Эволюция регуляторных и метаболических путей

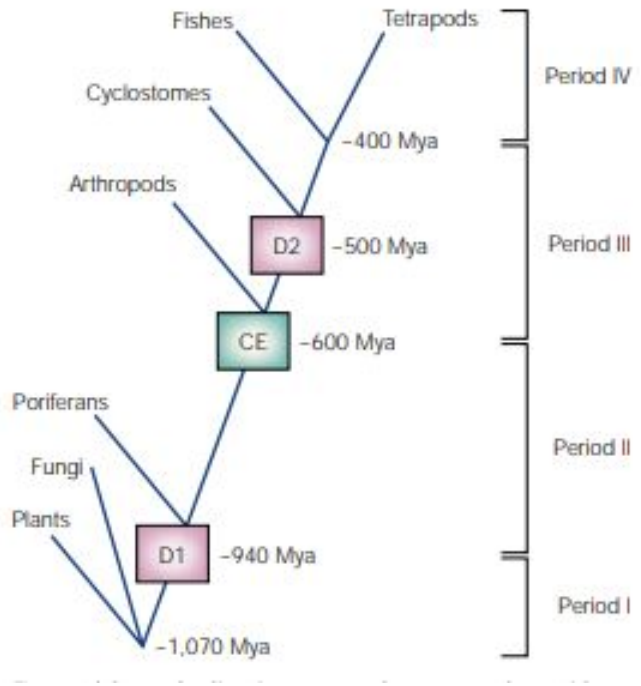
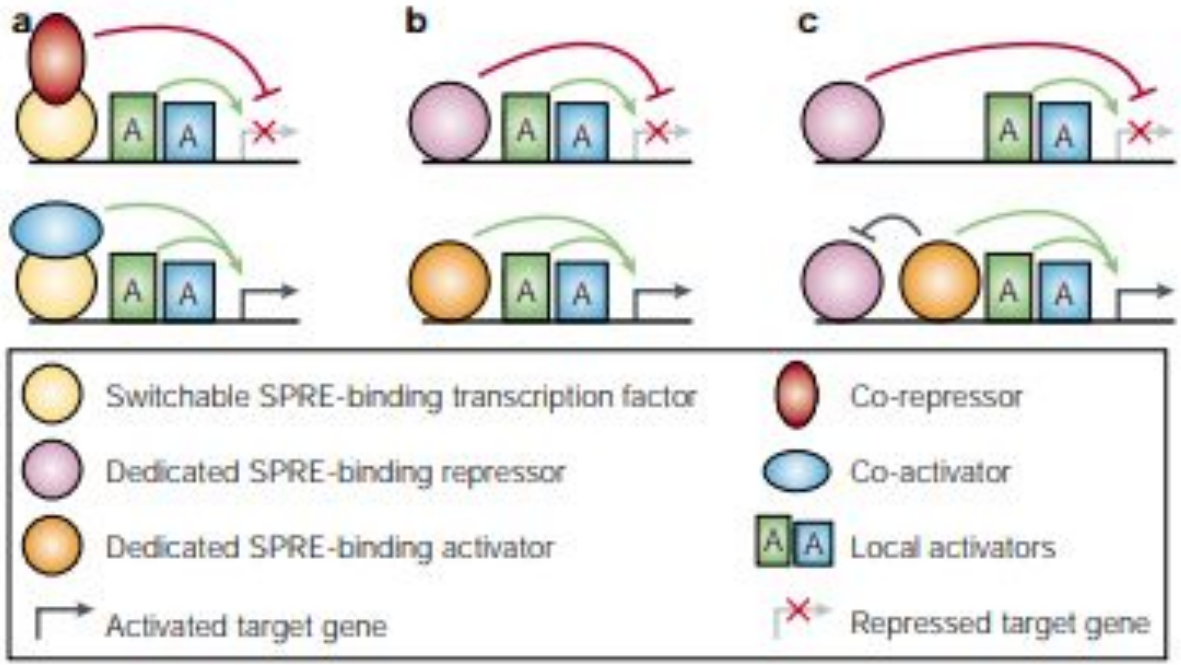


Table 1 | **Numbers of signalling molecules in selected pathways**

Signalling molecules	Species			
	Human	Fly	Worm	Yeast
Ligand				
RTK	48	3	4	0
TGF-β	29	6	4	0
Wnt	18	7	5	0
Notch	3	2	2	0
STAT	7	1	1	0
Receptor				
RTK	25	6	1	0
Wnt	12	6	5	0
NHR	59	25	270	1

NHR, nuclear hormone receptor; RTK, receptor tyrosine kinase; STAT, signal transducer and activator of transcription; TGF-β, transforming growth factor-β; Wnt, wingless related. The table contains selected entries from **REF. 44**.



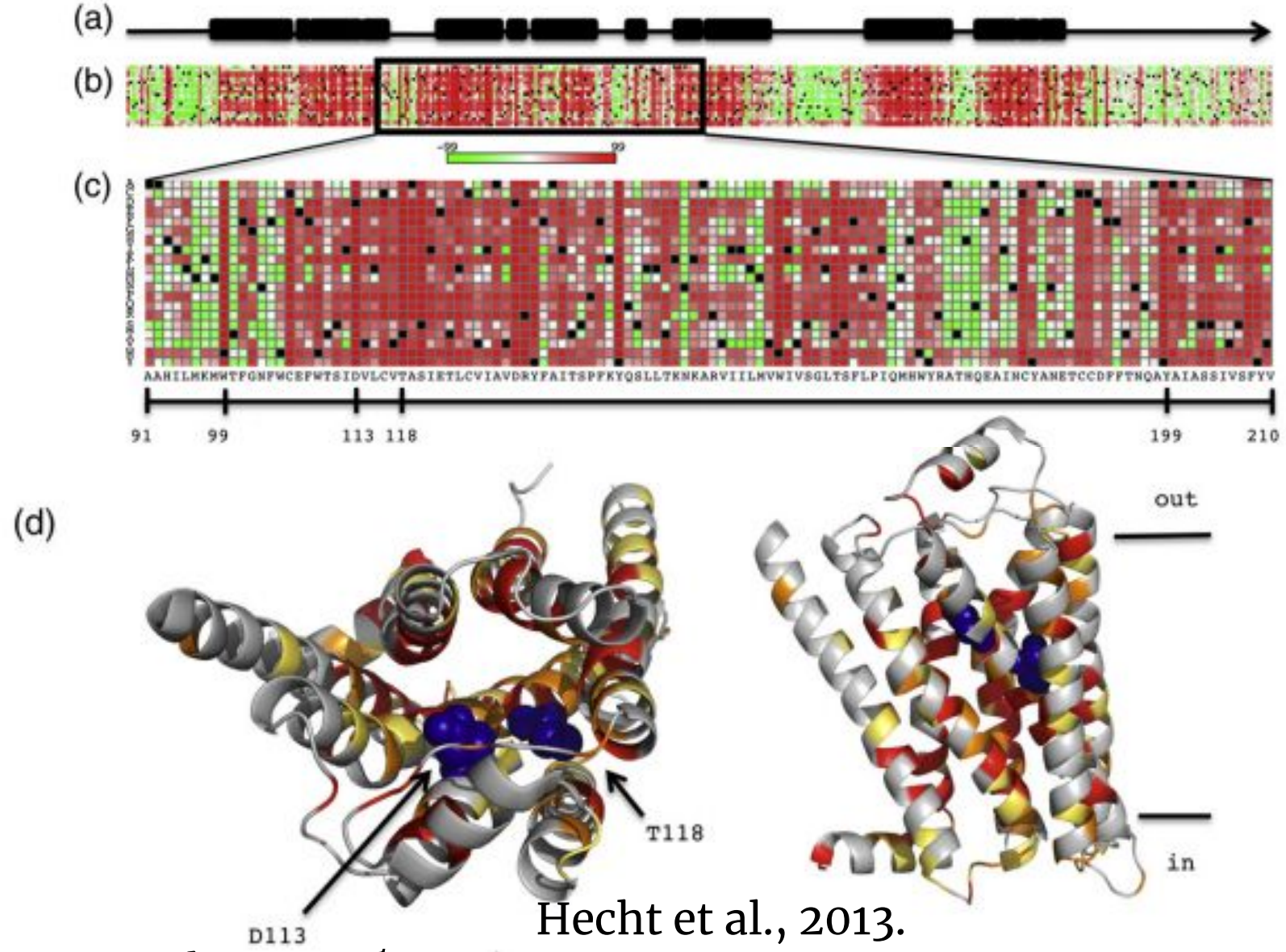
A) Элементы пути, отвечающие на сигнал и их транскрипционные факторы:

Wnt — Tcf/Lef

Notch — Su(H)

