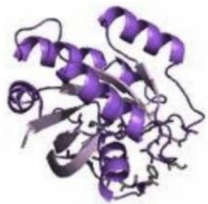
The Origin of Variation in Molecular Complexes:

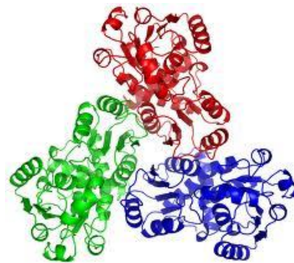
Driven by Adaptive Processes Unique to Individual Lineages,

a Simple Outcome of Stochastic Processes, or

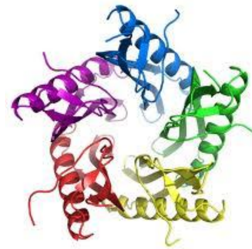
a Consequence of Biased Mutation Pressure and Biophysical Factors?



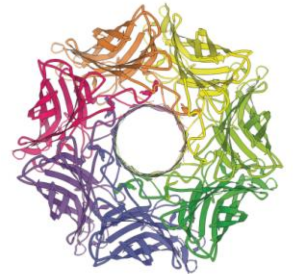
monomer



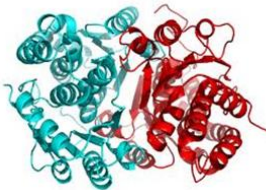
trimer



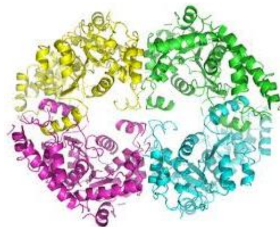
pentamer



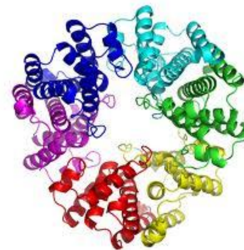
heptamer



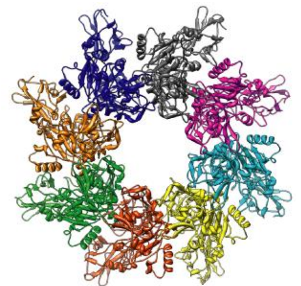
dimer



tetramer



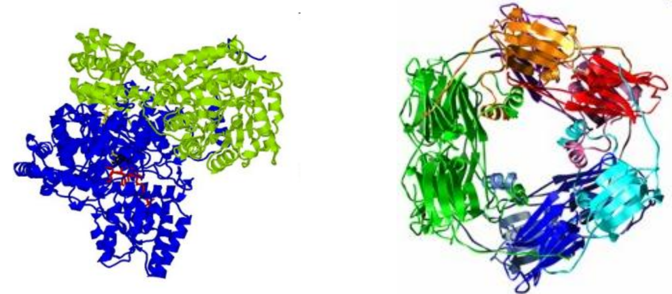
hexamer



octamer

- Most cellular components and pathways are assembled from protein subunits derived from the same gene or from related loci arising via gene duplication, rather than from products of unrelated genes:

  flagella, nuclear pore complexes, spliceosomes
  cytoskeleton
  chaperones and proteasomes
  receptors and ion channels
  nucleosomes and chromatin-remodeling complexes
  transcription factor complexes
  metabolic enzymes

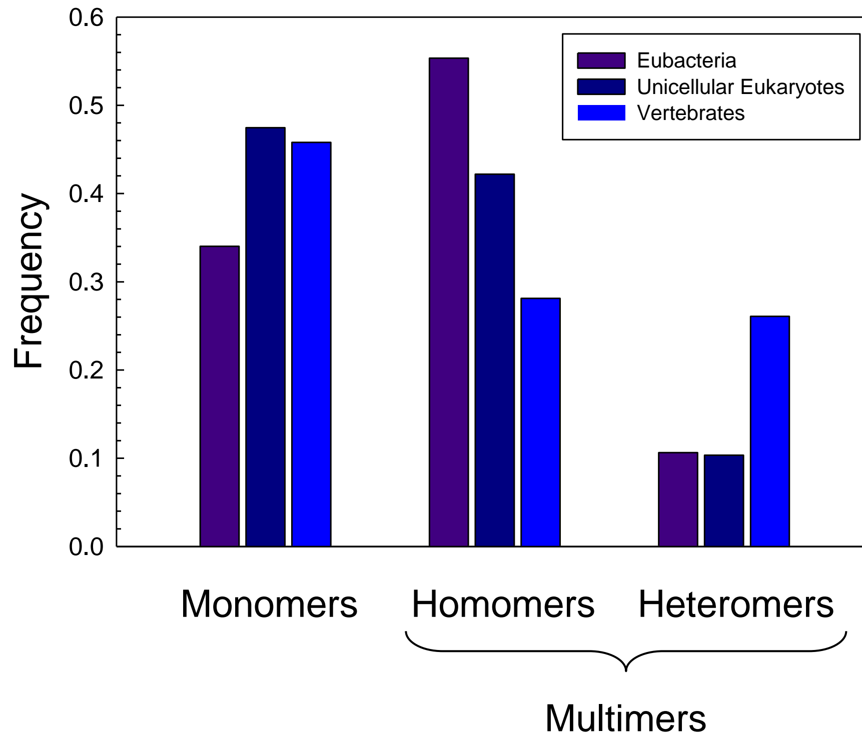


- Potential advantages to complex formation:

  increased structural diversity,
  increase enzyme size and reduced surface area will increase productive encounter rate with substrate,
  reduced problems of folding single large proteins,
  reduced vulnerability to denaturation and/or engagement in promiscuous interactions,
  reduced molecular motion at the catalytic site increases substrate specificity,
  increased flexibility for allosteric regulation,
  compensation for structural deficiencies in monomeric subunits?

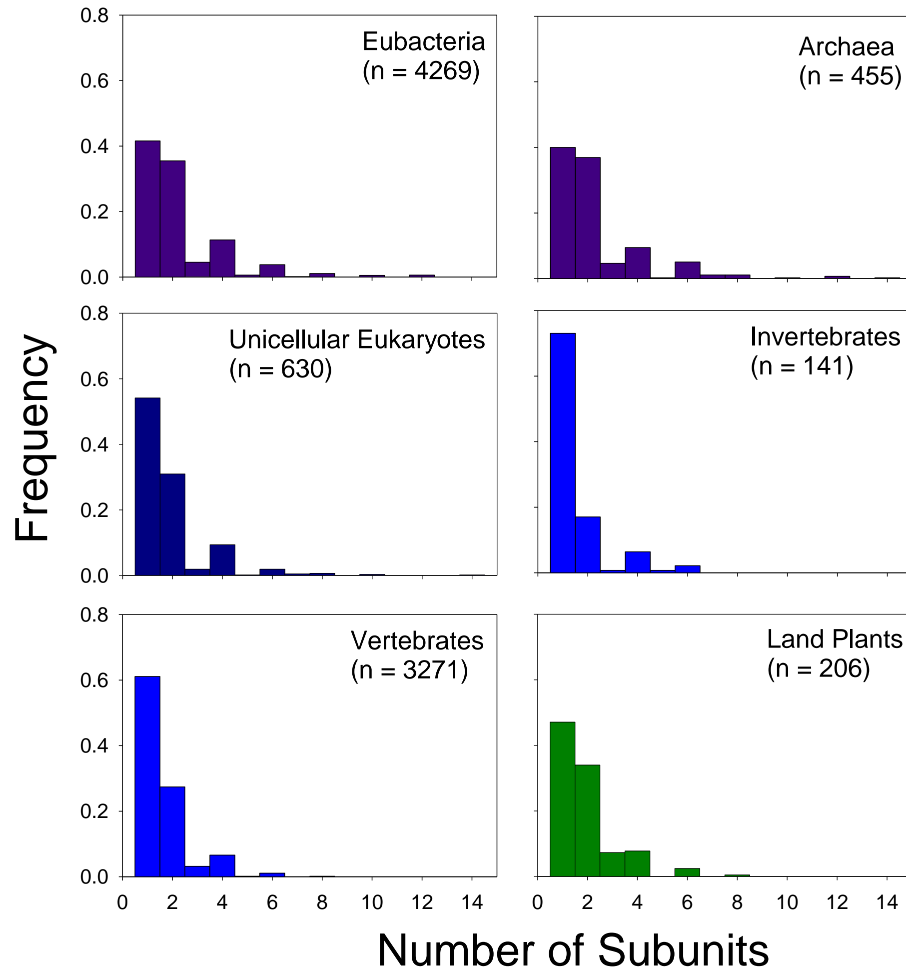
- Proteins with an affinity to oligomerize can also come at a cost:
  - Elevated production levels necessary for a critical encounter rate for successful multimerization.
  - Problems with harmful interactions between heterotypic molecules in heterozygotes in the establishment phase.
  - Concatenation into indefinite filaments – human disorders involving the production of inappropriate protein aggregates include Alzheimer’s, Parkinson’s, and amyotrophic lateral sclerosis (ALS).

Distribution of Complex Types



- Approximately two thirds of proteins with known structures exist as dimers or higher-order complexes.
- Most of these are homomers, with all subunits derived from the same locus.

