,000,000,000,000
43	398,710,664,840,751,000,000,000,000
44	563,862,029,680,584,000,000,000,000
45	17,941,979,917,833,800,000,000,000,000
46	25,373,791,335,626,300,000,000,000,000
47	843,273,056,138,187,000,000,000,000,000
48	1,192,568,192,774,430,000,000,000,000,000
49	41,320,379,750,771,300,000,000,000,000,000

50	58,435,841,445,947,300,000,000,000,000,000
51	2,107,339,367,289,340,000,000,000,000,000,000
52	2,980,227,913,743,310,000,000,000,000,000,000
53	111,688,986,466,335,000,000,000,000,000,000,000
54	157,952,079,428,395,000,000,000,000,000,000,000
55	6,142,894,255,648,410,000,000,000,000,000,000,000
56	8,687,364,368,561,750,000,000,000,000,000,000,000
57	350,144,972,571,959,000,000,000,000,000,000,000,000
58	495,179,769,008,020,000,000,000,000,000,000,000,000
59	20,658,553,381,745,600,000,000,000,000,000,000,000,000
60	29,215,606,371,473,200,000,000,000,000,000,000,000,000
61	1,260,171,756,286,480,000,000,000,000,000,000,000,000,000
62	1,782,151,988,659,860,000,000,000,000,000,000,000,000,000
63	79,390,820,646,048,400,000,000,000,000,000,000,000,000,000
64	112,275,575,285,571,000,000,000,000,000,000,000,000,000,000
65	5,160,403,341,993,140,000,000,000,000,000,000,000,000,000,000
66	7,297,912,393,562,140,000,000,000,000,000,000,000,000,000,000
67	345,747,023,913,541,000,000,000,000,000,000,000,000,000,000,000
68	488,960,130,368,663,000,000,000,000,000,000,000,000,000,000,000
69	23,856,544,650,034,300,000,000,000,000,000,000,000,000,000,000
70	33,738,248,995,437,800,000,000,000,000,000,000,000,000,000,000
71	1,693,814,670,152,440,000,000,000,000,000,000,000,000,000,000,000
72	2,395,415,678,676,080,000,000,000,000,000,000,000,000,000,000,000
73	123,648,470,921,128,000,000,000,000,000,000,000,000,000,000,000,000
74	174,865,344,543,354,000,000,000,000,000,000,000,000,000,000,000,000
75	9,273,635,319,084,590,000,000,000,000,000,000,000,000,000,000,000,000
76	13,114,900,840,751,600,000,000,000,000,000,000,000,000,000,000,000,000
77	714,069,919,569,512,000,000,000,000,000,000,000,000,000,000,000,000,000
78	1,009,847,364,737,870,000,000,000,000,000,000,000,000,000,000,000,000,000
79	56,411,523,645,991,500,000,000,000,000,000,000,000,000,000,000,000,000,000
80	79,777,941,814,291,600,000,000,000,000,000,000,000,000,000,000,000,000,000

81	4,569,333,415,325,320,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
82	6,462,013,286,957,630,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
83	379,254,673,472,001,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
84	536,347,102,817,483,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
85	32,236,647,245,120,100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
86	45,589,503,739,486,100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
87	2,804,588,310,325,440,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
88	3,966,286,825,335,280,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
89	249,608,359,618,965,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
90	352,999,527,454,840,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
91	22,714,360,725,325,800,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
92	32,122,956,998,390,500,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
93	2,112,435,547,455,290,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
94	2,987,435,000,850,310,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
95	200,681,377,008,254,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
96	283,806,325,080,780,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
97	19,466,093,569,800,600,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
98	27,529,213,532,835,600,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
99	1,927,143,263,410,250,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
100	2,725,392,139,750,730,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000

Bill Dembski stopped his calculations here,
at 100 proteins, even though real cells have at least
300 proteins. There was no point in going any further.

What metaphysics of explanation enables one to start with the whole organism?



Design.