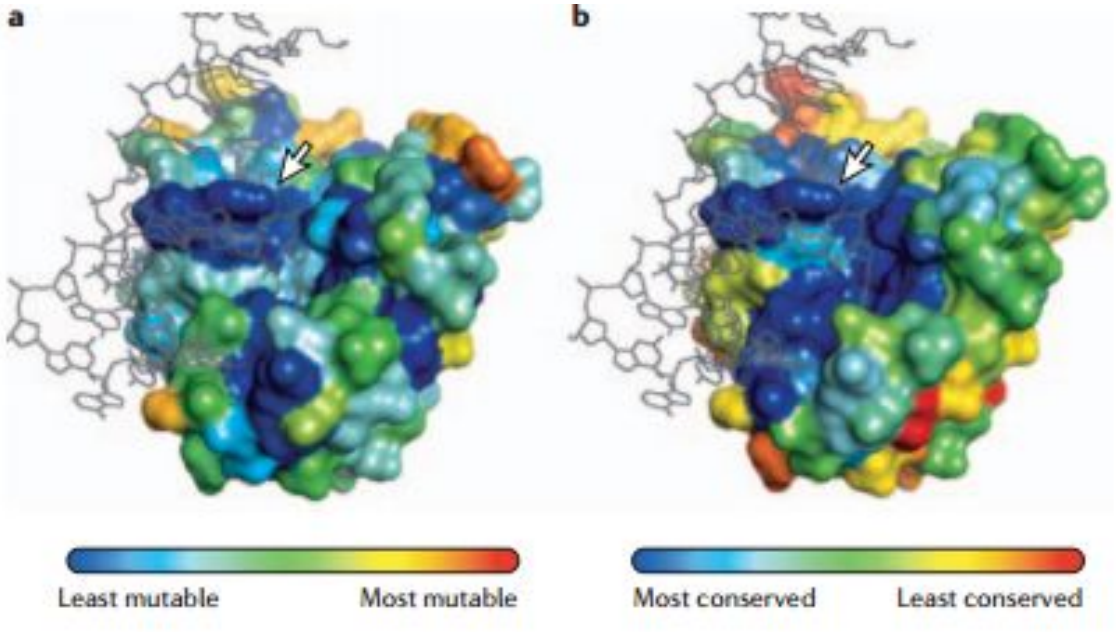
Hh (Hedgehog) — Gli/Ci

TGF-β и RTK пути имеют различные активаторы и репрессоры, для которых может быть один и тот же сайт связывания с транскрипционным фактором (второй тип, B), либо разные сайты для активатора и репрессора (третий тип, C)

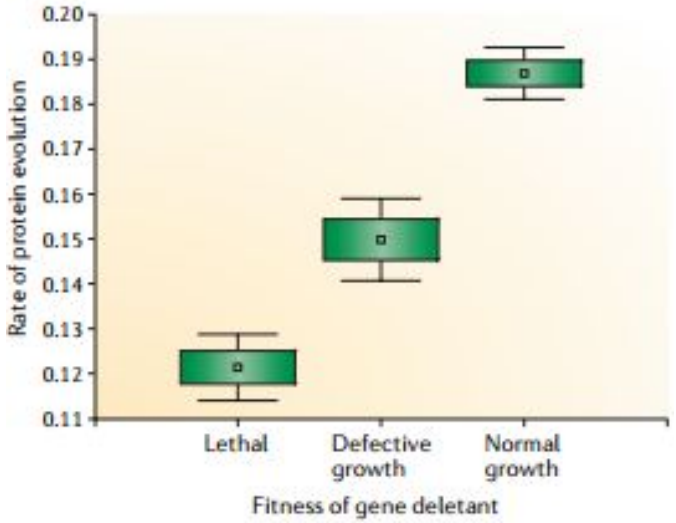


Hecht et al., 2013.

Ландшафт мутабильности адренергического рецептора ADRB2 человека. Смоделированы эффекты от замены аминокислоты «дикого типа» на все остальные (красный — структура и функция меняется, зеленый — нейтральность)



(a) — Распределение на структуре белка остатков, изменение которых с большей вероятностью (краснее) изменит функцию, и (b) — консервативность остатков среди организмов (по базе данных)



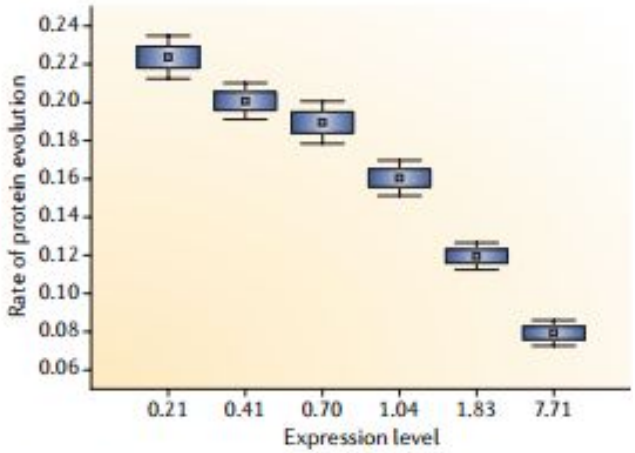
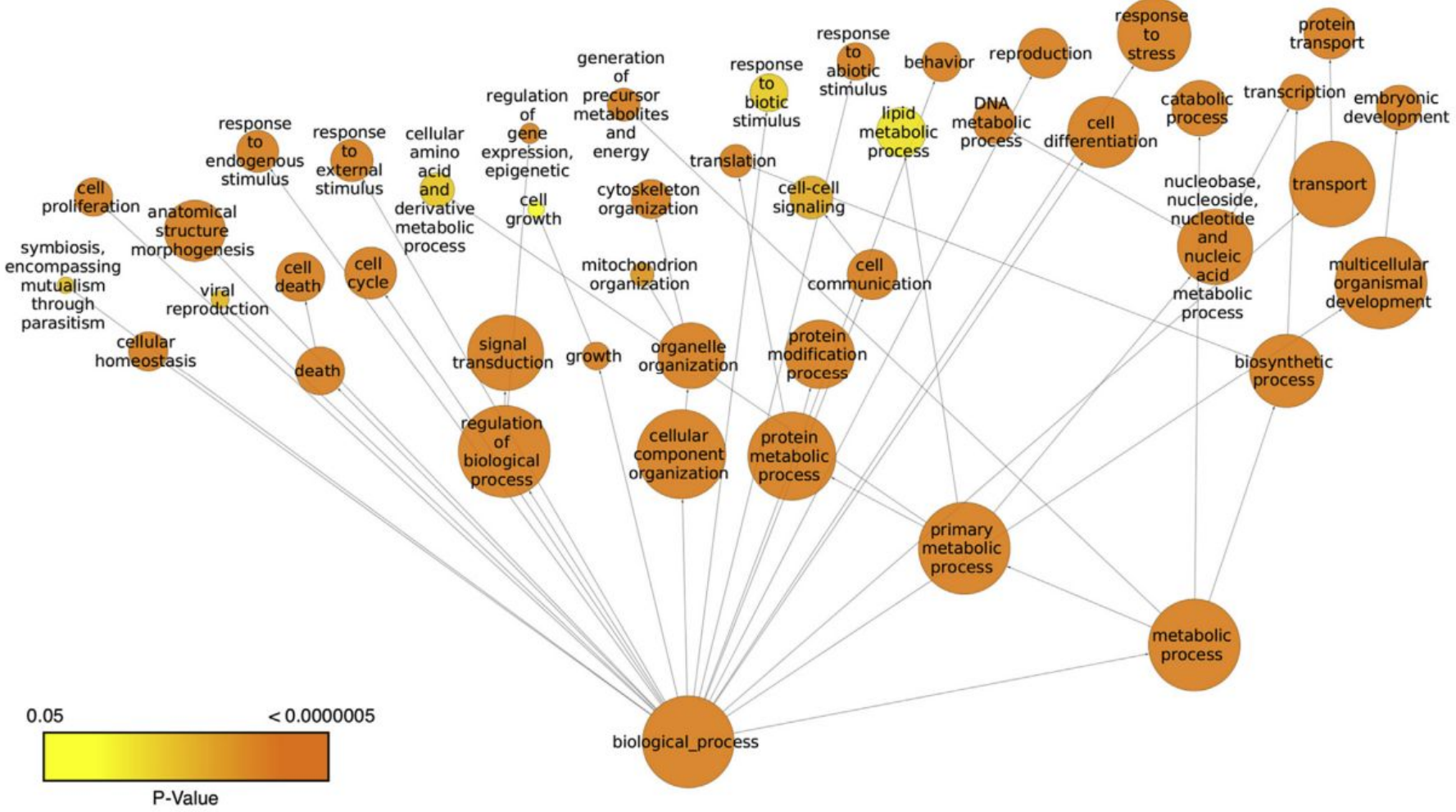
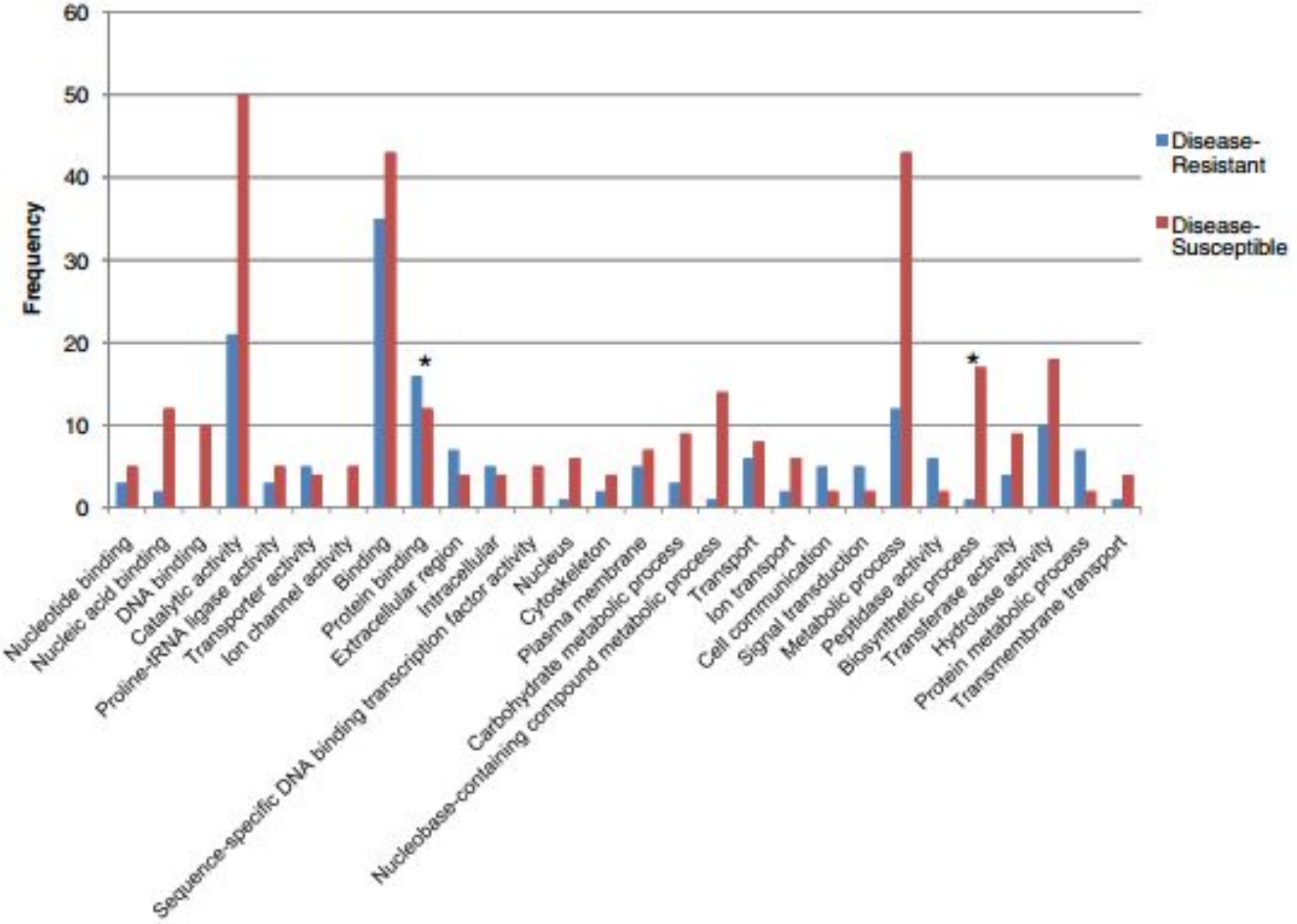


