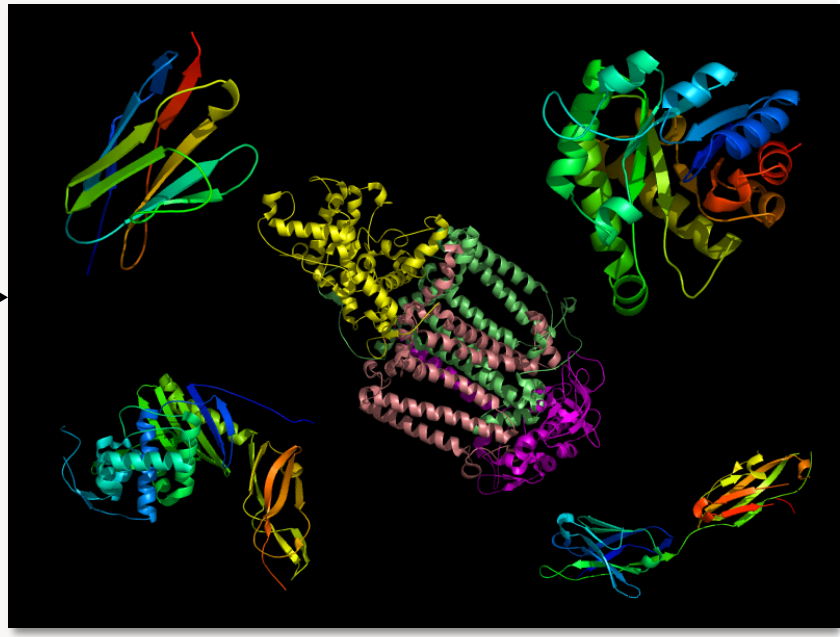
The Origin of Enzymes: All in the Family?

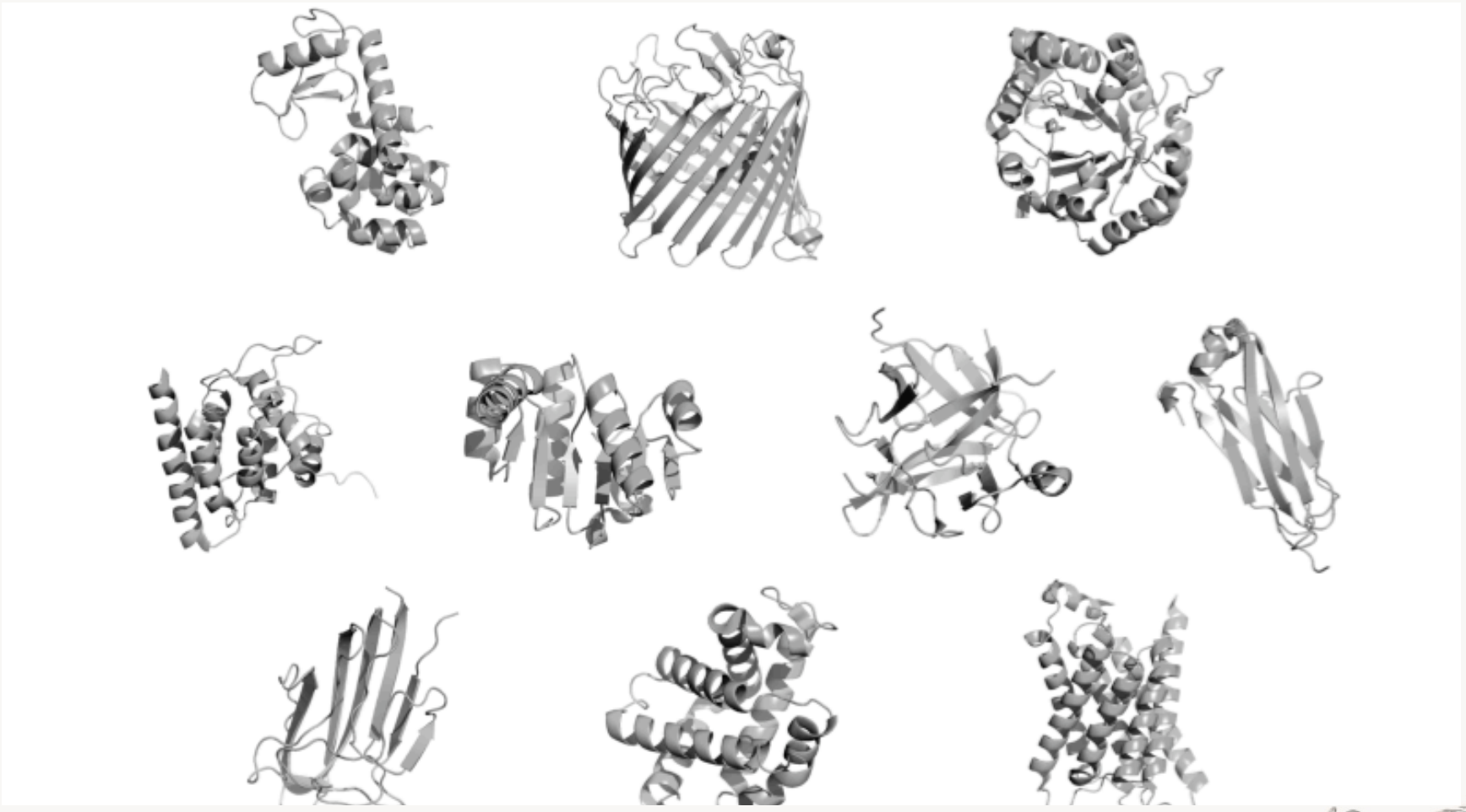
Ann Gauger

Biologic Institute

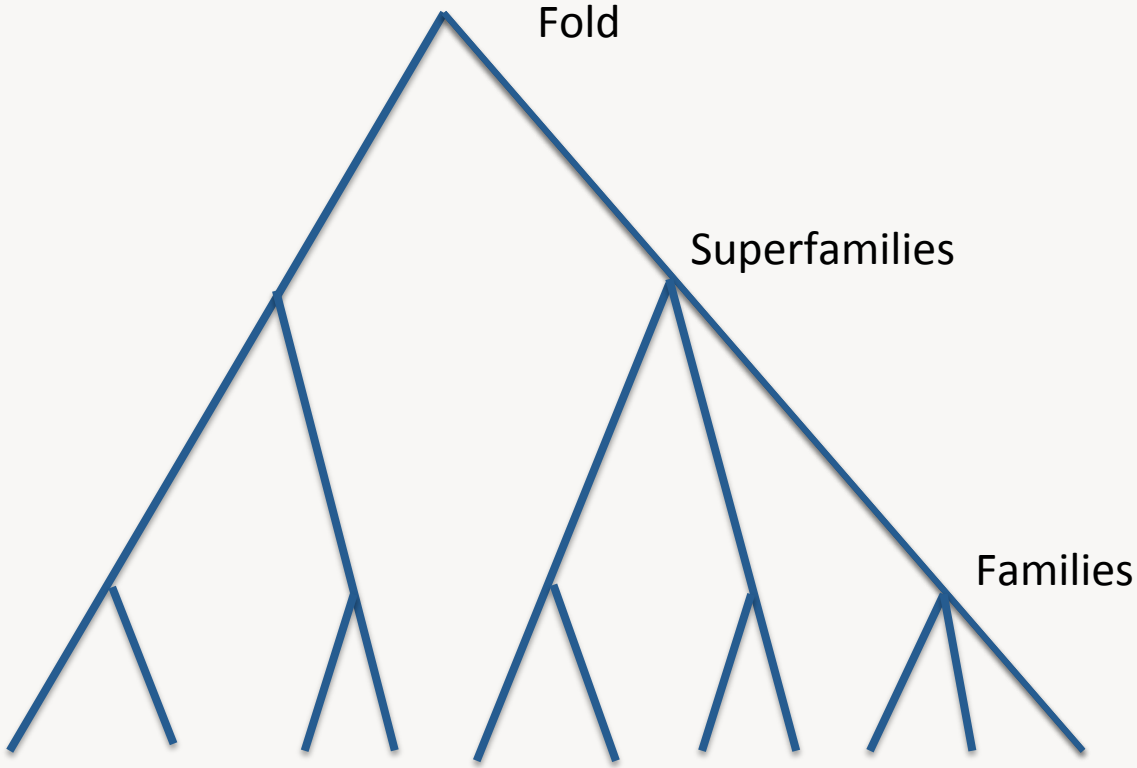
The image cannot be displayed. Your computer may not have enough memory to open the image, or the image may have been corrupted. Restart your computer, and then open the file again. If the red x still appears, you may have to delete the image and then insert it again.