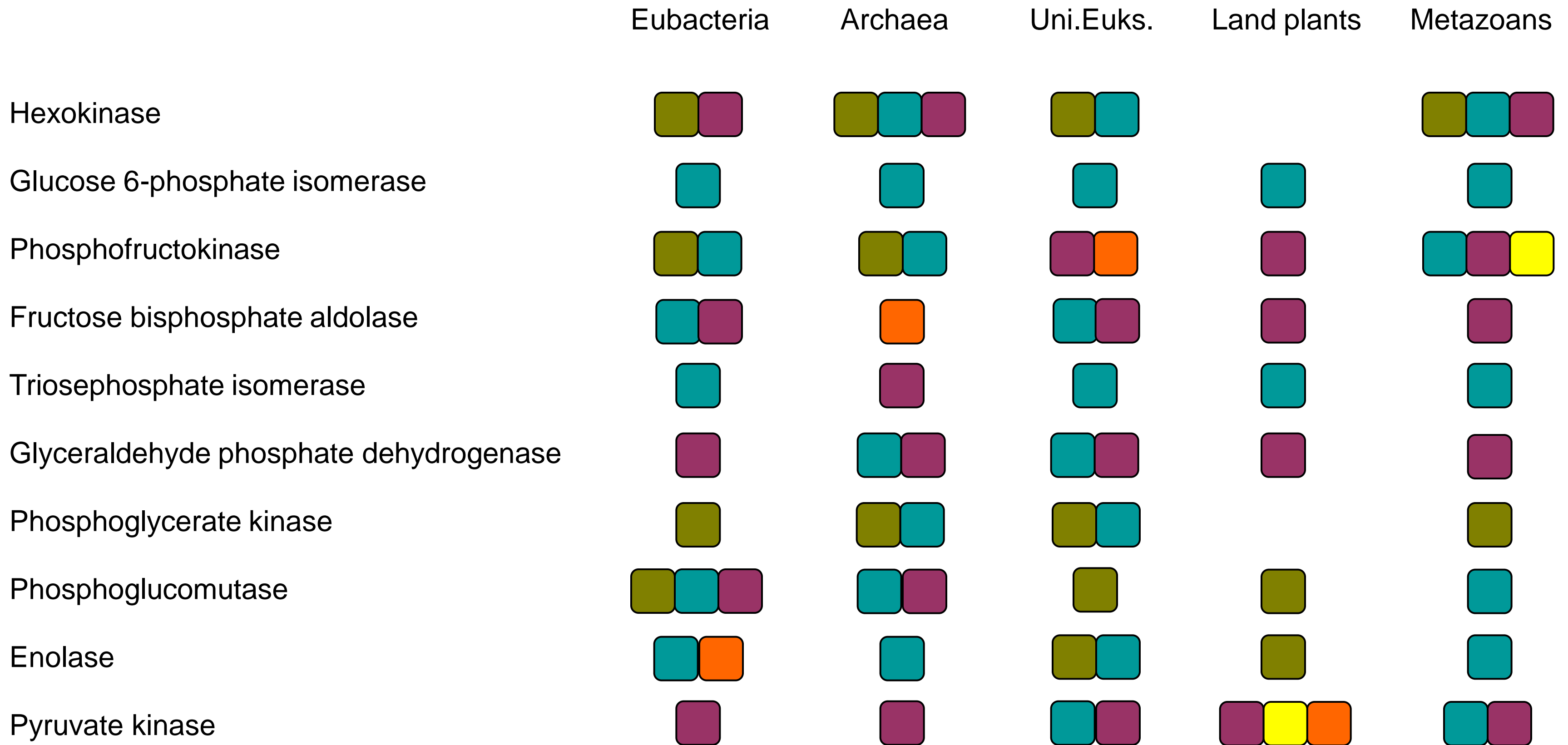
Distribution of Homomeric Types: approximate constancy across the tree of life.



- Roughly two thirds of multimers are dimeric.
- ~15% are tetramers, most of which are "dimers of dimers," most likely arising via an intermediate dimeric state.
- Odd-mers are greatly under-represented.

Data from: 3D Complex.org (Levy et al. 2008)

Known Oligomerization Structures for the Enzymes of Glycolysis



Monomer Dimer Trimer Tetramer Hexamer Octamer

Known Oligomerization Structures for the Enzymes of Citric Acid Cycle

	Eubacteria	Archaea	Uni.Euks.	Land plants	Metazoans
Citrate synthase	Dimer, Hexamer	Dimer	Dimer		Dimer
Isocitrate dehydrogenase	Monomer, Dimer, Tetramer	Dimer	Octamer		Dimer, Tetramer
Fumarase	Monomer, Tetramer	Dimer, Tetramer	Dimer, Tetramer		Tetramer
Malate dehydrogenase	Dimer, Tetramer	Dimer, Tetramer	Dimer	Dimer	Dimer

- There is substantial variation in the multimeric states both within and among phylogenetic groups.
- For enzymes, multimers are almost always homomeric.
- No tendency for more complex organisms to harbor more complex molecules – **in striking contrast to what is seen with the complexity of gene structure and genome architecture**.

Monomer Dimer Trimer Tetramer Hexamer Octamer

Enzymes with Identical Multimeric States Need Not Have the Same Structural Basis

Dihydrodipicolinate synthase (involved in lysine synthesis)

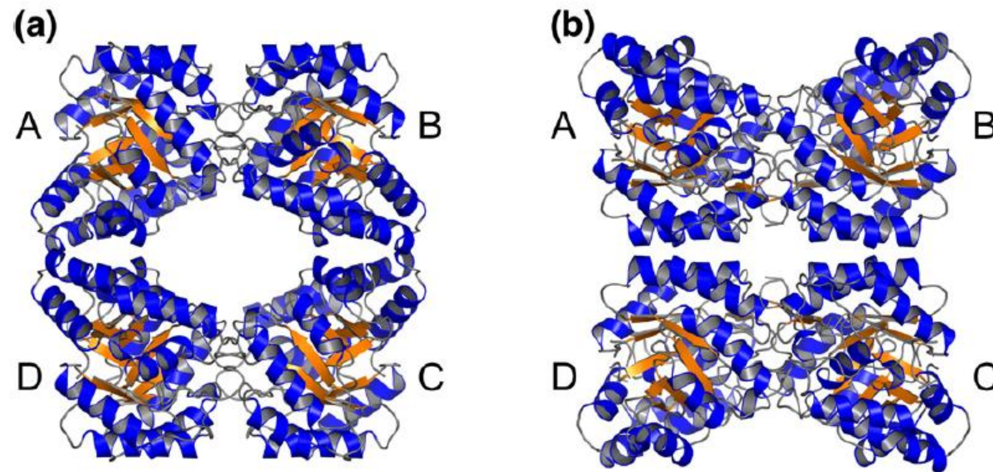
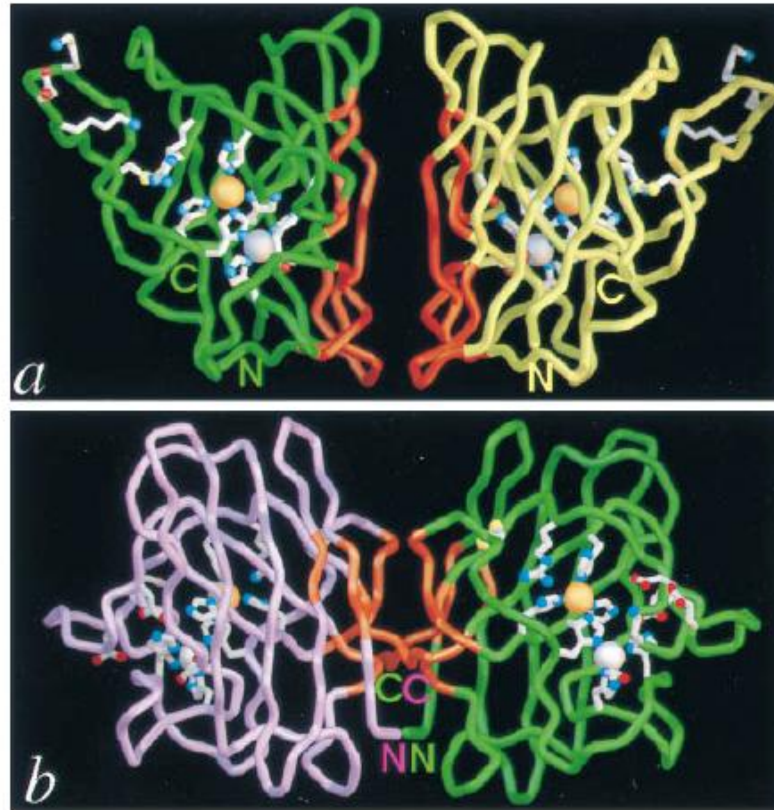


Fig. 1. The X-ray crystal structures of DHDPS from (a) *E. coli*[18,19] and (b) *N. sylvestris*.[23] Each enzyme is a homotetramer of $(\beta/\alpha)_8$-barrels composed of two tight-dimer units (A–B and C–D), but the arrangement of the two dimeric units is different.

Both species make homotetramers, but the dimer-dimer interfaces are completely nonoverlapping, face to face in the former, and back to back in the latter (Griffin et al. 2008).