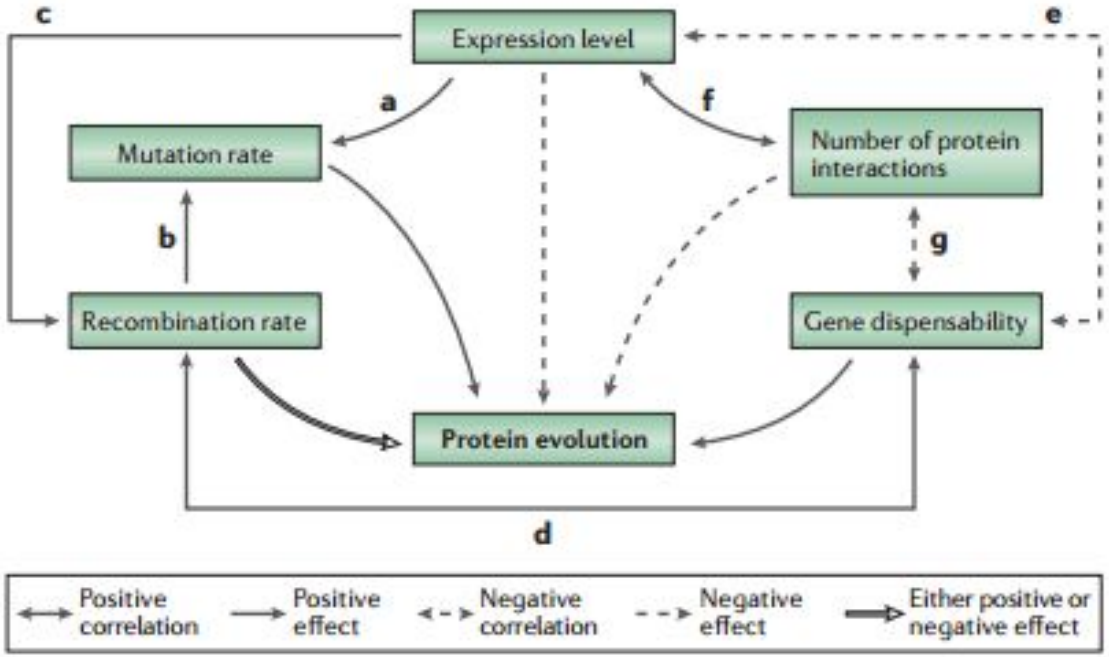
Figure 3 | **Gene-expression level and rate of protein evolution.** Gene-expression level (measured as mRNA abundance on a rich medium[137]) correlates strongly and negatively with the rate of protein evolution in yeast (R^2 = 0.29 for individual genes). Evolutionary rate (non-synonymous divergence) was calculated by Wall *et al.*[47] using sequences from four species of the *Saccharomyces* genus. The same number of genes was assigned to each bin. Boxes show mean ± standard error.



GO биологических процессов белков, вступающих в большое число белок-белковых взаимодействий



Yates, Sternberg, 2013.
Некоторые белковые домены более устойчивы к несинонимичным заменам (т.е. в данном случае реже проявляется клинический эффект), чем другие



(a) Транскрипция увеличивает долю спонтанных мутаций (показано на E. coli и S. cerevisiae)

(b) Рекомбинационная репарация двунитевых разрывов увеличивает частоту точечных мутаций

(c) Гены, которые у S. cerevisiae ближе к рекомбинационным точкам экспрессируются сильнее, чем большинство других

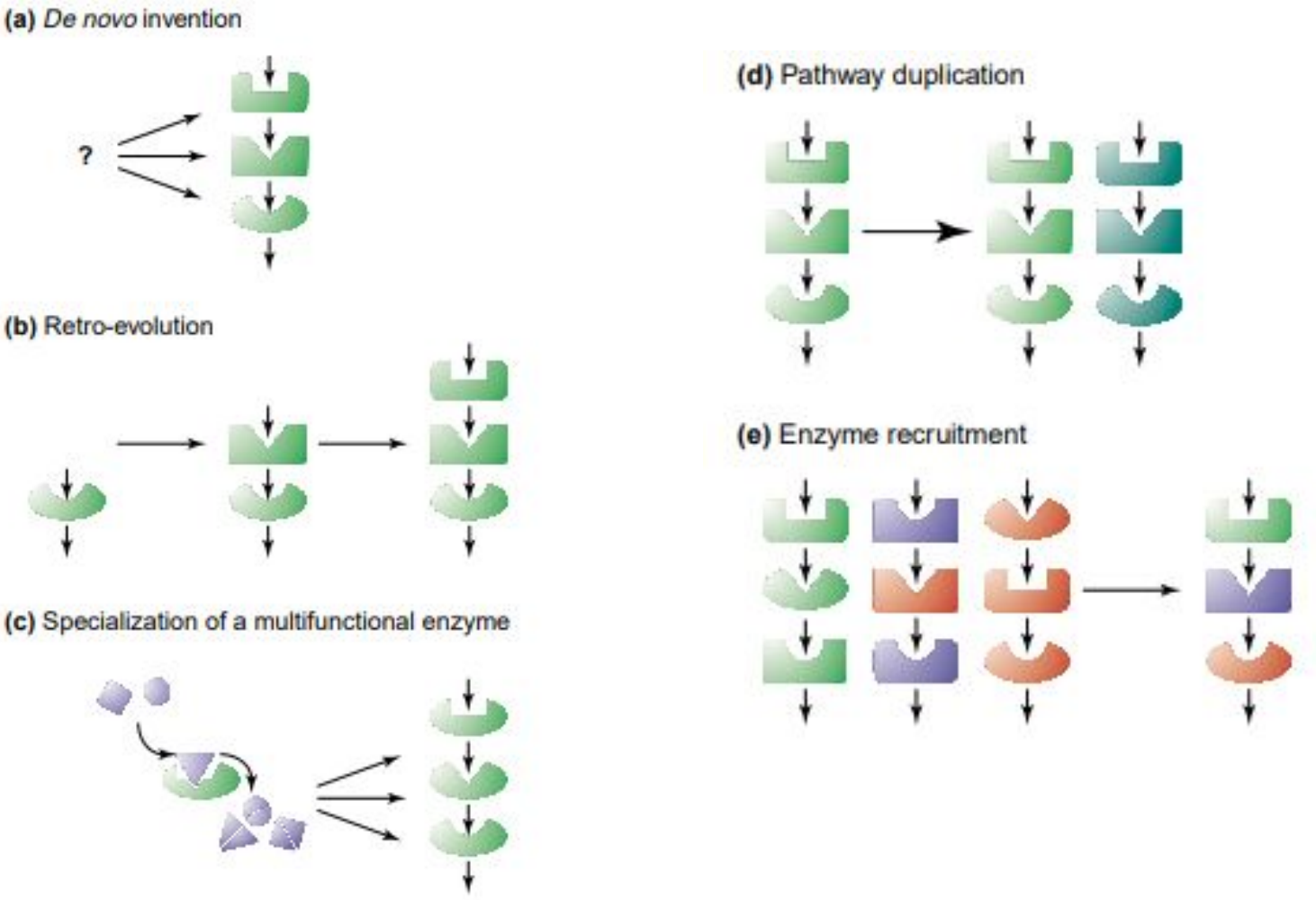
(d) Важные гены сконцентрированы в регионах с низкой рекомбинацией (показано на S. cerevisiae и C. elegans)