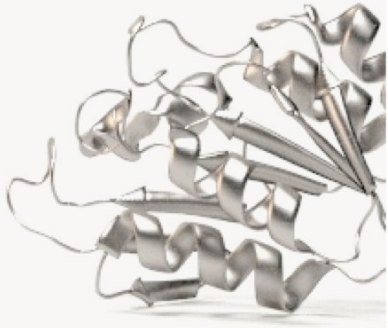


biologic INSTITUTE

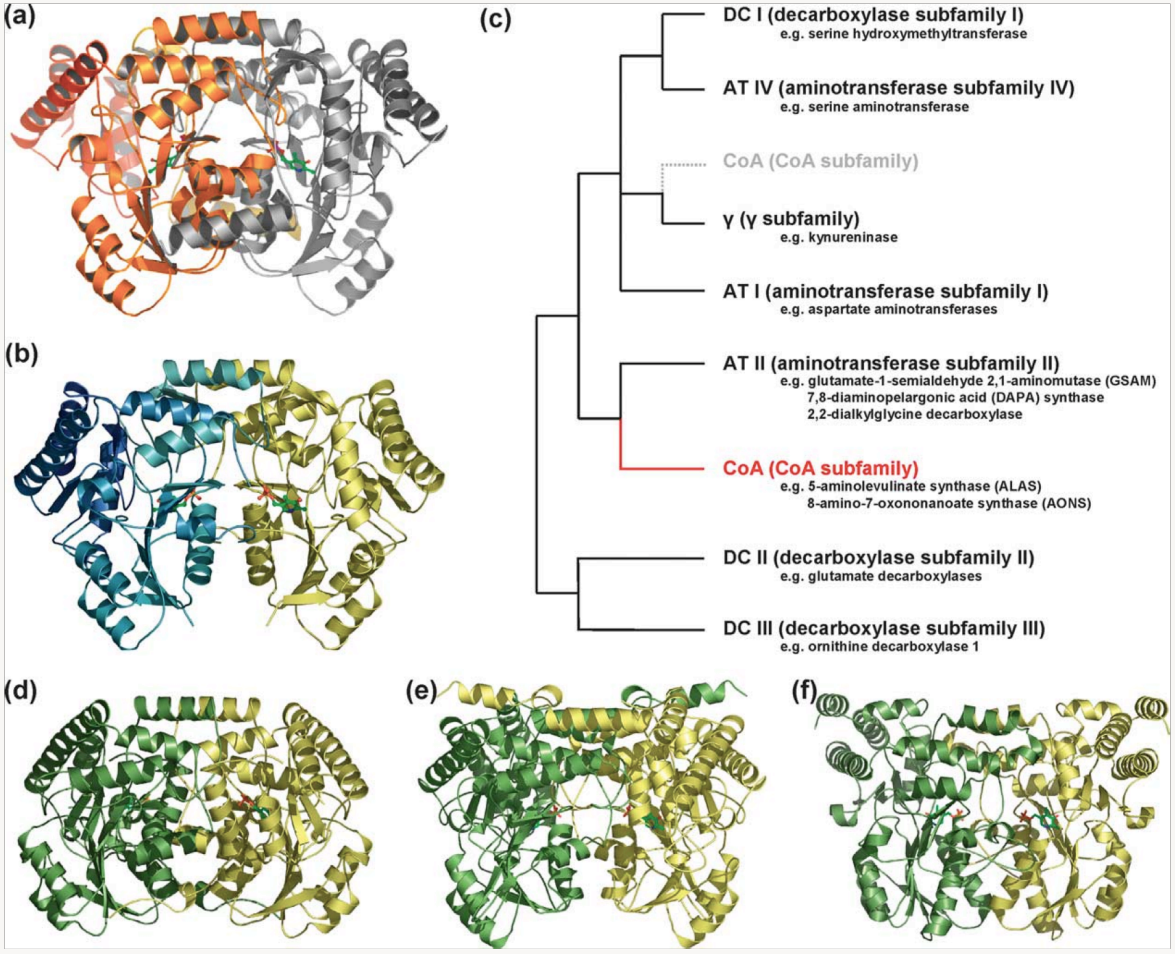


biologic INSTITUTE


Fold
Superfamilies
Families
Individual Proteins



PLP-dependent Families:

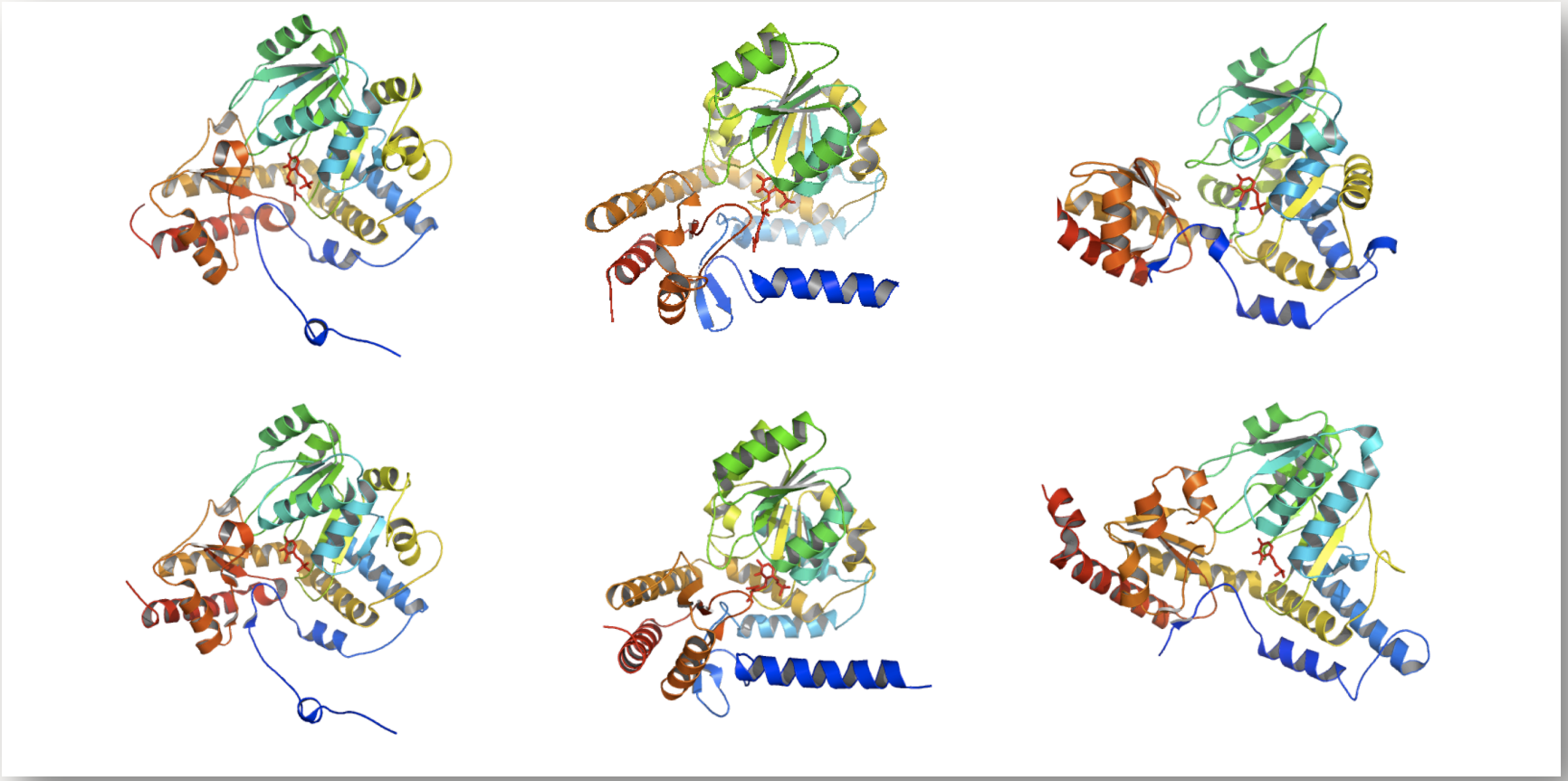


From Schulze et al. (2006), J. Mol. Biol. 358, 1212.