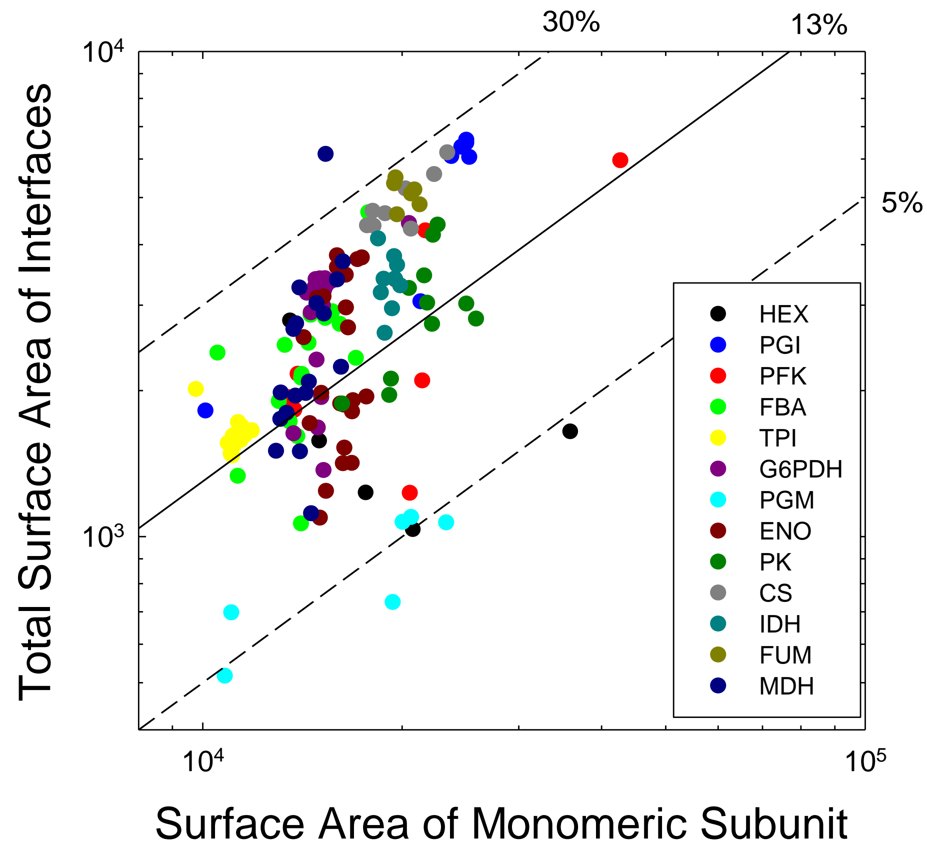
Cu,Zn Superoxide Dismutase:

Dimer interfaces in *Photobacterium* (above) and cow (below) are constructed from diametrically opposite beta-barrel elements (Bourne et al. 2008).



• Dayhoff et al. (2010) estimate that about two-thirds of protein families containing homomers exhibit phylogenetic variation in the binding interfaces.

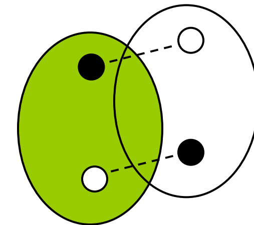
Substantial Variation in Interface Sizes Exist Among Species



Some Pure Biophysical Explanations for Frequent Homomers

- To ensure stable complexation, interfaces must overcome the energetic cost of thermal motion. Random symmetrical interfaces are more likely to generate extremes of binding strength than random asymmetric interfaces (Lukatsky et al. 2007; Andre et al. 2008).

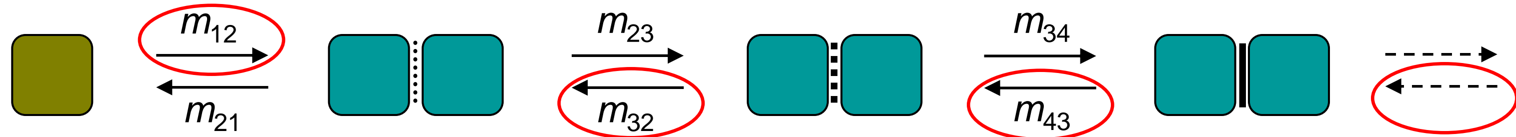
  Two for the price of one: any pair of adhesive residues in a symmetric interface must be present twice (Monod et al. 1965).



- Colocalization ensures the opportunity for coevolution.

- Protein surfaces are generally selected to be hydrophilic, which generates a natural mutational bias in the direction of hydrophobic (sticky) residues.

Evolution of a Dimeric Structure

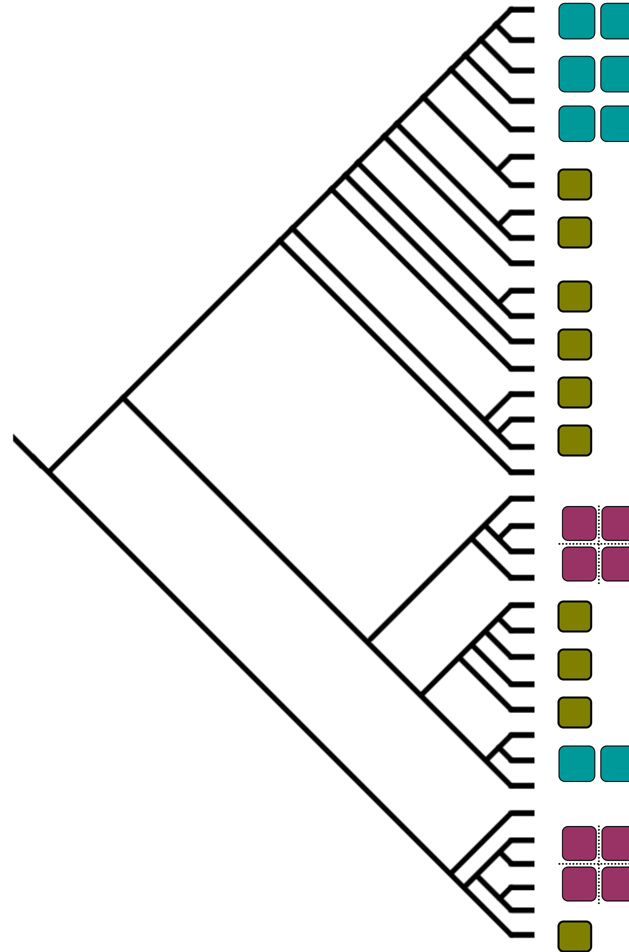


- Each transition rate is equal to the product of the number of relevant mutations arising per generation and the fixation probability.

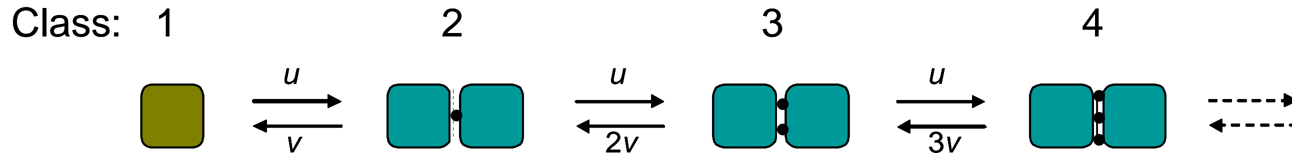
- At steady state, the flux rate must be equal in both directions. This means that the net rate of establishment of dimers from monomers must equal the reverse rate.

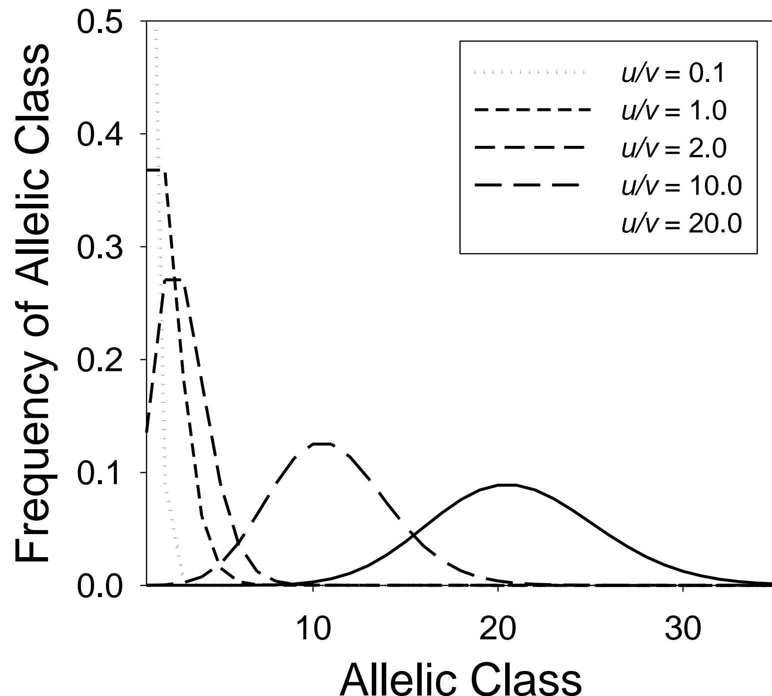
- The equilibrium probability of each state is simply proportional to the product of the total set of transition rates towards the state from both directions.

A Hypothetical Steady-state Distribution Unveiled on a Phylogeny



The Neutral Expectation: the steady-state distribution of alternative allelic states is Poisson, a simple function of the ratio of upward and downward mutation rates, independent of population size.