(e) Менее важные гены чаще экспрессированы на более низком уровне, чем более важные

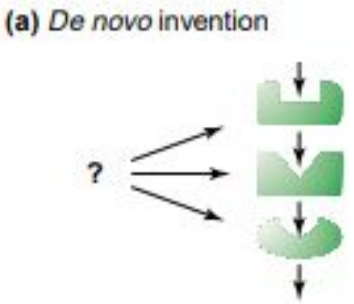
(f) Более высоко экспрессированные белки обладают большим числом белков, с которыми они вступают во взаимодействие (на S. cerevisiae, не подтверждено некоторыми методами)

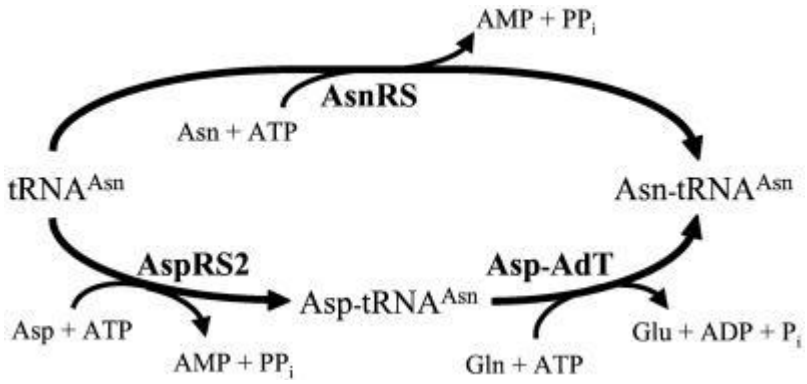
(g) У более важных генов в среднем больше взаимодействий с другими.