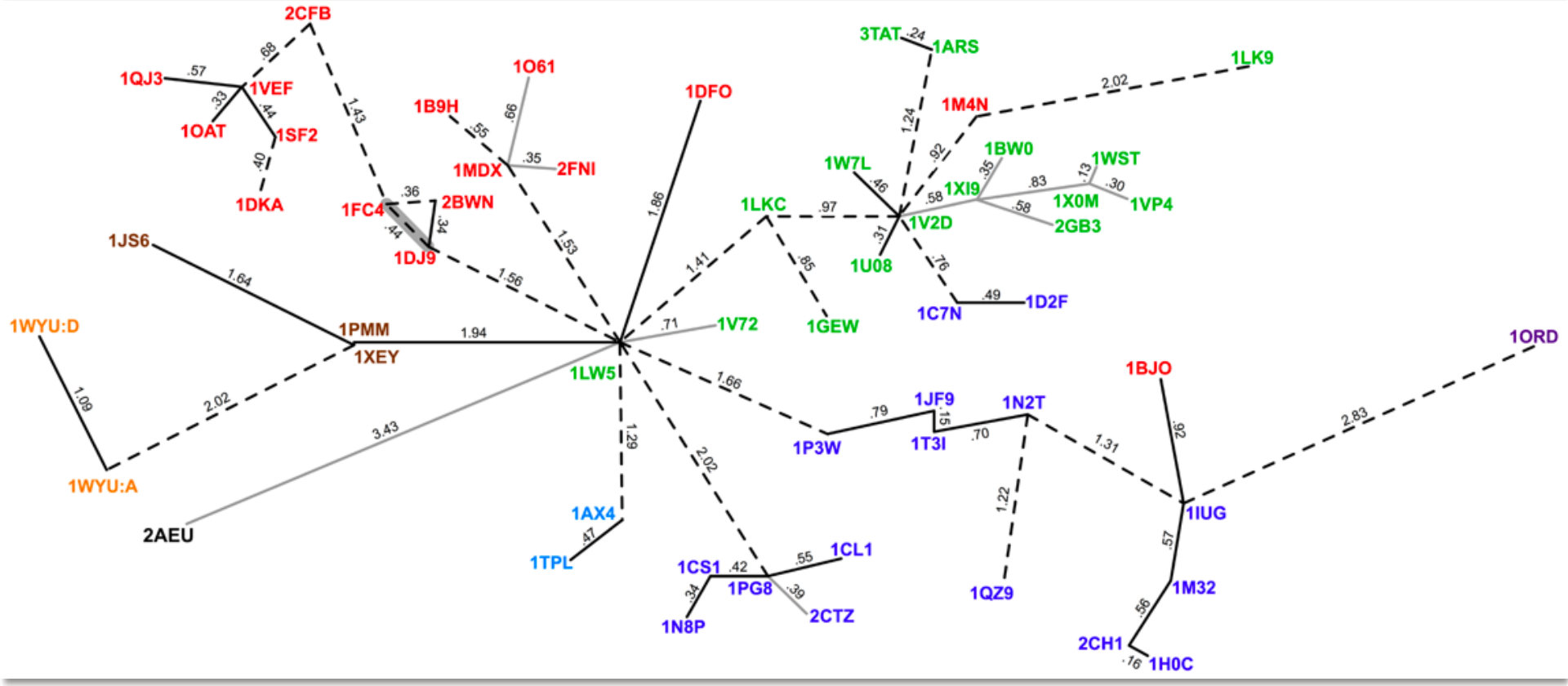
Similarity of form is assumed
to indicate common descent.

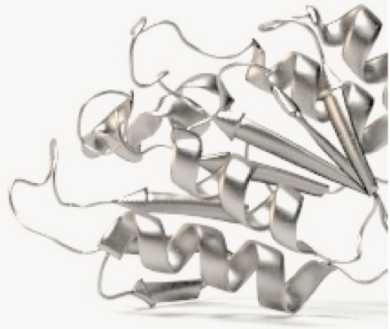
But is the proposed mechanism sufficient?

We propose to test it.



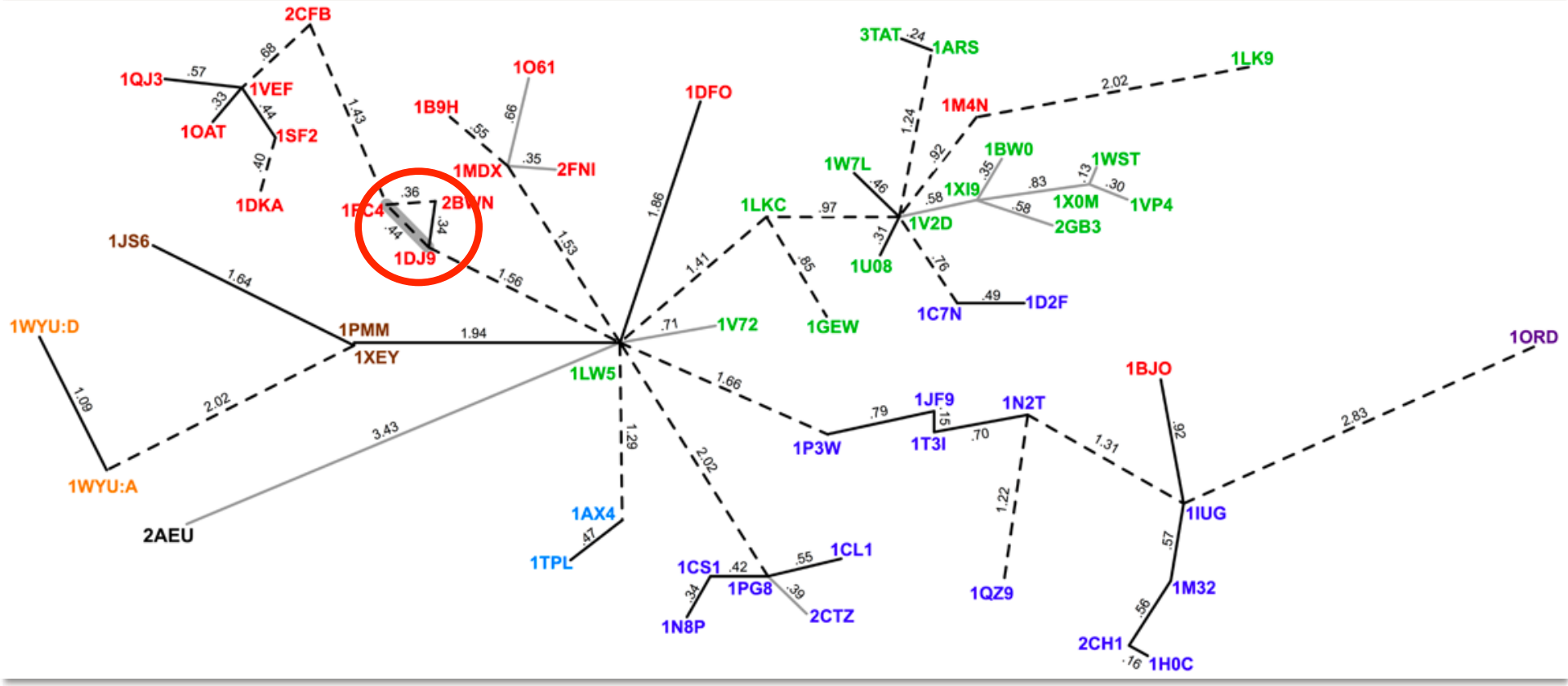
biologic INSTITUTE

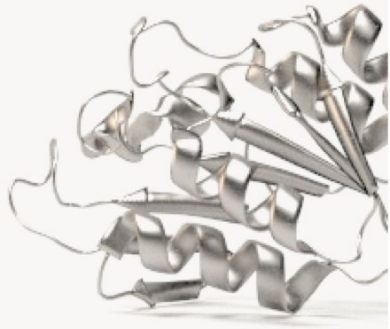

2CFB
1QJ3
1VEF
1OAT
1SF2
1DKA
1FC4
2BWN
1DJ9
1JS6
1PMM
1XEY
1WYU:D
1WYU:A
2AEU
1B9H
1O61
1MDX
2FNI
1DFO
1LW5
1V72
1LKC
1GEW
3TAT
1ARS
1W7L
1M4N
1LK9
1BW0
1WST
1XI9
1X0M
1VP4
1V2D
2GB3
1U08
1C7N
1D2F
1P3W
1JF9
1T3I
1N2T
1ORD
1BJO
1IUG
1M32
2CH1
1H0C
1AX4
1TPL
1CS1
1PG8
1CL1
2CTZ
1N8P
1QZ9