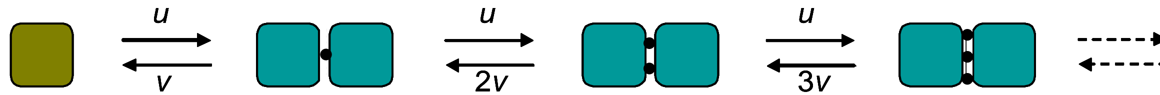
$$\widetilde{P}_i = \frac{(u/v)^{i-1}}{(i-1)!C}$$

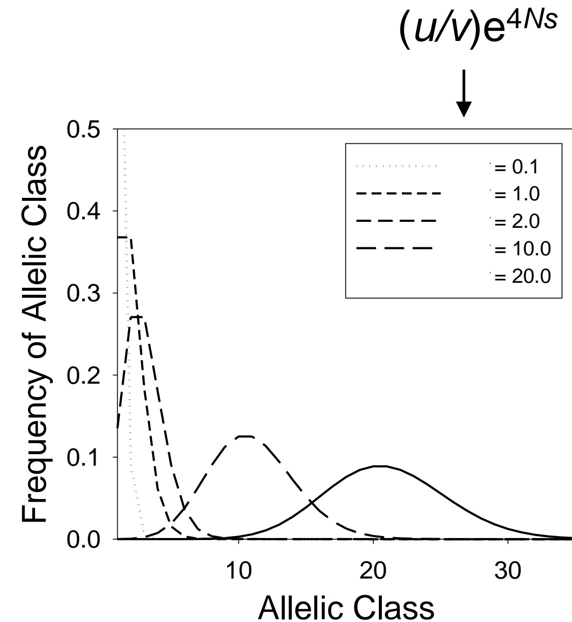
$$C = \sum_{i=0}^{\infty} (u/v)^i / i! = e^{u/v}$$

u/v is the mutation bias.

Expected frequency of monomers = $e^{-u/v}$

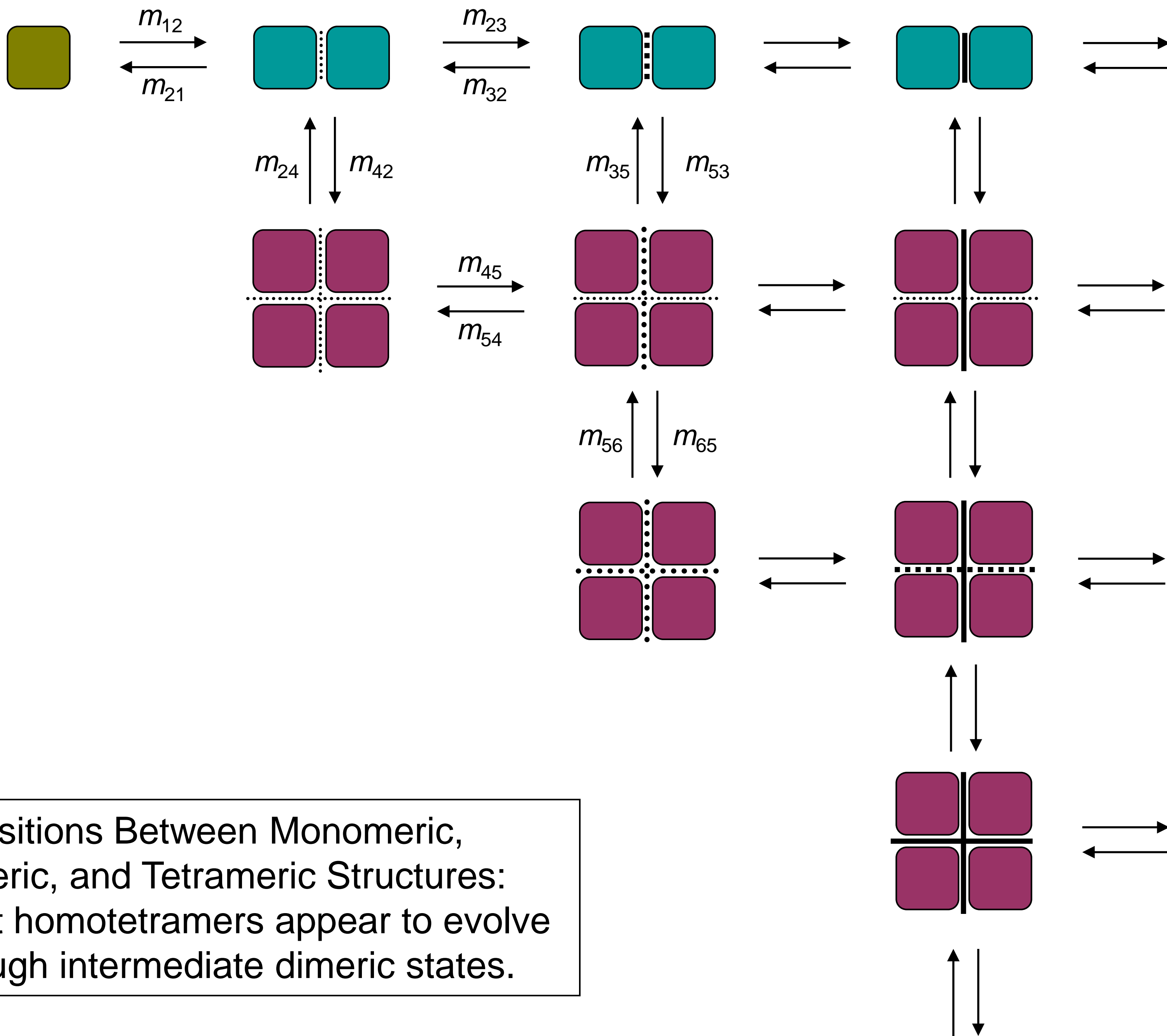
Adding in Selection: s is the selective advantage (or disadvantage) of each incrementing allele.





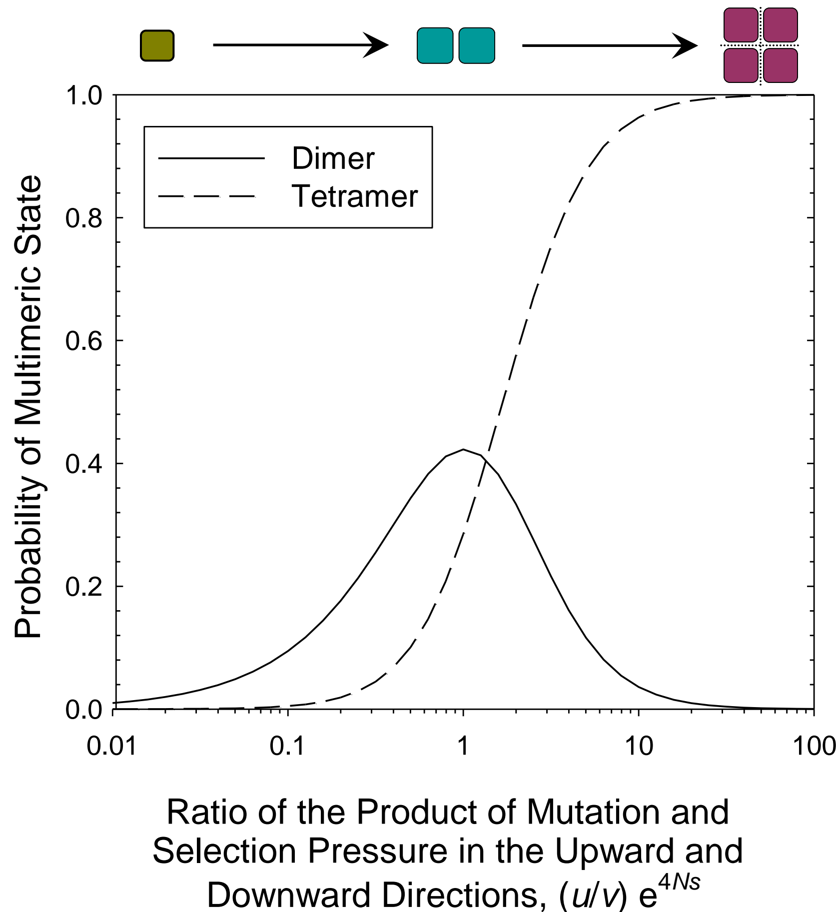
- The distribution is again Poisson, but now the key parameter is $(u/v)e^{4Ns}$.
- The quantity e^{4Ns} is the ratio of fixation probabilities of beneficial and detrimental mutations with the same absolute s – the strength of selection bias.
- The effects of selection, drift, and mutation bias cannot be disentangled with the steady-state distribution alone.

- Even if the direction of selection is constant in all lineages, there will be substantial phenotypic variation unless the joint directional bias due to the forces of mutation and selection are << 0.1 or >> 5.0.
- Even with *negative* selection against dimers, they will still be common provided the mutational bias is sufficiently large.
- If the ratio of the power of drift to selection, $s / [1/(2N)] = 2Ns$, is < 1.0, the phenotypic distribution is entirely driven by mutation bias.



Transitions Between Monomeric, Dimeric, and Tetrameric Structures: most homotetramers appear to evolve through intermediate dimeric states.