Пути эволюции метаболических цепей

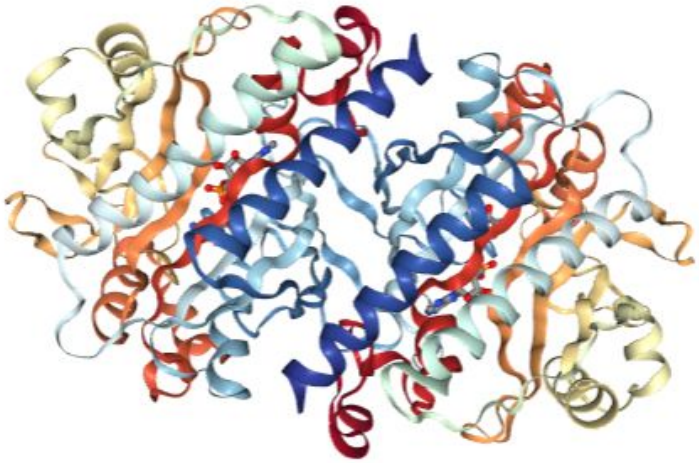


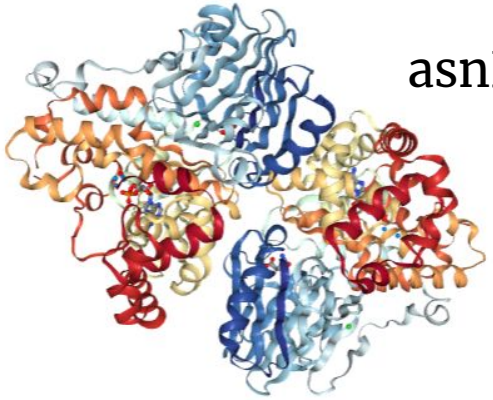
Пути эволюции метаболических цепей





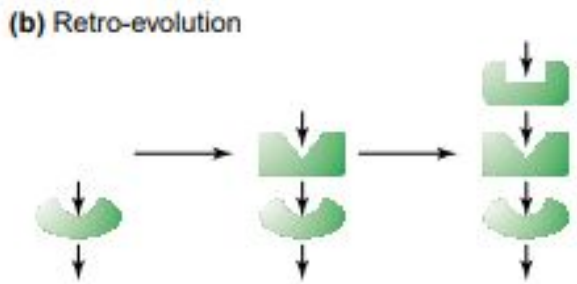
asnA



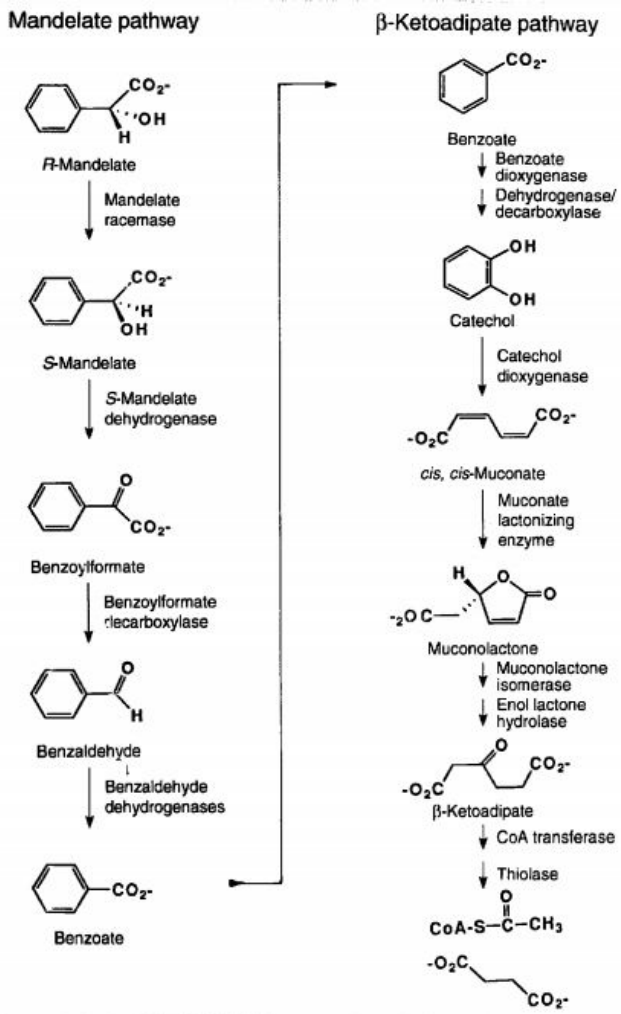


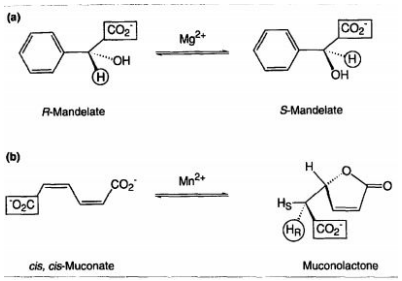
asnB

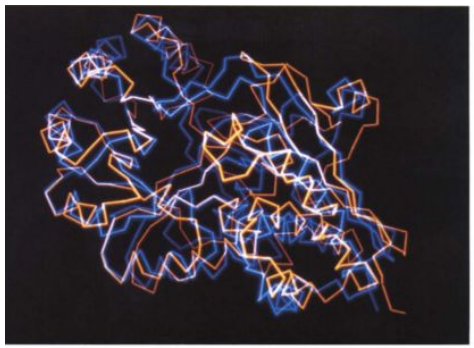
Пути эволюции метаболических цепей

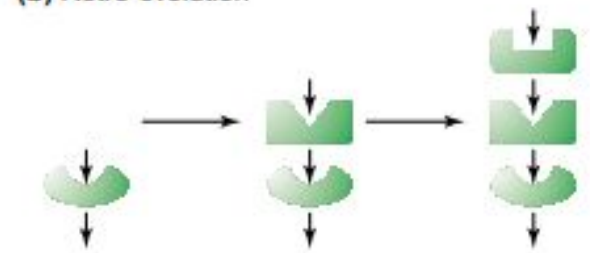


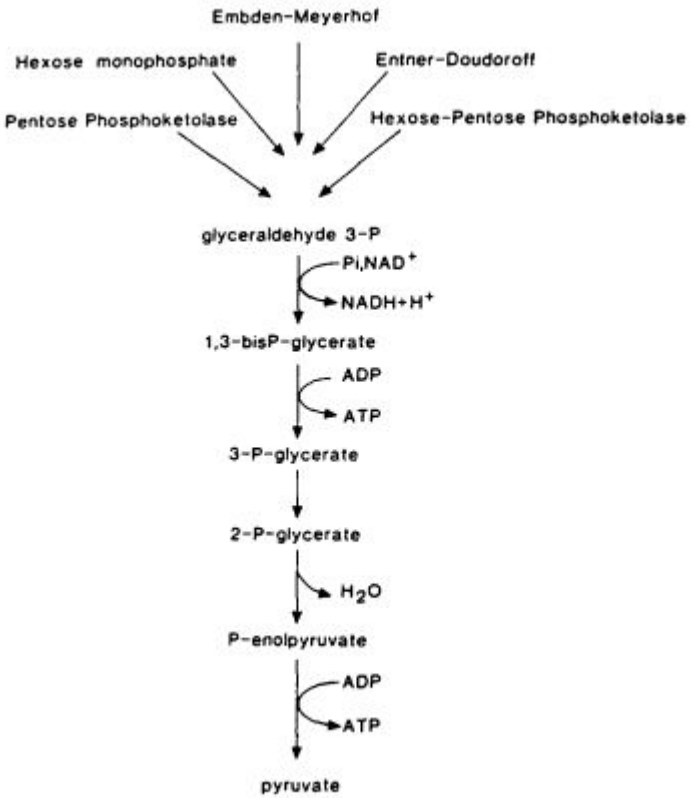
Модель «ретро-эволюции» предполагает, что отбор действует, в основном, на выход конечного продукта и на «достраивание» цепи ферментов для увеличения возможности синтезировать


Mandelate pathway
β-Ketoadipate pathway
R-Mandelate
Mandelate racemase
S-Mandelate
S-Mandelate dehydrogenase
Benzoylformate
Benzoylformate decarboxylase
Benzaldehyde
Benzaldehyde dehydrogenases
Benzoate
Benzoate
Benzoate dioxygenase
Dehydrogenase/ decarboxylase
Catechol
Catechol dioxygenase
cis, cis-Muconate
Muconate lactonizing enzyme
Muconolactone
Muconolactone isomerase
Enol lactone hydrolase
β-Ketoadipate
CoA transferase
Thiolase

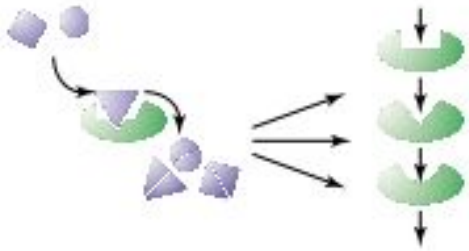

(a)
Mg2+
R-Mandelate
S-Mandelate
(b)
Mn2+
cis, cis-Muconate
Muconolactone



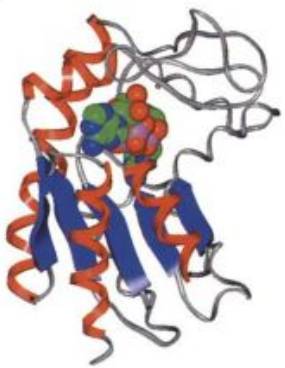

(b) Retro-evolution


Embden-Meyerhof
Hexose monophosphate
Entner-Doudoroff
Pentose Phosphoketolase
Hexose-Pentose Phosphoketolase
glyceraldehyde 3-P
Pi,NAD+
NADH+H+
1,3-bisP-glycerate
ADP
ATP
3-P-glycerate
2-P-glycerate
H2O
P-enolpyruvate
ADP
ATP
pyruvate

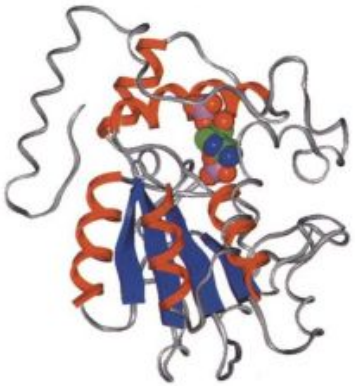
(c) Specialization of a multifunctional enzyme



(a)



Adenylate kinase

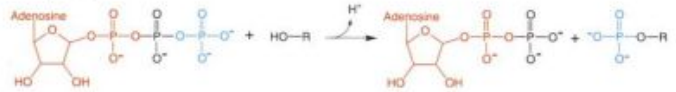


Estrogen sulfotransferase

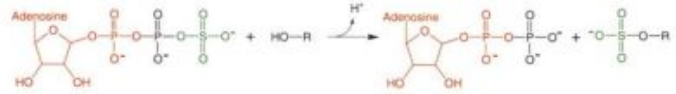
(b)

Adenylate kinase

Phosphoryl transfer

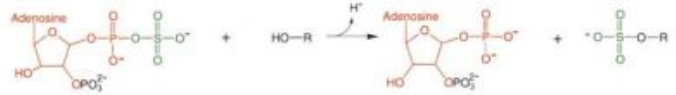


Promiscuous sulfuryl transfer



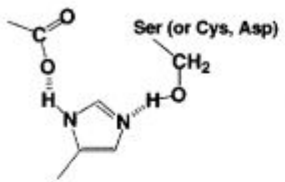
Estrogen sulfotransferase

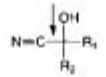
Sulfuryl transfer



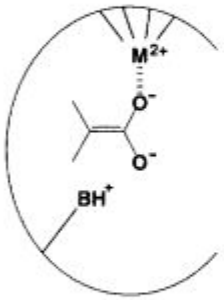
Chemistry & Biology

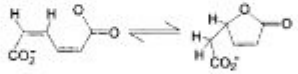
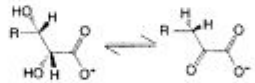
(a)



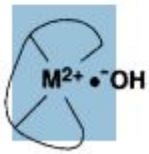
Enzyme	Substrate
Acetyl-cholinesterase	
Hydroxynitrile lyase	
Semialdehyde hydrolase	
Haloalkane dehalogenase	

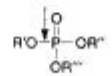
(b)



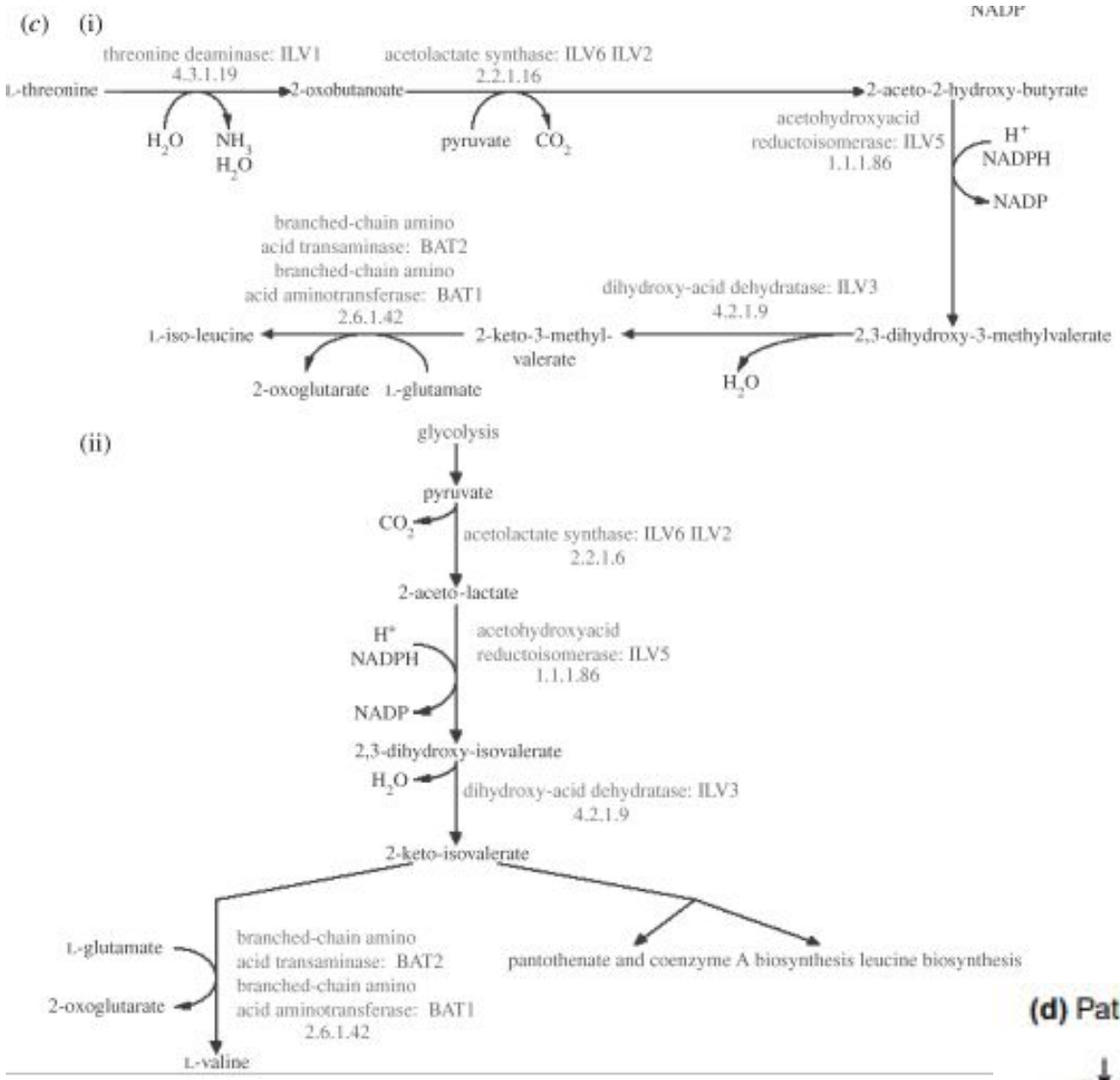
Enolase	
Mandelate racemase	
Muconate lactonizing enzyme	
Galactonate dehydratase	