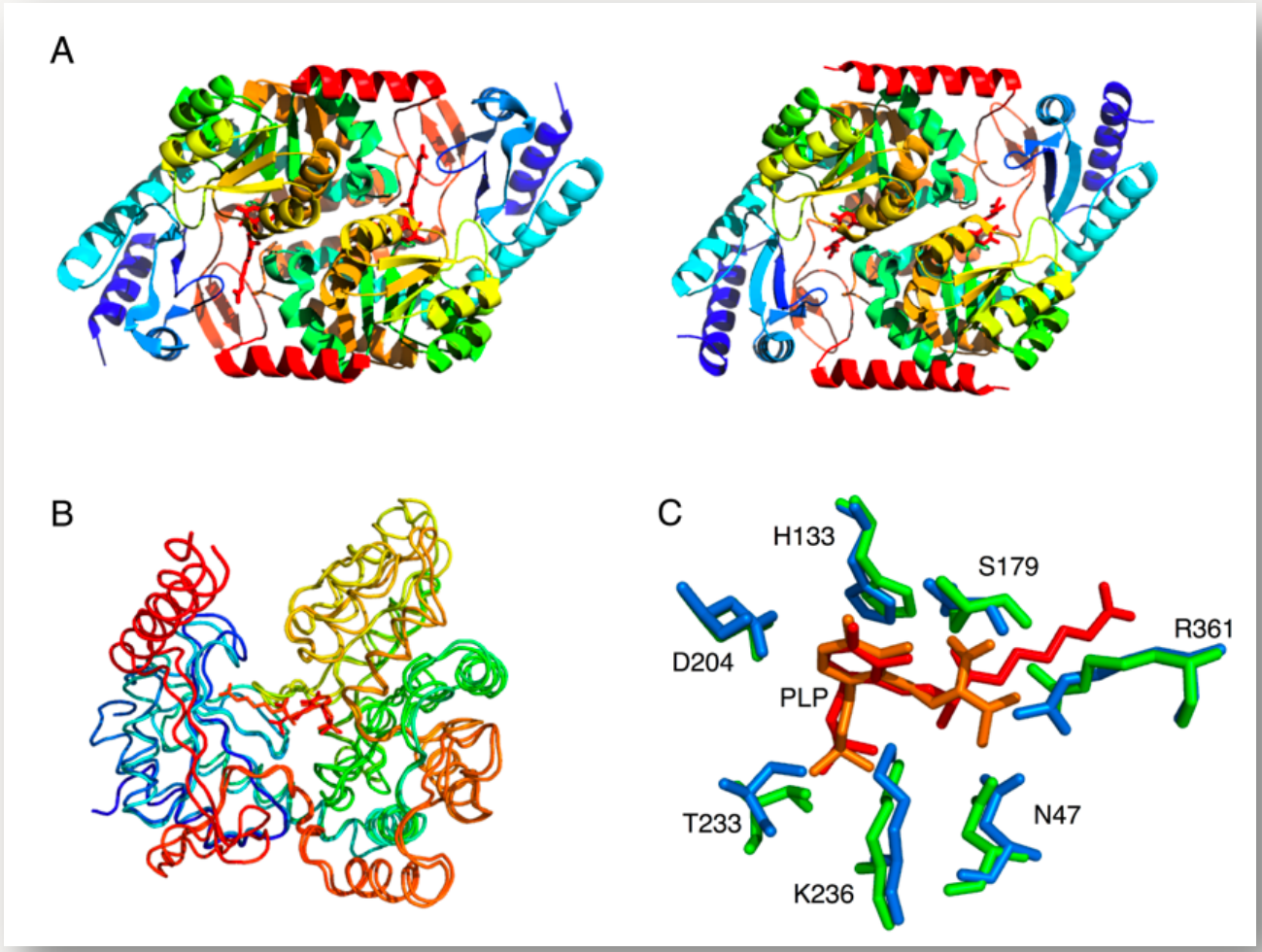


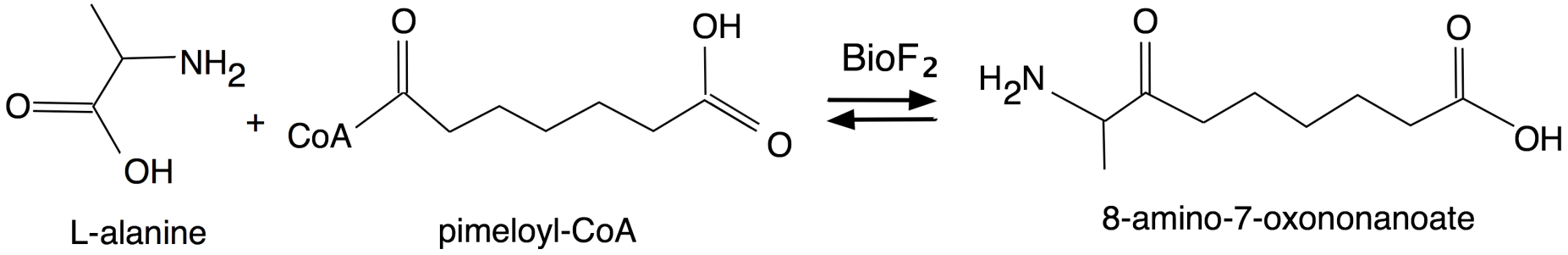
Candidate pair with

- Close structural similarity
- In E. coli
- Distinct chemistry


2CFB
1QJ3
1VEF
1OAT
1SF2
1DKA
1B9H
1O61
1MDX
2FNI
1PC4
2BWN
1DJ9
1JS6
1PMM
1XEY
1WYU:D
1WYU:A
2AEU
1LW5
1DFO
1LKC
1V72
1GEW
3TAT
1ARS
1LK9
1M4N
1W7L
1BW0
1WST
1XI9
1X0M
1VP4
2GB3
1V2D
1U08
1C7N
1D2F
1ORD
1BJO
1JF9
1N2T
1P3W
1T3I
1IUG
1M32
2CH1
1H0C
1AX4
1TPL
1CS1
1PG8
1CL1
2CTZ
1N8P
1QZ9