Steady-state Distributions of Monomeric, Dimeric, and Tetrameric Structures



$$\widetilde{P}_1 = 12(2+\theta)/C$$

$$\widetilde{P}_2 = 12\theta(2+\theta)/C$$

$$\widetilde{P}_3 = 4\theta^2(3+\theta)/C$$

$$\widetilde{P}_4 = 4\theta^2(3+2\theta)/C$$

$$\widetilde{P}_5 = 4\theta^3(2+\theta)/C$$

$$\widetilde{P}_6 = \theta^4(2+\theta)/C$$

$$\theta = (u/v)e^{4Ns}$$

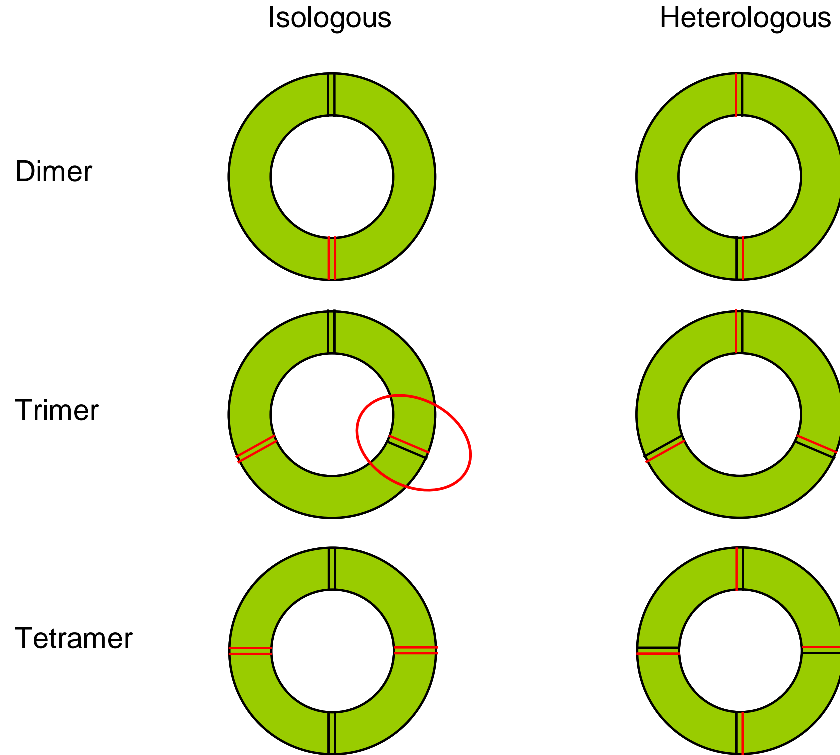
- Again, the probability distribution of alternative phenotypes is determined by a single composite function of the power of mutation, drift, and selection.

- If this parameter is in the range of 0.1 to 10.0, all three structures are likely to be common across the tree of life, even in the face of constant natural selection.

Simple Geometric Limitations on Oddmers:

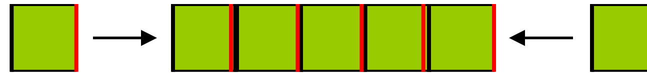
• Can't take the "two for the price of one" route because an isologous structure cannot be completed.

• Heterologous structures may often require twice the number mutations.



• Unless equipped with the correct angular orientation for creating a closed loop, heterologous interfaces encourage concatenation into endless fibrils.

Closed dimer



Open-ended with two interfaces

General Conclusions

- Substantial phenotypic variation can arise among lineages owing to the joint stochasticity of the forces of mutation, drift, and selection, ***even when selection and mutation is operating in an identical manner in all lineages***.

- Patterns of phylogenetic variation in molecular cooperativity cannot be understood by focusing on selection alone.

- There is a fundamental distinction between the mechanisms responsible for the origin of a biological feature and natural selection’s subsequent involvement in molding it into an adaptation.

"Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneous with being zoologists and botanists? Let us hope so."

H. J. Muller (American Naturalist, 1922)

Research Funding:

Desks, Stimulation, Salary:

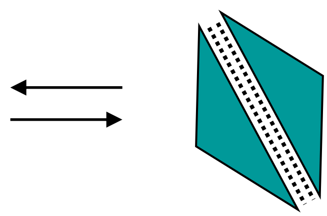
Are Multimeric Molecules Functionally Superior to Monomers?

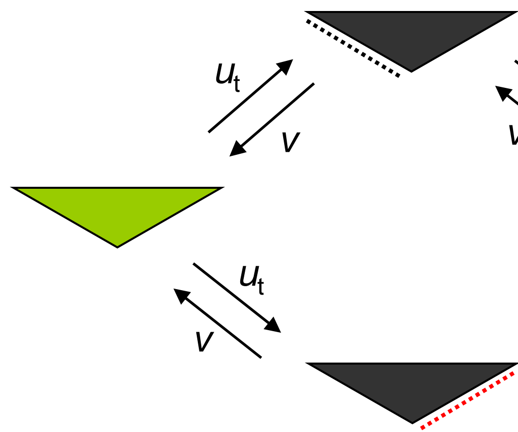
- The mismatch-repair machinery in eubacteria employs monomers, whereas that in eukaryotes employs dimers, and yet MMR efficiency is greater in eubacteria.

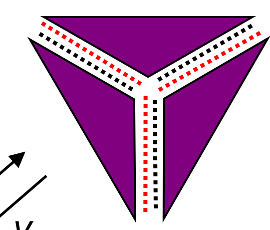
- Sliding clamps used in DNA replication are homodimers in eubacteria, but homotrimeric in eukaryotes, and yet replication-fork progression rates are an order of magnitude greater in eubacteria.

- The protein repertoire of eukaryotic ribosomes is substantially more complex than that in prokaryotes, and yet the level of translation fidelity is no greater (and probably lower) in eukaryotes.

- Class II amino-acyl tRNA synthetases are dimeric or tetrameric, yet monomeric class I synthetases are much less error-prone with respect to amino-acid charging.

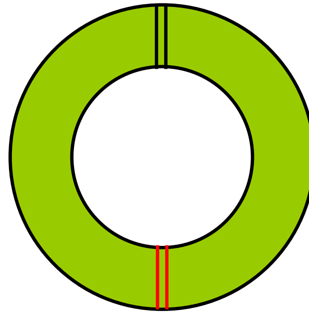
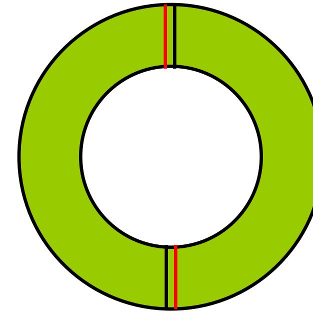
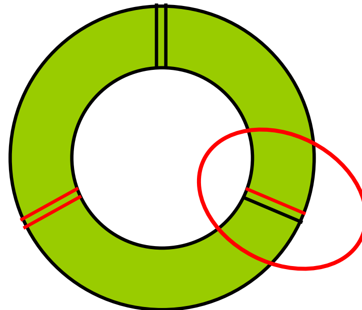
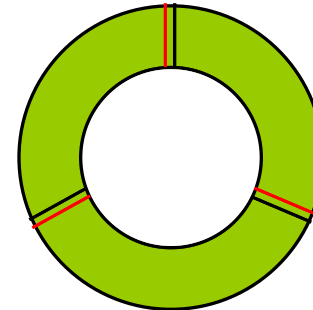
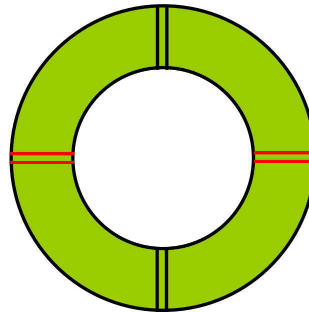
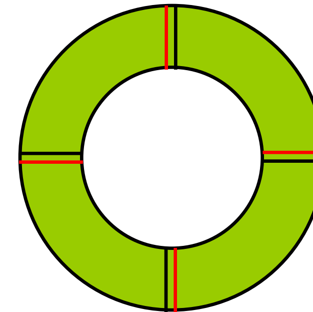