(c)



Urease	
Phosphotriesterase	
Adenosinedeaminase	
Chlorohydrolase	

Chemistry & Biology


(c) (i)
NADP
threonine deaminase: ILV1
4.3.1.19
acetolactate synthase: ILV6 ILV2
2.2.1.16
L-threonine
2-oxobutanoate
2-aceto-2-hydroxy-butyrate
H2O
NH3
H2O
pyruvate
CO2
acetohydroxyacid
reductoisomerase: ILV5
1.1.1.86
H+
NADPH
NADP
branched-chain amino
acid transaminase: BAT2
branched-chain amino
acid aminotransferase: BAT1
2.6.1.42
dihydroxy-acid dehydratase: ILV3
4.2.1.9
L-iso-leucine
2-keto-3-methyl-
valerate
2,3-dihydroxy-3-methylvalerate
2-oxoglutarate
L-glutamate
H2O
(ii)
glycolysis
pyruvate
CO2
acetolactate synthase: ILV6 ILV2
2.2.1.6
2-aceto-lactate
H+
NADPH
acetohydroxyacid
reductoisomerase: ILV5
1.1.1.86
NADP
2,3-dihydroxy-isovalerate
H2O
dihydroxy-acid dehydratase: ILV3
4.2.1.9
2-keto-isovalerate
L-glutamate
branched-chain amino
acid transaminase: BAT2
branched-chain amino
acid aminotransferase: BAT1
2.6.1.42
2-oxoglutarate
L-valine
pantothenate and coenzyme A biosynthesis leucine biosynthesis

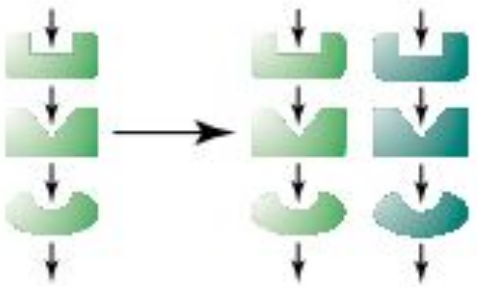

(d) Pathway duplication

Fig. 1. HisA and TrpF catalyze similar reactions in histidine and tryptophan biosynthesis. HisA and TrpF catalyze the isomerizations of the aminoaldoses *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) and phosphoribosylanthranilate (PRA) to the aminoketoses *N*′-[(5′-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR) and 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5-phosphate (CdRP).

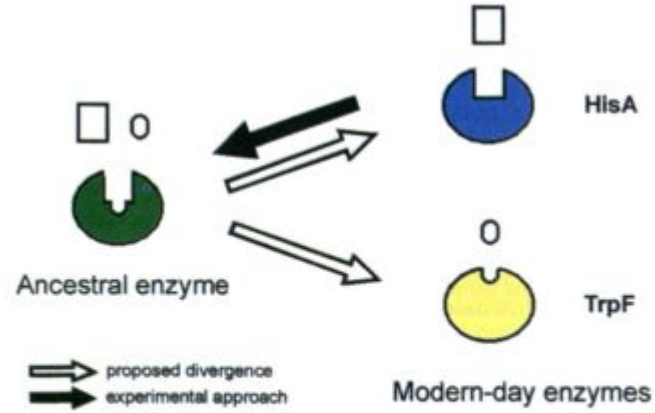
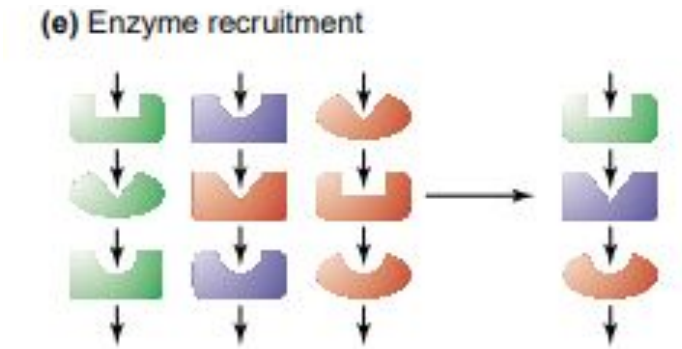
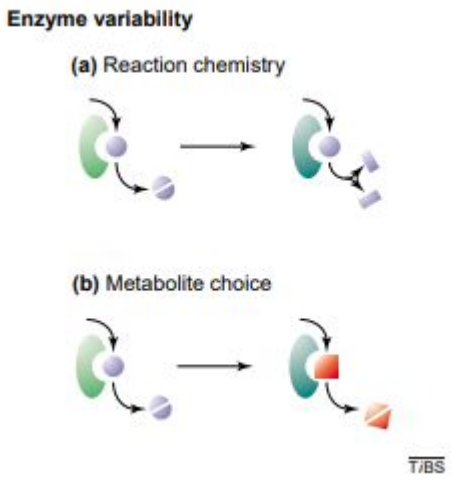
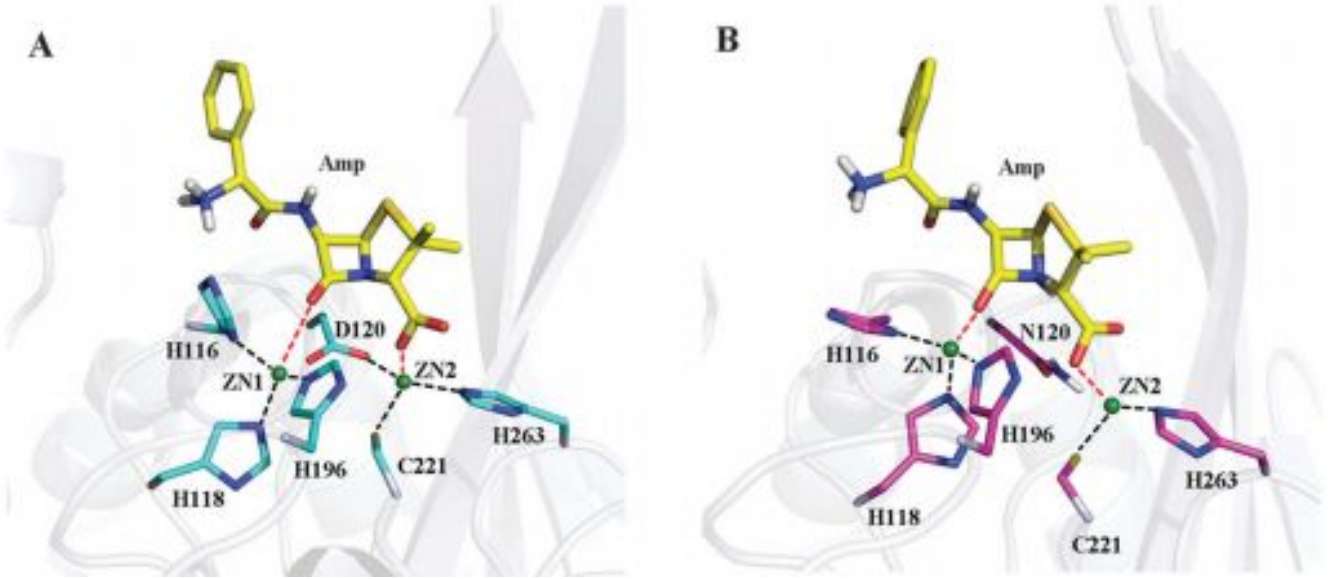


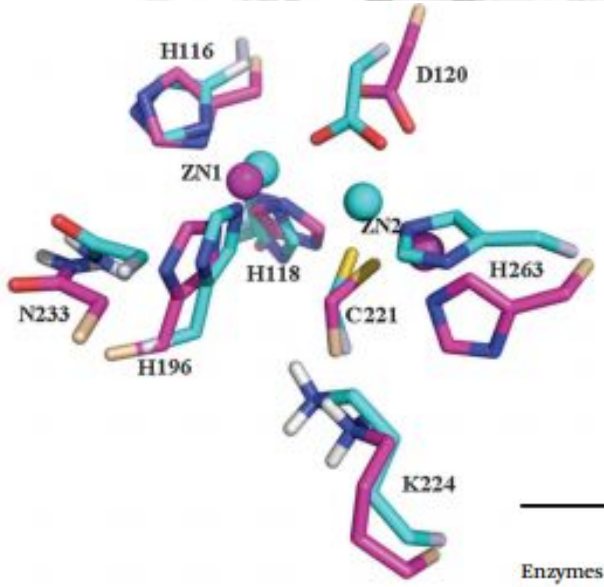
Fig. 2. Experimental approach for testing the patchwork hypothesis (4) of enzyme evolution. Modern-day enzymes such as HisA and TrpF are highly specific catalysts that may have evolved from a common ancestor enzyme that was less specific. Starting from HisA, we tried to reverse the postulated evolutionary path, creating an enzyme capable of catalyzing both the HisA and the TrpF reaction.