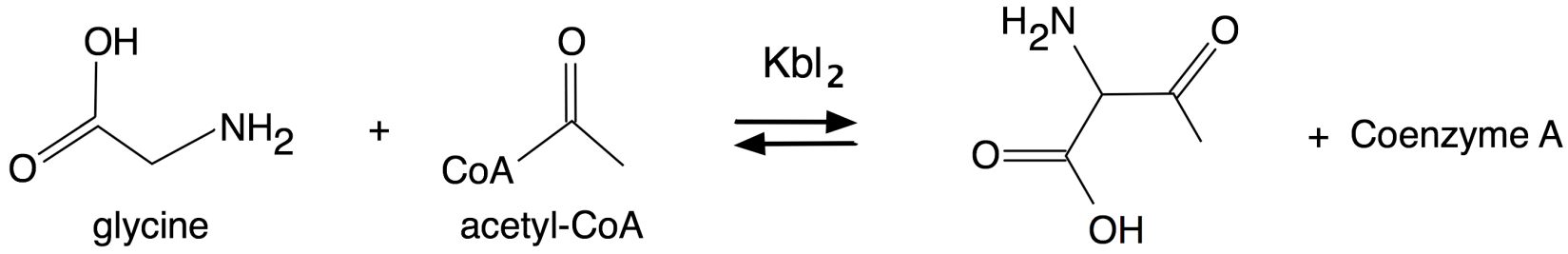

A
B
C
H133
S179
R361
D204
PLP
T233
N47
K236



L-alanine pimeloyl-CoA

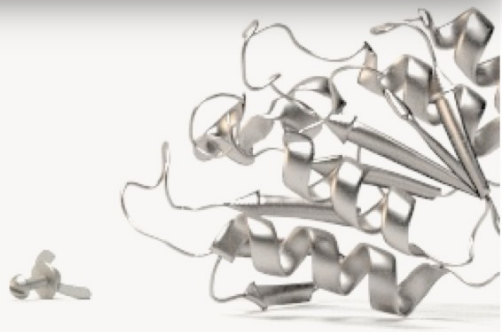
8-amino-7-oxononanoate

+ CO_2 + Coenzyme A



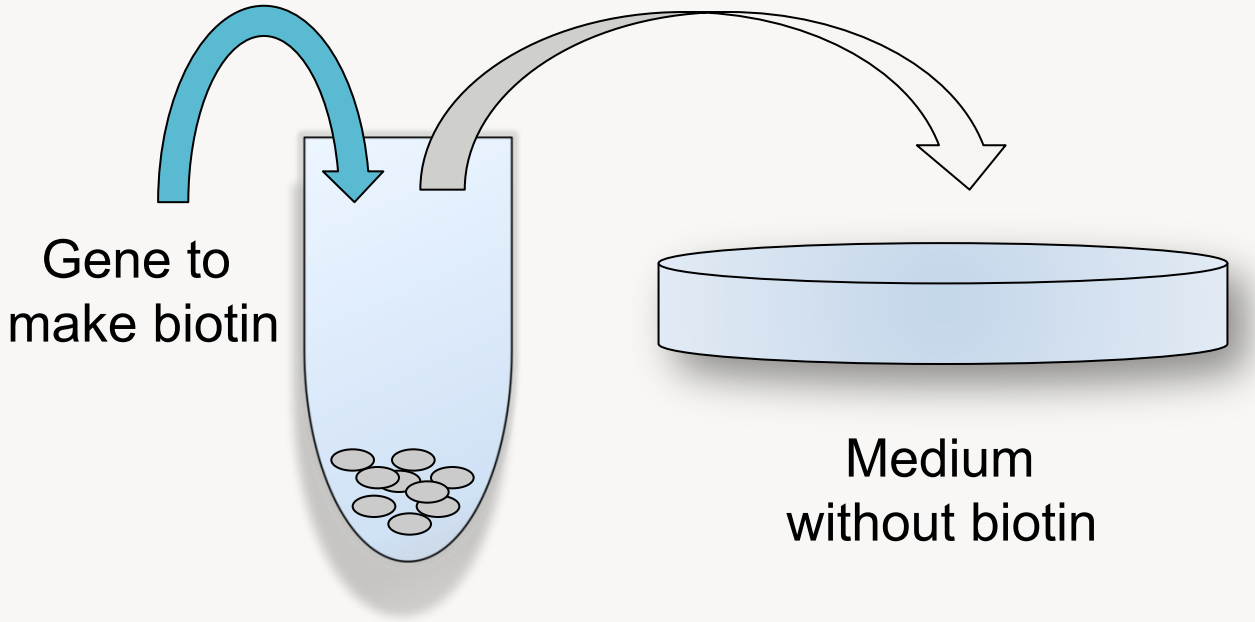
glycine acetyl-CoA

2-amino-3-ketobutyrate



- Identify what positions in BioF are essential for BioF activity
- Convert Kbl to resemble BioF at those loci

Functional test



biologic INSTITUTE

```
BioF     1 M--SWQEKINAALDARRAADALRRRYPVAQGAGRWL-VADDRQYLNFSSN    47
           |           |   ||                     |||     ||  |
Kbl      1 MRGEFYQQLTNDLETARAEGLFKEERIITSAQQADITVADGSHVINFCAN    50

BioF    48 DYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWL    97
            ||||  ||  | |   |    | |       |    |  ||  ||  |
Kbl     51 NYLGLANHPDLIAAAKAGMDSHGFGMASVRFICGTQDSHKELEQKLAAFL   100

BioF    98 GYSRALLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQL   147
           |   | |  | | ||          || |  | | |||      |
Kbl    101 GMEDAILYSSCFDANGGLFETLLGAEDAIISDALNHASIIDGVRLCKAKR   150

BioF   148 RRFAHNDVTHL-ARLLASPCPGQQ--MVVTEGVFSMDGDSAPLAEIQQVT   194
           |  |   |   | |||      |       | ||||||  | |
Kbl    151 YRYANNDMQELEARLKEAREAGARHVLIATDGVFSMDGVIANLKGVCDLA   200

BioF   195 QQHNGWLMVDDAHGTGVIGEQGRGSCWLQKV--KPELLVVTFGKGFG-VS   241
              |||| |  |  || ||||     |         | ||  |  |
Kbl    201 DKYDALVMVDDSHAVGFVGENGRGSHEYCDVMGRVDIITGTLGKALGGAS   250

BioF   242 GAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQALRASLAVIRSDE-GDA   290
           |       |   | |  |  | |  ||      ||  |    | |
Kbl    251 GGYTAARKEVVEWLRQRSRPYLFSNSLAPA---IVAASIKVLEMVEAGSE   297

BioF   291 RREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQLAEKLRQ   340
           |  | |    ||       ||||    || |   ||   |   |  |
Kbl    298 LRDRLWANARQFREQMSAAGFTLAGADHAIIPVMLGDAVVAQKFARELQK   347

BioF   341 QGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG------   384
           | ||  |  |||| | || |    |||    | |  |     |
Kbl    348 EGIYVTGFFYPVVPKGQARIRTQMSAAHTPEQITRAVEAFTRIGKQLGVI

BioF   384 -    384

Kbl    398 A    398
```

Results

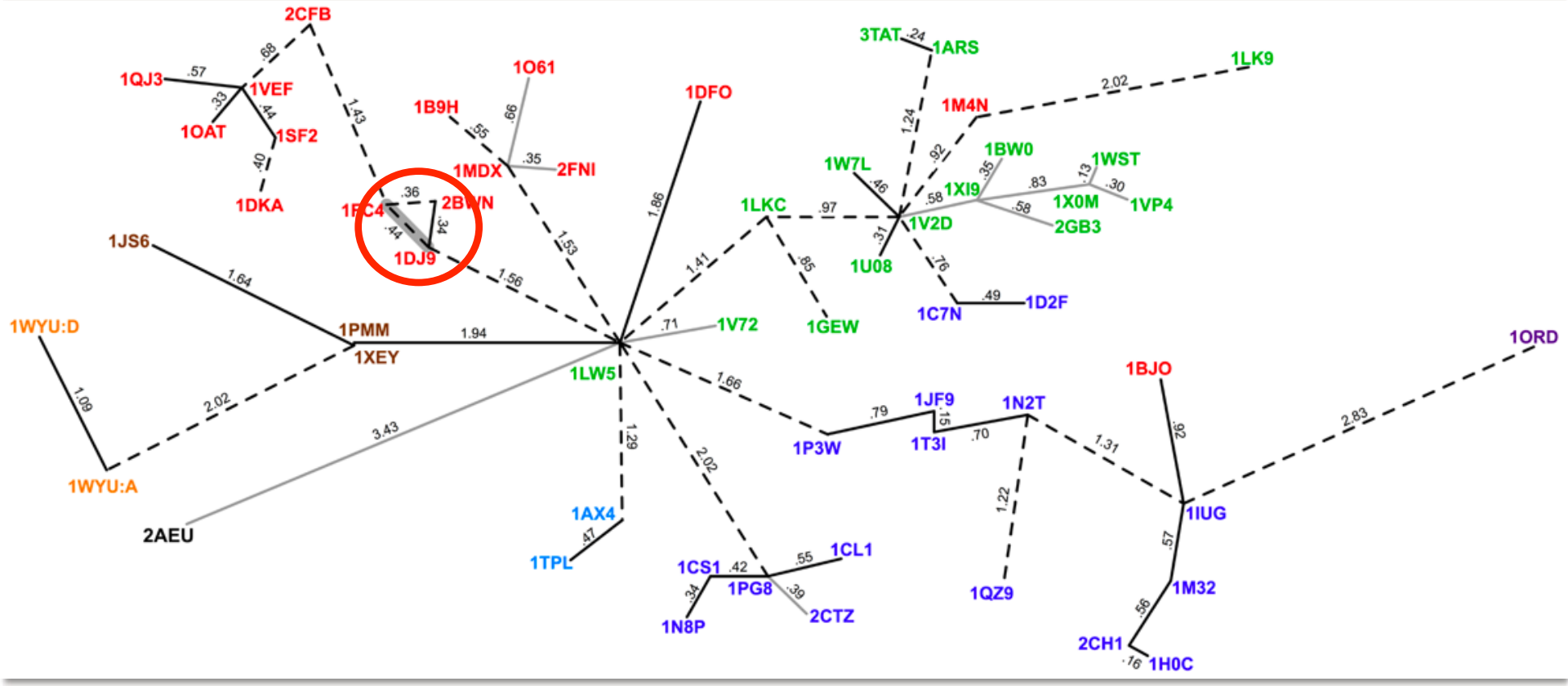
Critical loci changed	Able to make biotin?
N155H	No
Group 1	No
Group 2	No
Group 3	No
All combined	No
All plus random mutagenesis	No