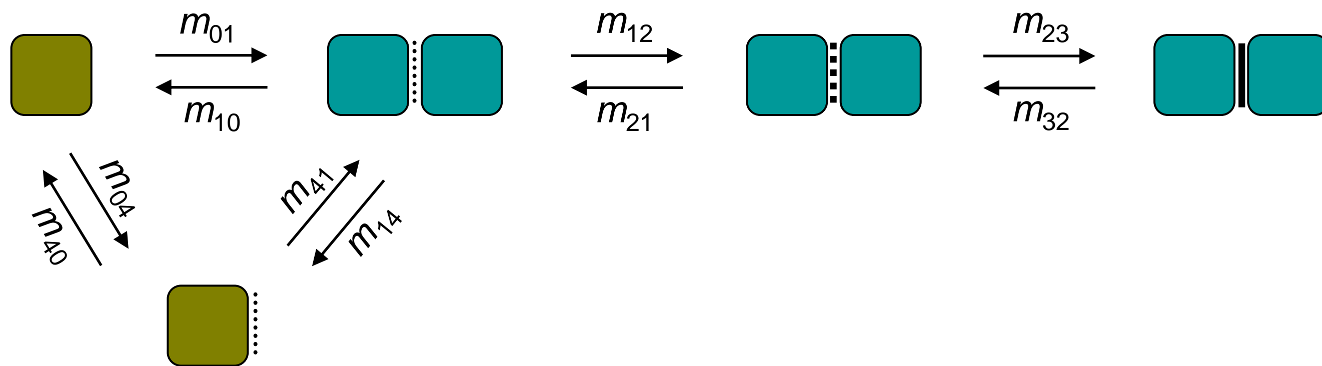


Selection: s_d 0 s_0 s_t

The Problems with Oddmers

	Isologous	Heterologous
Dimer		
Trimer		
Tetramer		

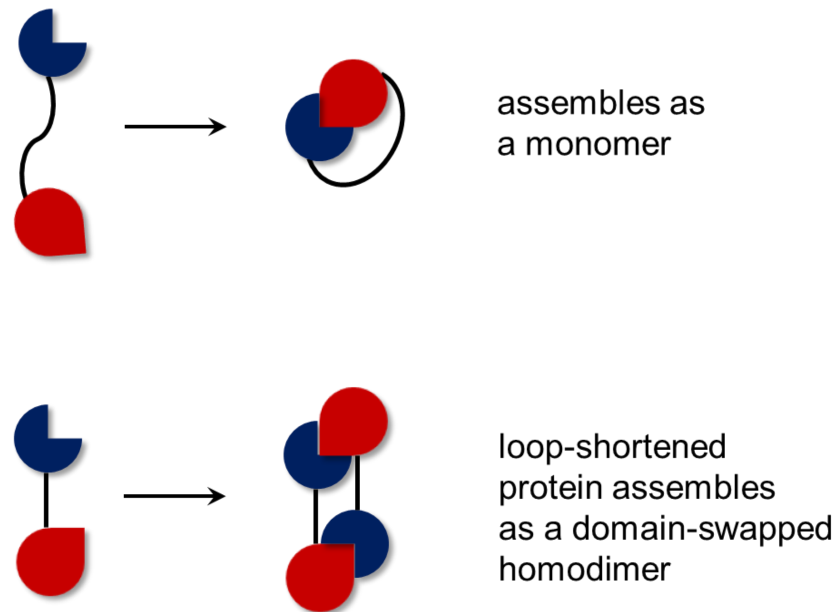
Evolution of a Dimeric Structure, Allowing for a Deleterious Monomer



The origin of protein interactions and allostery in colocalization

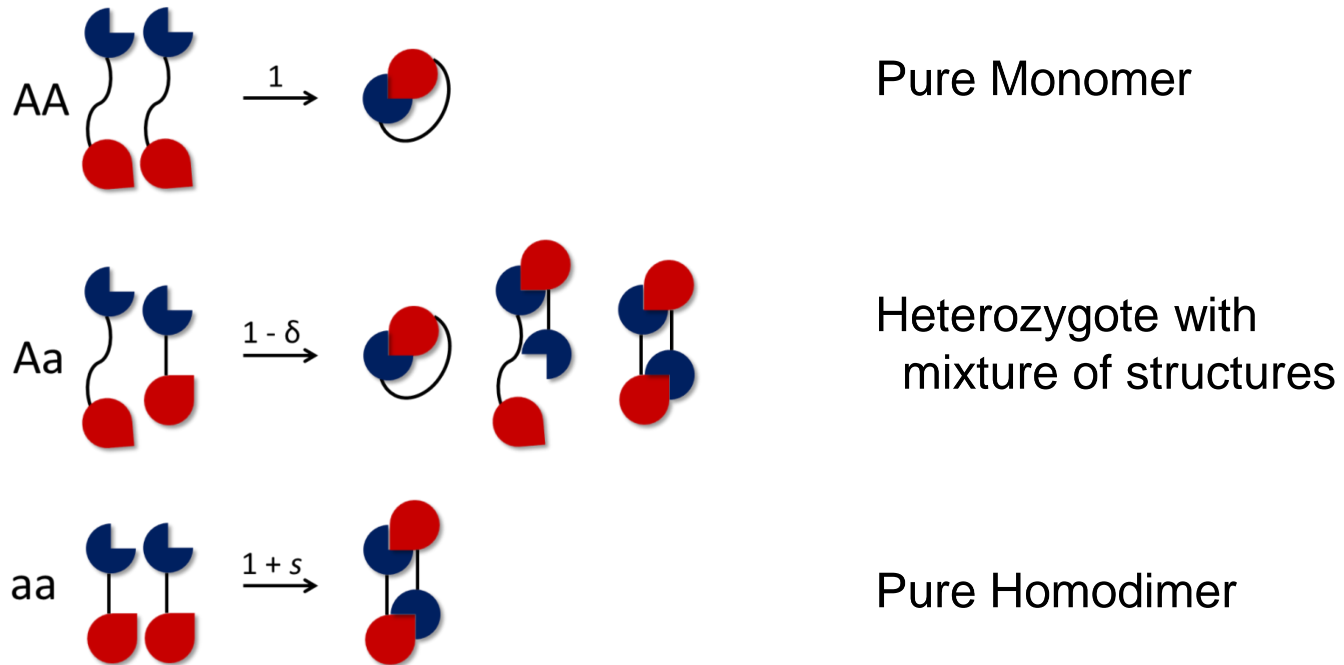
John Kuriyan[1,2] & David Eisenberg[3]

NATURE | Vol 450 | 13 December 2007 | doi:10.1038/nature06524



The Domain-swapping Model

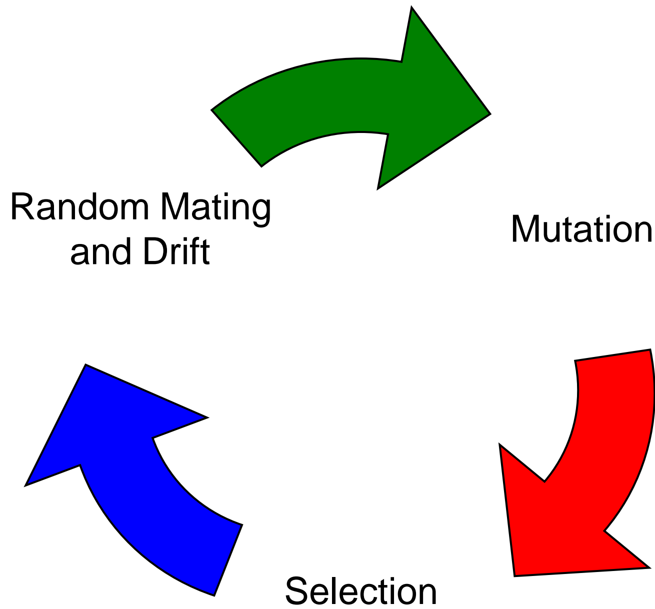
The Population-genetic Conditions for the Origin of Domain Swapping



- Advantages: preadapted to complexation, and only requires a single mutation.

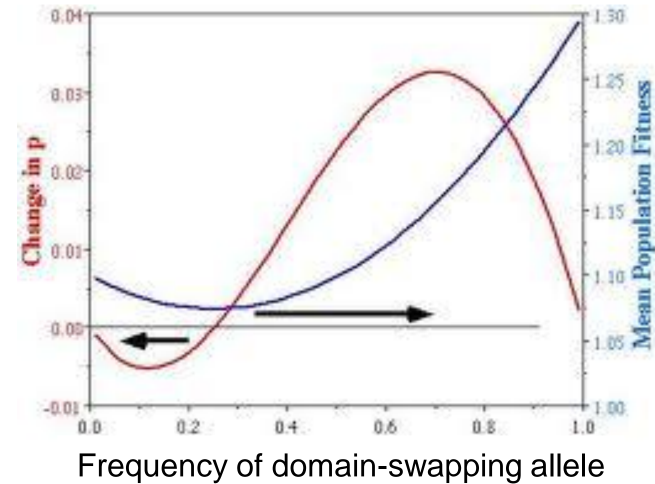
- Disadvantage: reduced heterozygote fitness may impose a strong barrier to fixation;
  the homozygote might also be weakly disadvantageous due to the diffusion barrier to assembly.

With Recurrent Mutational Introduction of the Domain-swapping Allele, How Long Does It Take To Establish (Go To Fixation) In Populations?



- Function of the population size, the mutation rate, and the fitness effects of the domain-swapping allele in heterozygotes and homozygotes

  = rate of input (2Nu) x probability of fixation.