Мутации в активном центре

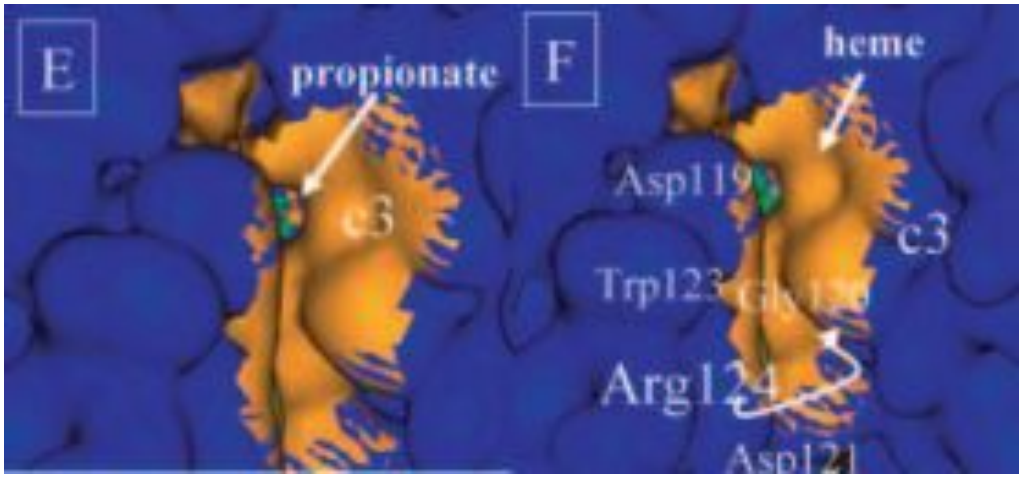


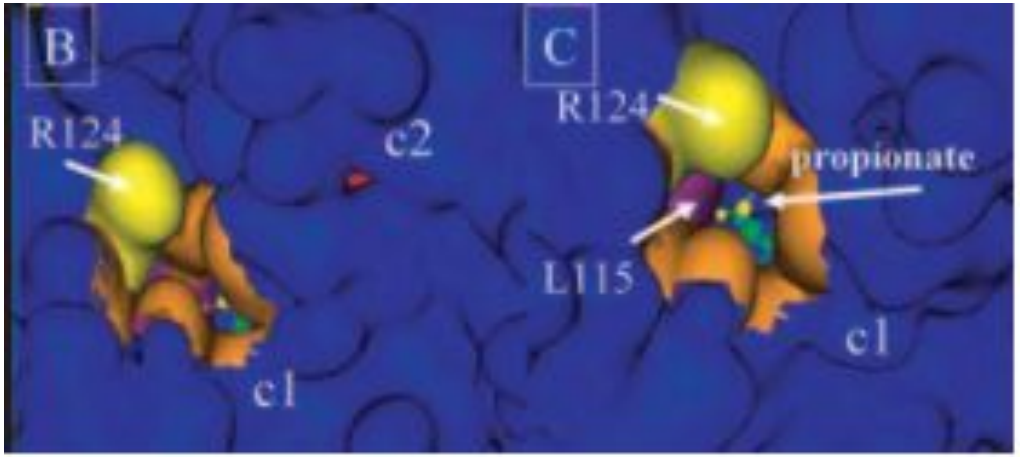


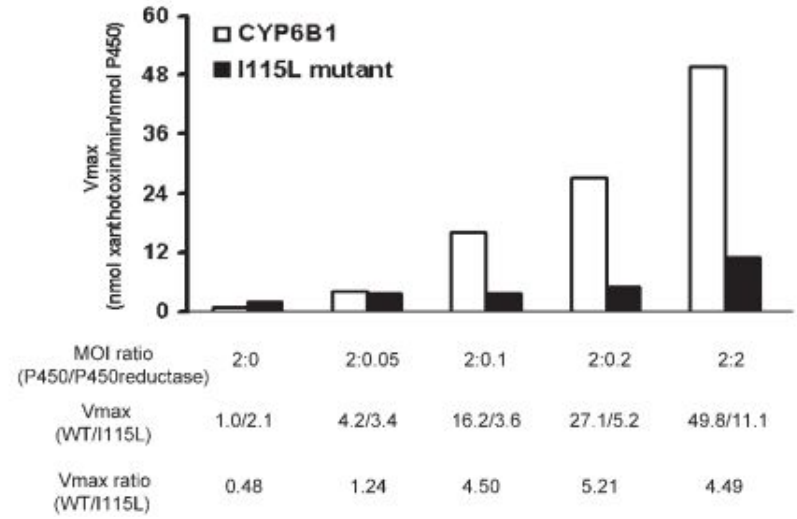
Мутация аспарагиновой кислоты (120) в аспарагин в металло-бета-лактамазе (*E. coli*) привела изменению расстояния и перераспределению заряда между ионами цинка, что привело к невозможности расщепления антибиотиков.

Enzymes harbored in *E. coli* BL21 (DE3)	Antibiotics (μg ml^{-1})				
	Penicillin G	Ampicillin	Cefuroxime	Ceftizoxime	Meropenem
Wild-type NDM-1	512	>512	>512	64	16
D120N mutant	8	4	1	1	<0.5

Мутации, меняющие доступ к активному центру



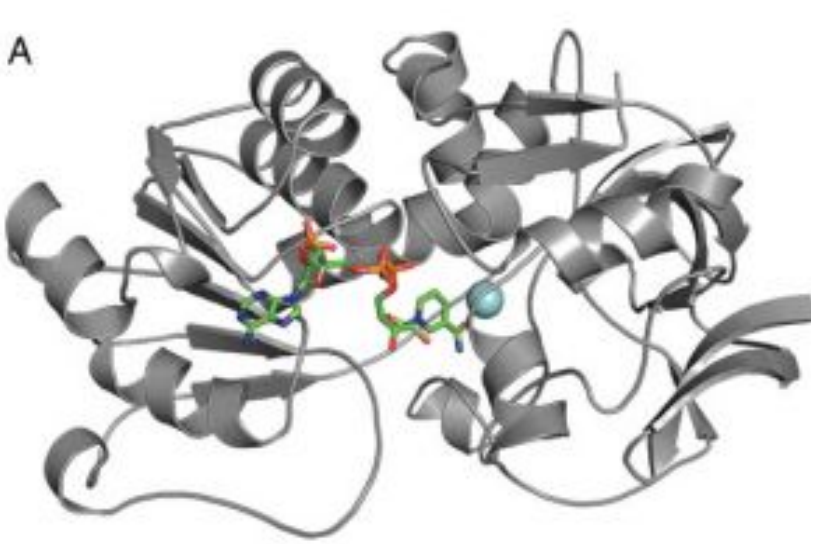


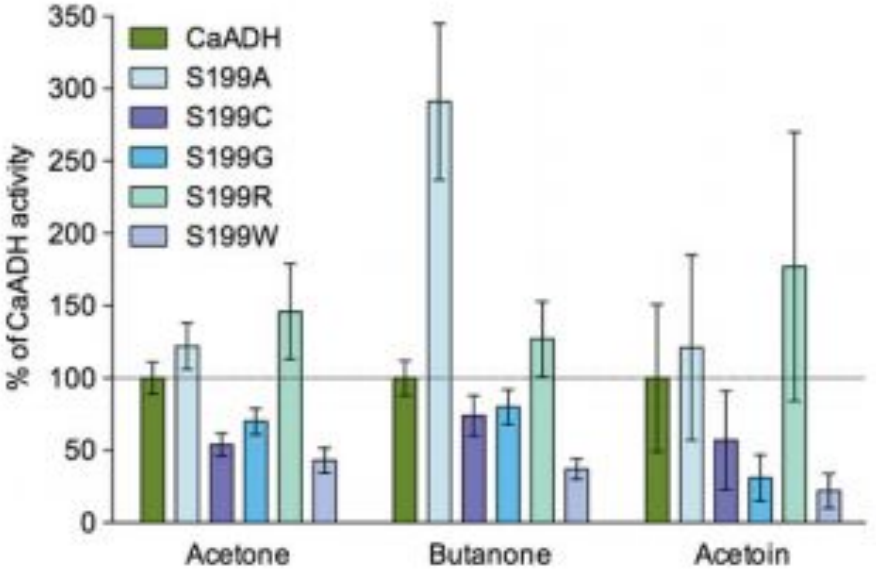


MOI ratio (P450/P450reductase)	2:0	2:0.05	2:0.1	2:0.2	2:2
Vmax (WT/I115L)	1.0/2.1	4.2/3.4	16.2/3.6	27.1/5.2	49.8/11.1
Vmax ratio (WT/I115L)	0.48	1.24	4.50	5.21	4.49

Замена изолейцина на лейцин в CYP6B1 *Papilio polyxenes* ограничила доступ субстрата к активному центру, что привело к невозможности метаболизировать фуранокумарин.