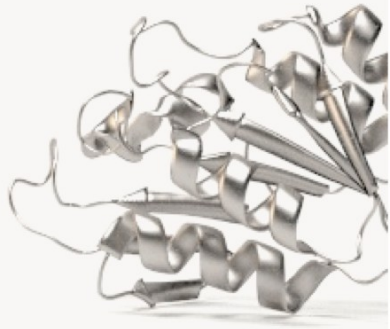
20 changes

Minimum estimate of mutations required

7 genetic events

biologic INSTITUTE


2CFB
1QJ3
1VEF
1OAT
1SF2
1DKA
1B9H
1O61
1MDX
2FNI
1PC4
2BWN
1DJ9
1JS6
1PMM
1XEY
1WYU:D
1WYU:A
2AEU
1LW5
1DFO
1LKC
1V72
1GEW
3TAT
1ARS
1LK9
1M4N
1W7L
1BW0
1WST
1XI9
1X0M
1VP4
2GB3
1V2D
1U08
1C7N
1D2F
1ORD
1BJO
1JF9
1N2T
1P3W
1T3I
1IUG
1M32
2CH1
1H0C
1AX4
1TPL
1CS1
1PG8
1CL1
2CTZ
1N8P
1QZ9



OPEN ACCESS

BIO-complexity

Research Article

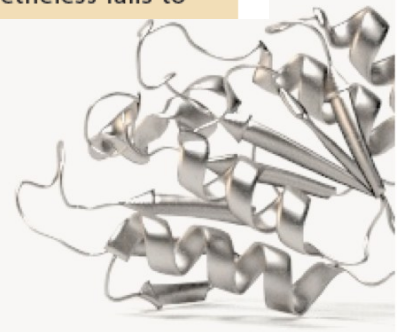
The Evolutionary Accessibility of New Enzyme Functions: A Case Study from the Biotin Pathway

Ann K. Gauger and Douglas D. Axe*

Biologic Institute, Redmond, WA, USA

Abstract

Enzymes group naturally into families according to similarity of sequence, structure, and underlying mechanism. Enzymes belonging to the same family are considered to be *homologs*—the products of evolutionary divergence, whereby the first family member provided a starting point for conversions to new but related functions. In fact, despite their similarities, these families can include remarkable functional diversity. Here we focus not on minor functional variations within families, but rather on *innovations*—transitions to genuinely new catalytic functions. Prior experimental attempts to reproduce such transitions have typically found that many mutational changes are needed to achieve even weak functional conversion, which raises the question of their evolutionary feasibility. To further investigate this, we examined the members of a large enzyme superfamily, the PLP-dependent transferases, to find a pair with distinct reaction chemistries and high structural similarity. We then set out to convert one of these enzymes, 2-amino-3-ketobutyrate CoA ligase (Kbl_2), to perform the metabolic function of the other, 8-amino-7-oxononanoate synthase ($BioF_2$). After identifying and testing 29 amino-acid changes, we found three groups of active-site positions and one single position where Kbl_2 side chains are incompatible with $BioF_2$ function. Converting these side chains in Kbl_2 makes the residues in the active-site cavity identical to those of $BioF_2$, but nonetheless fails to



A growing realization:

- Interchanging reactions catalyzed by members of mechanistically diverse superfamilies might be envisioned as "easy" exercises in (re)design: *if Nature did it, why can't we?* ....Anecdotally, many attempts at interchanging activities in mechanistically diverse superfamilies have since been attempted, but few successes have been realized.

Gerlt and Babbitt (2009) Enzyme (Re)Design: Lessons from Natural Evolution and Computation. Curr Opin Chem Biol 13:10-18.

A growing realization:

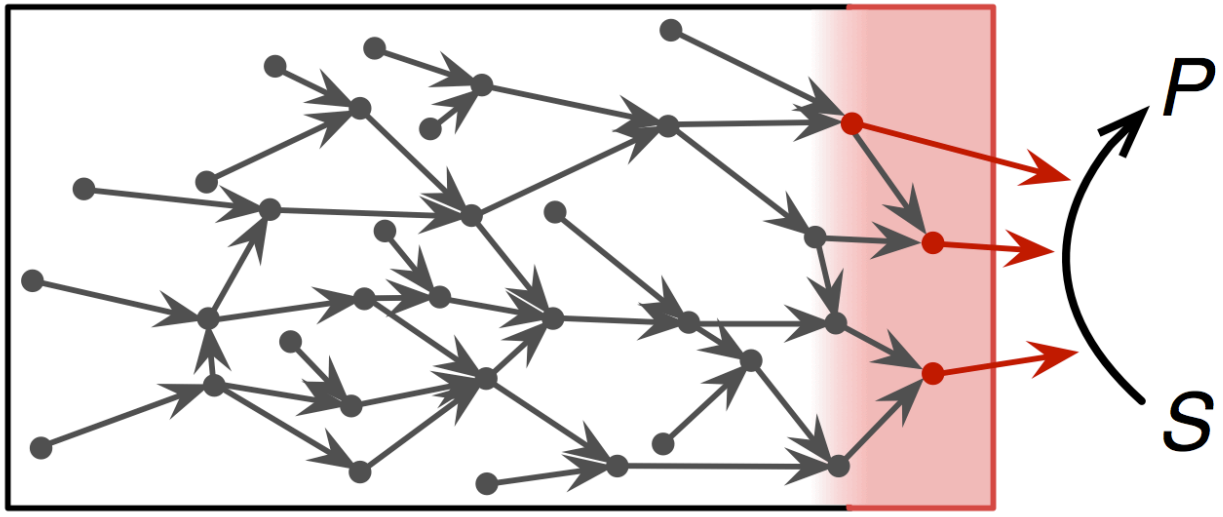
Some functions...simply can not be reached through a series of small uphill steps and instead require longer jumps that include mutations that would be neutral or even deleterious when made individually. Examples of functions that might require multiple simultaneous mutations include the appearance of *a new catalytic activity or an activity for which the parent and its single mutants show no measurable activity*.

Romero and Arnold (2009) Exploring Protein Fitness Landscapes by Directed Evolution. Nature Reviews: Molecular Cell Biology 10:866-875.

biologic INSTITUTE

Why so hard?

Epistasis and long-range interactions



biologic INSTITUTE

Proteins are global interactive designs, not static structures built of interchangeable parts.

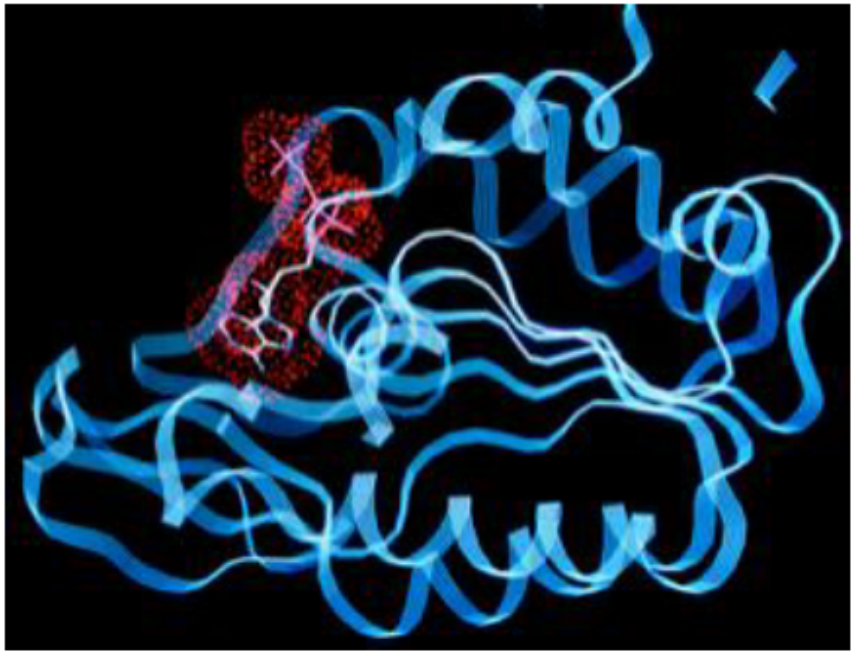
You can't put a new active site onto an old scaffold and expect it to work.