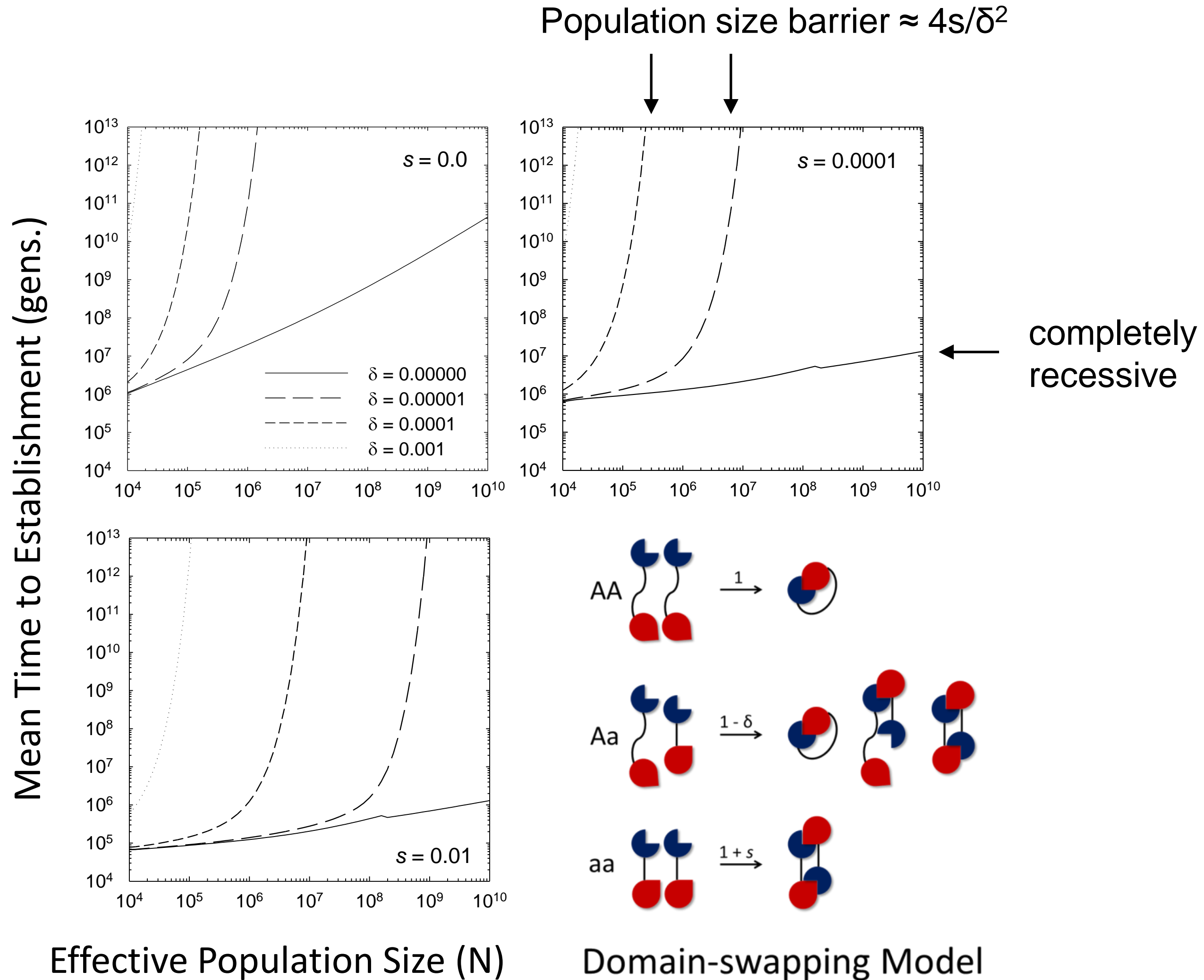
Frequency of domain-swapping allele

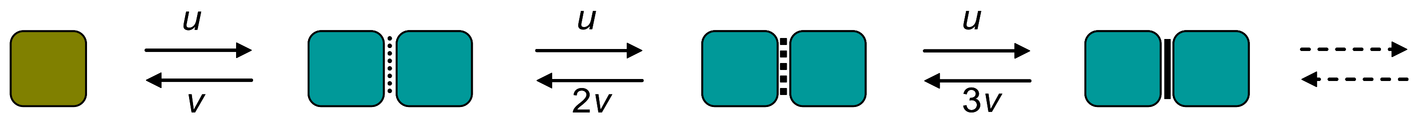
Probability of fixation of an underdominant mutation = $$\frac{\mathrm{erf}\{[p_0 - (0.5/(1+\omega))]\sqrt{4\theta(1+\omega)}\} + \mathrm{erf}\{\sqrt{\theta/(1+\omega)}\}}{\mathrm{erf}\{[1 - (0.5/(1+\omega))]\sqrt{4\theta(1+\omega)}\} + \mathrm{erf}\{\sqrt{\theta/(1+\omega)}\}},$$

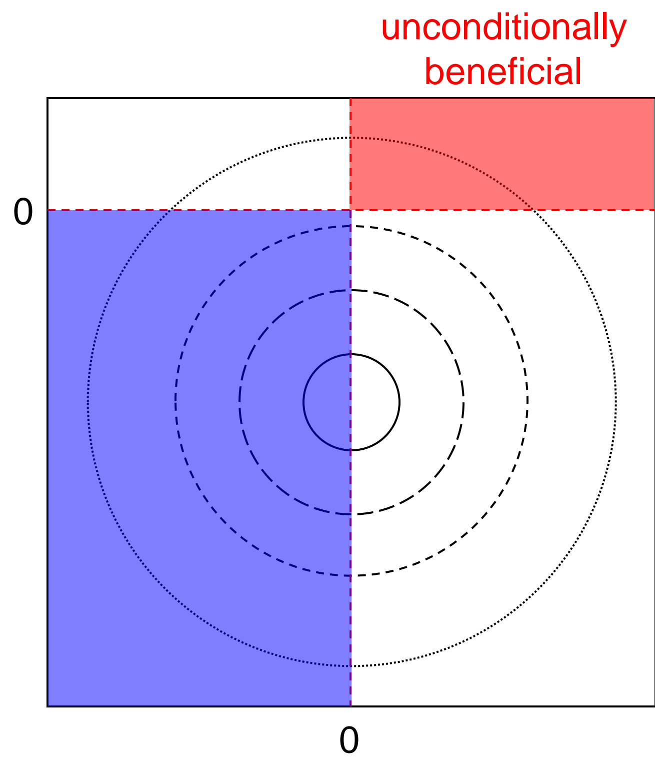
where $\theta = Ns$, $\omega = s/(2\delta)$ (Walsh 1982), and

$$\mathrm{erf}(x) = \int_0^x e^{-y^2}\,dy,$$

Evolution of Domain-swapping Homodimers is Strongly Inhibited in Large Populations, Unless the Heterozygote Disadvantage is Extremely Weak




Class: 1
2
3
4
u
v
u
2v
u
3v



Selective Advantage in a Monomer

Selective Advantage in an Interface

Adding in selection:

The general result:

$$\widetilde{P}_i = \frac{(u/v)^{i-1} e^{-4Ns_i}}{(i-1)!C}$$

- s_i = selective disadvantage of allele *i* (*s* = 0 implies molecular perfection).
- Population size now becomes important, unless the ratio of the power of drift to selection, $s / [1/(2N)] = 4Ns$ is always smaller than 1.0.

Weak and consistent positive selection:

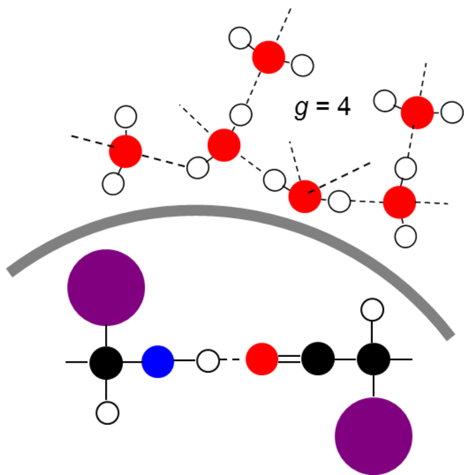
$$\widetilde{P}_i = \frac{[(u/v)e^{4Ns}]^{i-1}}{(i-1)!C}$$

- The distribution is again Poisson, with the probability of the monomeric state being $\exp[-(u/v)e^{4Ns}]$.
- The quantity e^{4Ns} is equivalent to the ratio of fixation probabilities of beneficial and detrimental mutations with the same absolute *s* – the strength of selection bias.
- The effects of selection, drift, and mutation bias cannot be disentangled.
- Even with *negative* selection against dimers, they will still be common provided the mutational bias is sufficiently large.

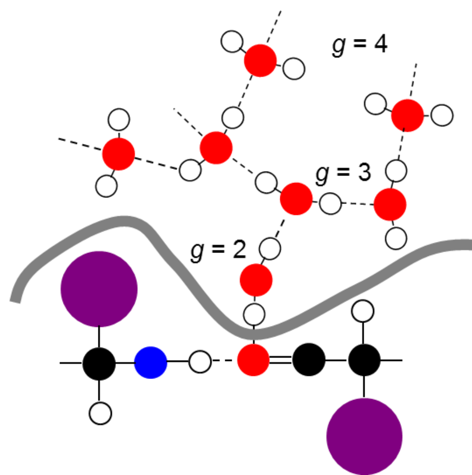
The drift barrier and the asymptotic approach to molecular perfection:

$$\widetilde{P}_i = \frac{(u/v)^{i-1} exp[4Ns_1(1 - e^{-k(i-1)})}{(i-1)!C}$$

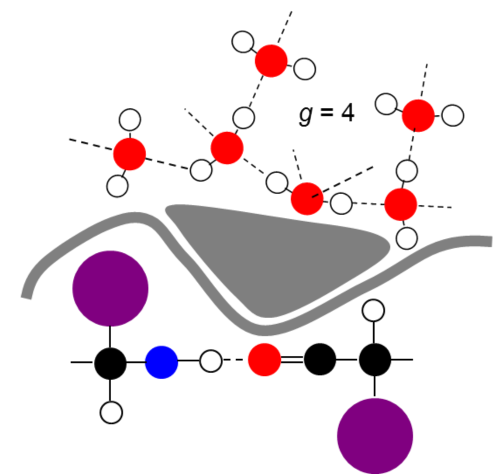
Can Nonadaptive Processes Lead to the Evolution of Protein Complexity?



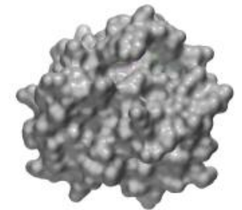
Well-wrapped



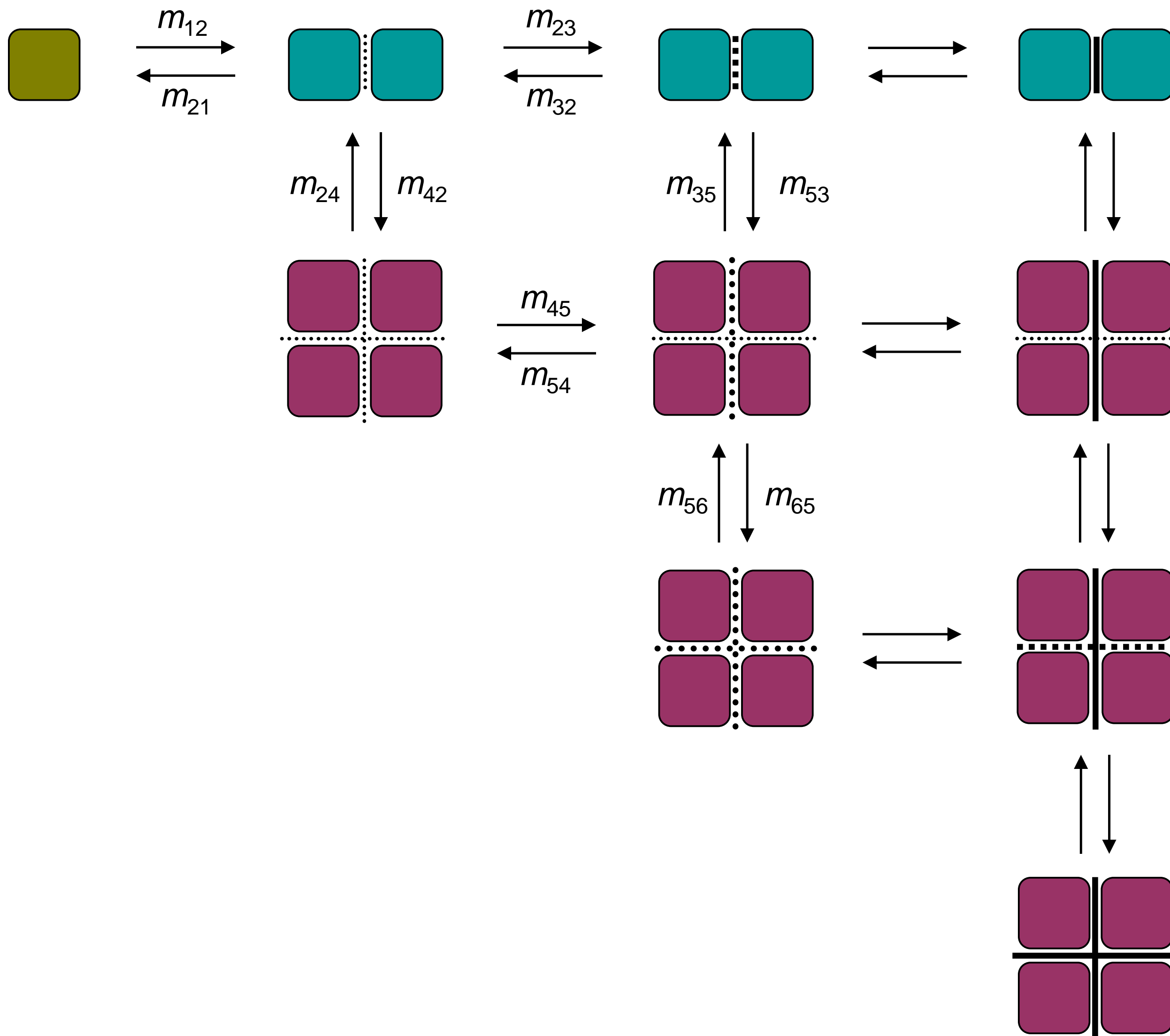
Exposed



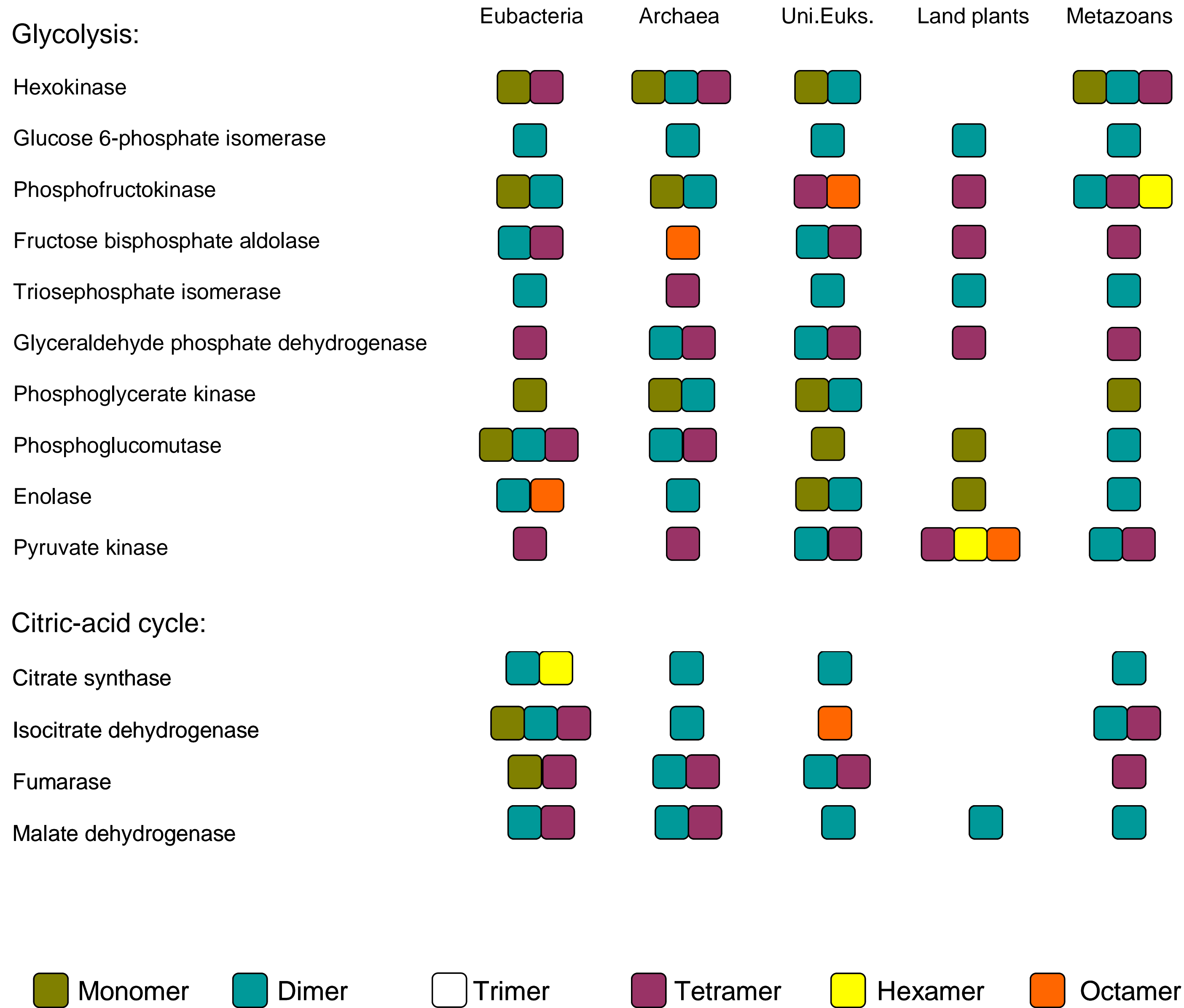
Tension relieved



“Water molecules can intrude and compete for the hydrogen bonds, like lovers undermining a marriage.” Philip Ball



m_{12}
m_{21}
m_{23}
m_{32}
m_{24}
m_{42}
m_{35}
m_{53}
m_{45}
m_{54}
m_{56}
m_{65}


Eubacteria
Archaea
Uni.Euks.
Land plants
Metazoans
Glycolysis:
Hexokinase
Glucose 6-phosphate isomerase
Phosphofructokinase
Fructose bisphosphate aldolase
Triosephosphate isomerase
Glyceraldehyde phosphate dehydrogenase
Phosphoglycerate kinase
Phosphoglucomutase
Enolase
Pyruvate kinase
Citric-acid cycle:
Citrate synthase
Isocitrate dehydrogenase
Fumarase
Malate dehydrogenase
Monomer
Dimer
Trimer
Tetramer
Hexamer
Octamer