From New Scientist this week:

Designer enzymes: Learning to rival nature's wizardry

16 July 2012 by **James Mitchell Crow**
Magazine issue 2873. **Subscribe and save**

Biologically inspired catalysts could usher in a revolution in chemistry – if only we can figure out the best way to create them



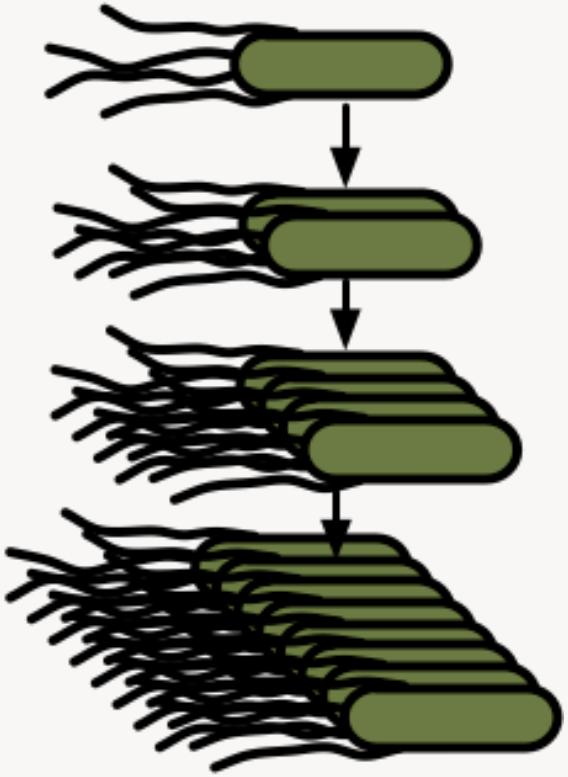
An enzyme, nature's own catalyst *(Image: Oxford Molecular Biophysics laboratory/Science Photo Library)*

So how feasible is an unguided neo-Darwinian search
for seven mutations that work together
to make a new function.

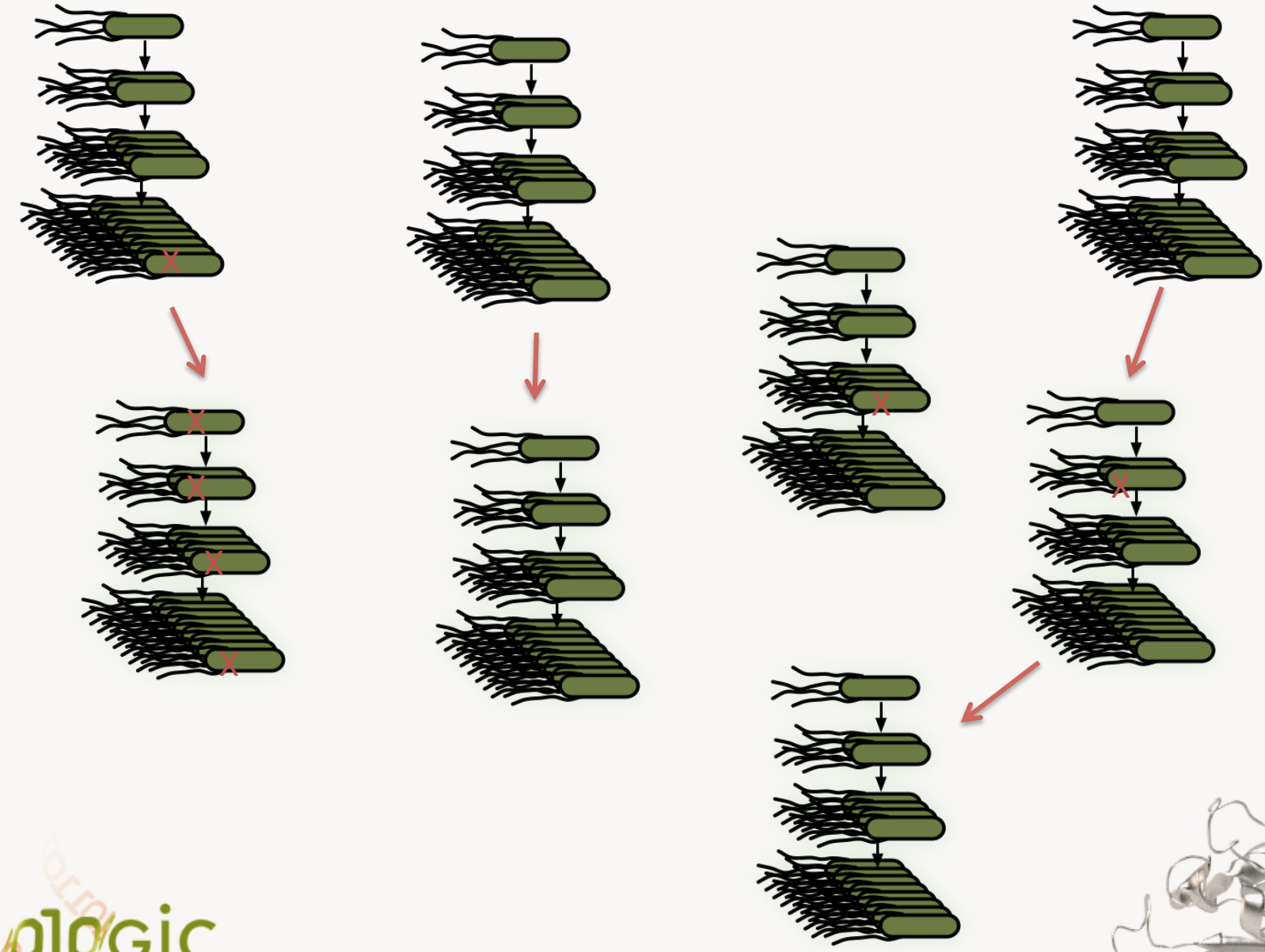
No function until all seven appear.

biologic INSTITUTE

Population Model



biologic INSTITUTE



biologic INSTITUTE

OPEN ACCESS

BIO-complexity

Research Article

The Limits of Complex Adaptation: An Analysis Based on a Simple Model of Structured Bacterial Populations

Douglas D. Axe*

Biologic Institute, Redmond, Washington, USA

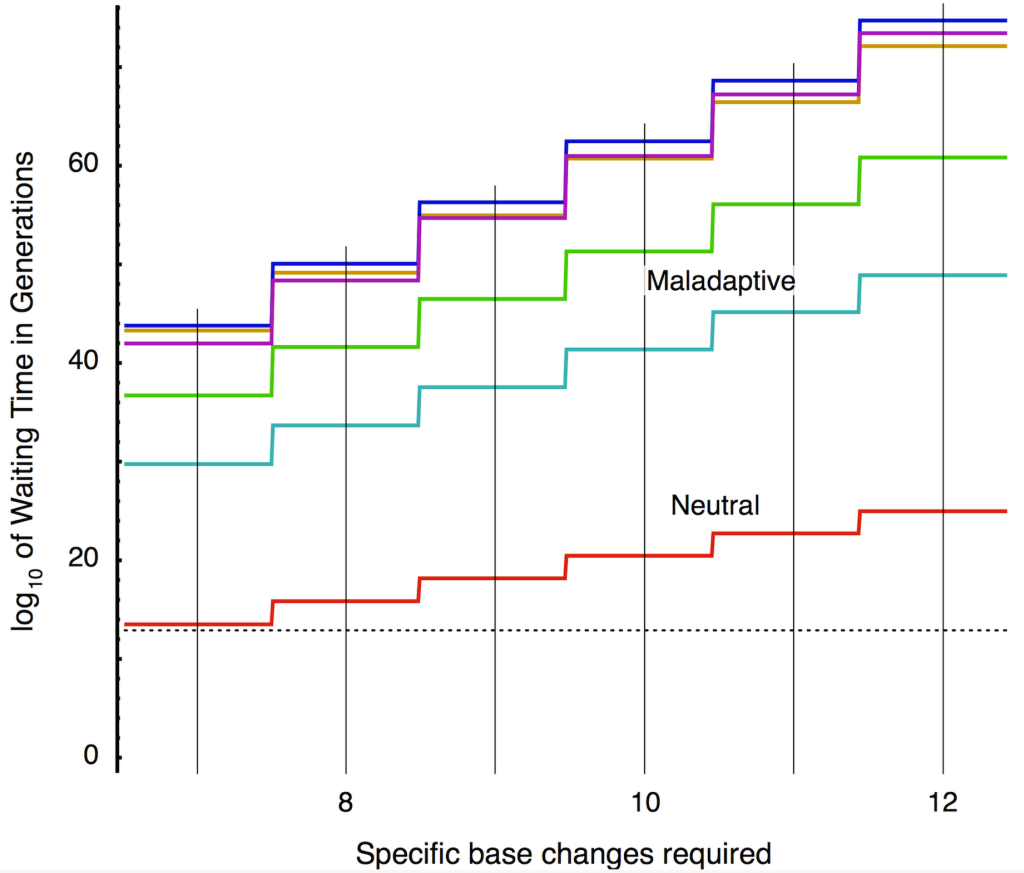
Abstract

To explain life's current level of complexity, we must first explain genetic innovation. Recognition of this fact has generated interest in the evolutionary feasibility of complex adaptations—adaptations requiring multiple mutations, with all intermediates being non-adaptive. Intuitively, one expects the waiting time for arrival and fixation of these adaptations to have exponential dependence on d, the number of specific base changes they require. Counter to this expectation, Lynch and Abegg have recently concluded that in the case of selectively neutral intermediates, the waiting time becomes independent of d as d becomes large. Here, I confirm the intuitive expectation by showing where the analysis of Lynch and Abegg erred and by developing new treatments of the two cases of complex adaptation—the case where intermediates are selectively maladaptive and the case where they are selectively neutral. In particular, I use an explicit model of a structured bacterial population, similar to the island model of Maruyama and Kimura, to examine the limits on complex adaptations during the evolution of paralogous genes—genes related by duplication of an ancestral gene. Although substantial functional innovation is thought to be possible within paralogous families, the tight limits on the value of d found here ($d \leq 2$ for the maladaptive case, and $d \leq 6$ for the neutral case) mean that the mutational jumps in this process cannot have been very large. Whether the functional divergence commonly attributed to paralogs is feasible within such tight limits is far from certain, judging by various experimental attempts to interconvert the functions of supposed paralogs. This study provides a mathematical framework for interpreting experiments of that kind, more of which will needed before the limits to functional divergence become clear.

Cite as: Axe DD (2010) The limits of complex adaptation: An analysis based on a simple model of structured bacterial populations. BIO-Complexity 2010(4):1-10. **doi:10.5048/BIO-C.2010.4**

biologic INSTITUTE

Population model:



Conclusions

Unguided genetic innovations have to be surprisingly common (occur within a few steps) if they are eventually to become established.

biologic INSTITUTE

Innovations requiring more than three specific mutations are probably not feasible.

This seems to exclude most functional conversions for which there is no starting activity.

biologic INSTITUTE

Similarity of structure does not guarantee ease of conversion or common ancestry.

Objection

Need to use ancestral protein as starting point

biologic INSTITUTE

Response

If evolution is so constrained that only a few starting points will do, this is a tacit admission that conversion is *hard.*

biologic INSTITUTE

Response

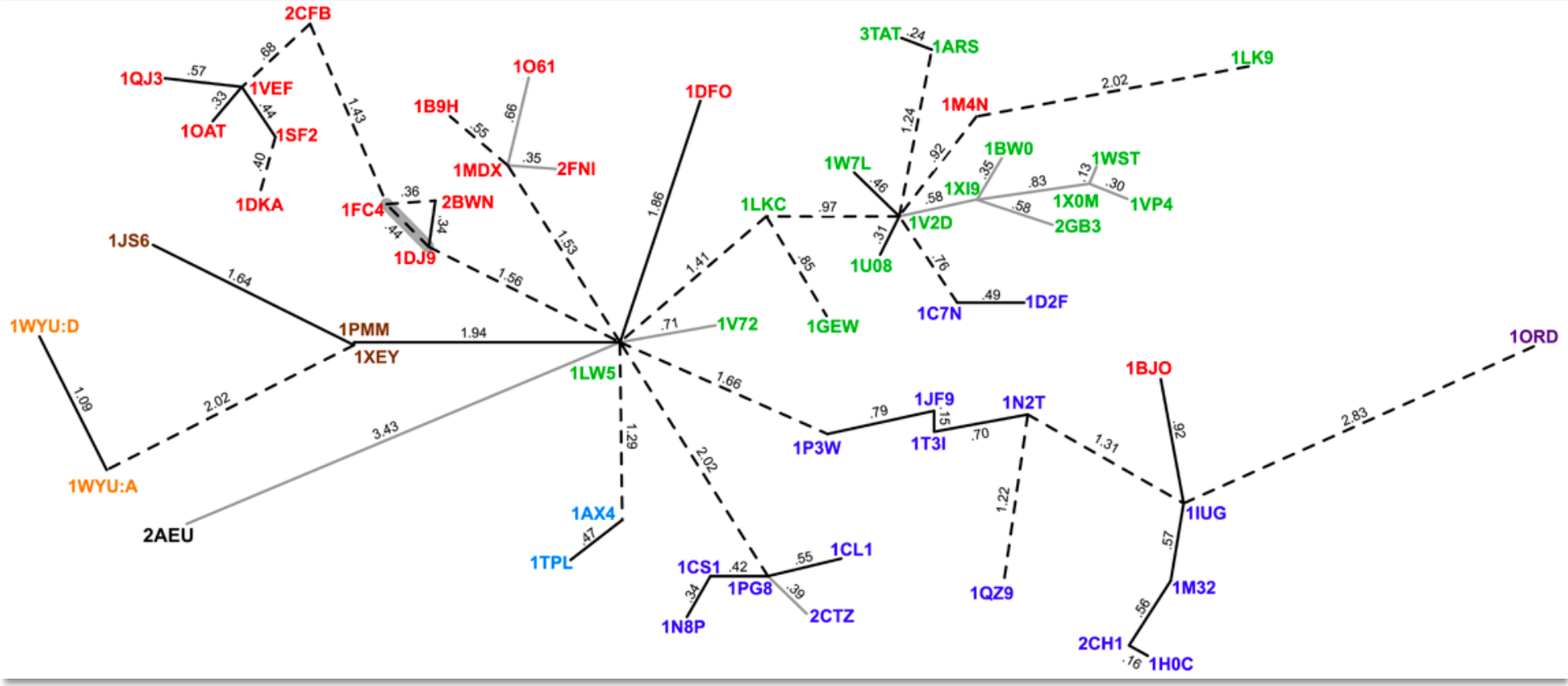
If evolution is so constrained that only a few starting points will do, this is a tacit admission that conversion is *hard.*

To go from several hundred proteins in the first cell, to the thousands we have today would require either *highly unusual ur-proteins* or *guidance*.

biologic INSTITUTE

One PLP-dependent transferase in minimal set

Over fifty unique PLP-dependent enzymes now


2CFB
1QJ3
1VEF
1OAT
1SF2
1DKA
1FC4
2BWN
1DJ9
1JS6
1PMM
1XEY
1WYU:D
1WYU:A
2AEU
1B9H
1O61
1MDX
2FNI
1DFO
1LW5
1V72
1LKC
1GEW
3TAT
1ARS
1W7L
1M4N
1LK9
1BW0
1WST
1XI9
1X0M
1VP4
1V2D
2GB3
1U08
1C7N
1D2F
1P3W
1JF9
1T3I
1N2T
1ORD
1BJO
1IUG
1M32
2CH1
1H0C
1AX4
1TPL
1CS1
1PG8
1CL1
2CTZ
1N8P
1QZ9