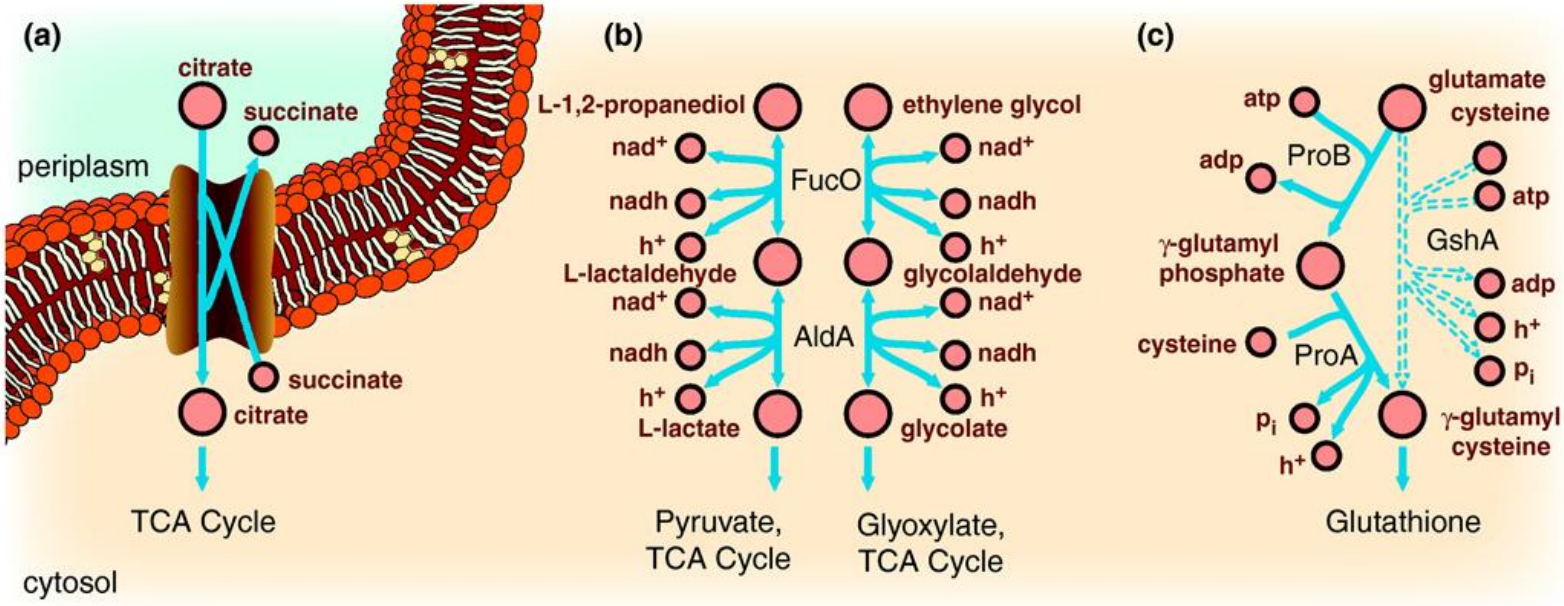
Мутации в сайте связывания с коферментом





Изменение в месте связывания с НАДФН, расположенному в удалении от активного центра, в алкогольдегидрогеназе *Clostridium autoethanogenum* привело к изменению в специфичности к субстрату. В некоторых случаях менялась специфичность к коферменту (НАДН вместо НАДФН)

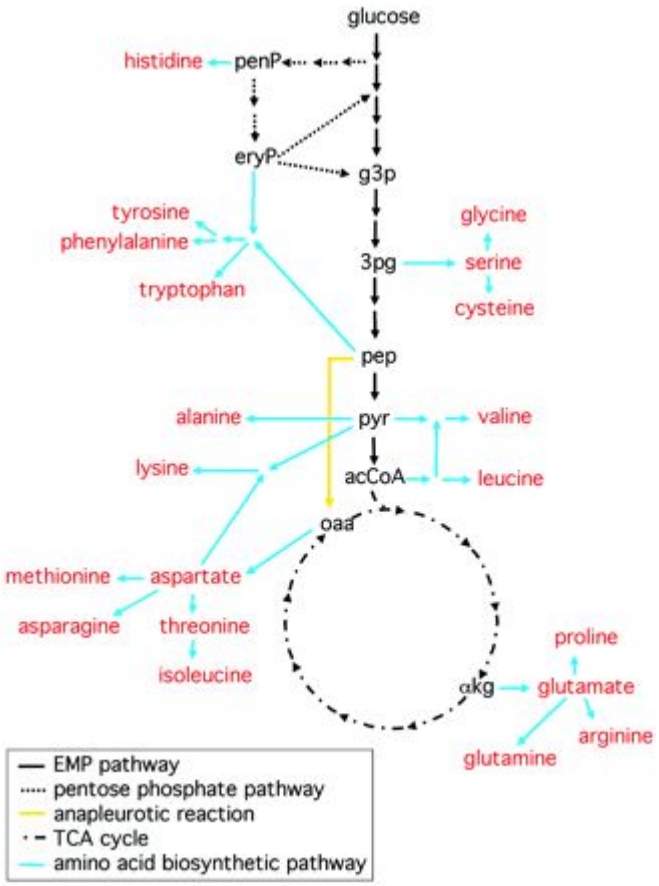
Экспериментальное изучение эволюции метаболических путей



Extension of metabolic pathways through laboratory evolution

Laboratory evolution under a defined selective pressure has identified a few cases in which a new metabolic function arose. Three of these examples include the ability that *E. coli* gained to (a) transport citrate after 33,000 generations, (b) metabolize L-1,2-propanediol and ethylene glycol, and (c) synthesize glutathione when a key enzyme in its synthesis was deleted.

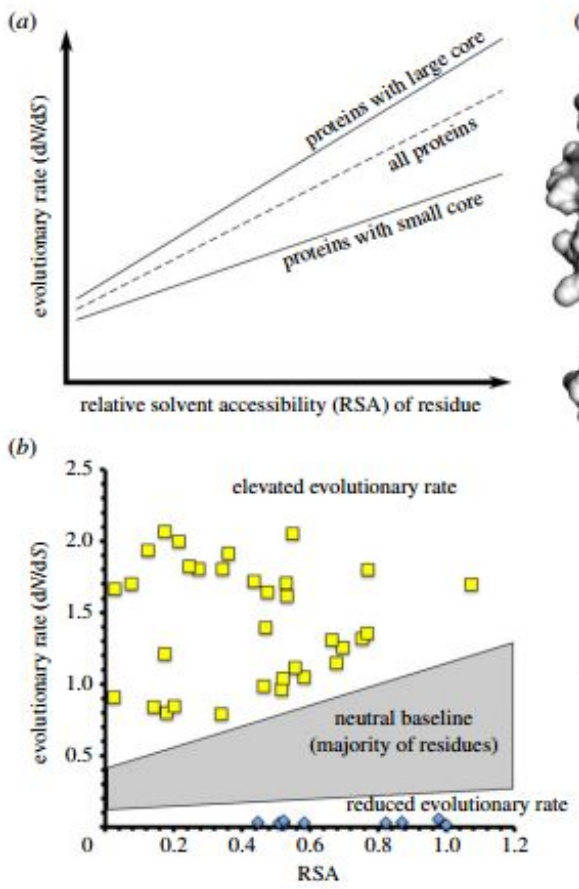
Цена за аминокислоту

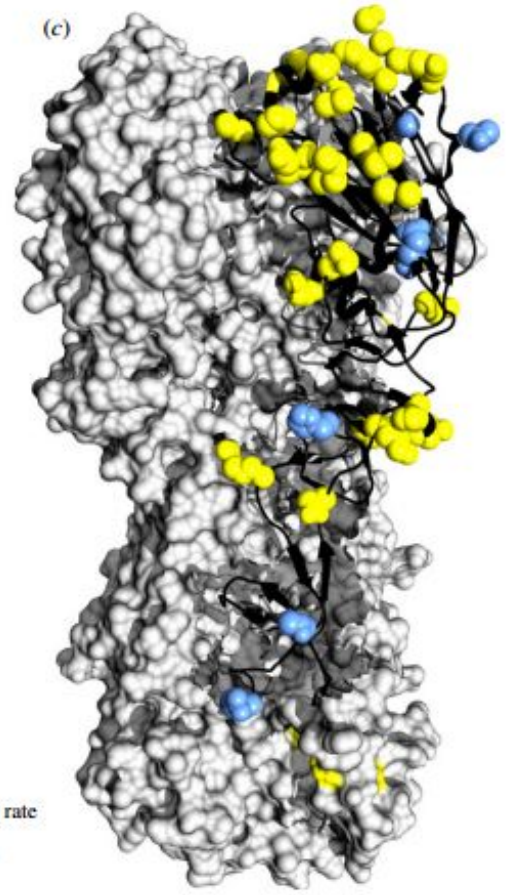


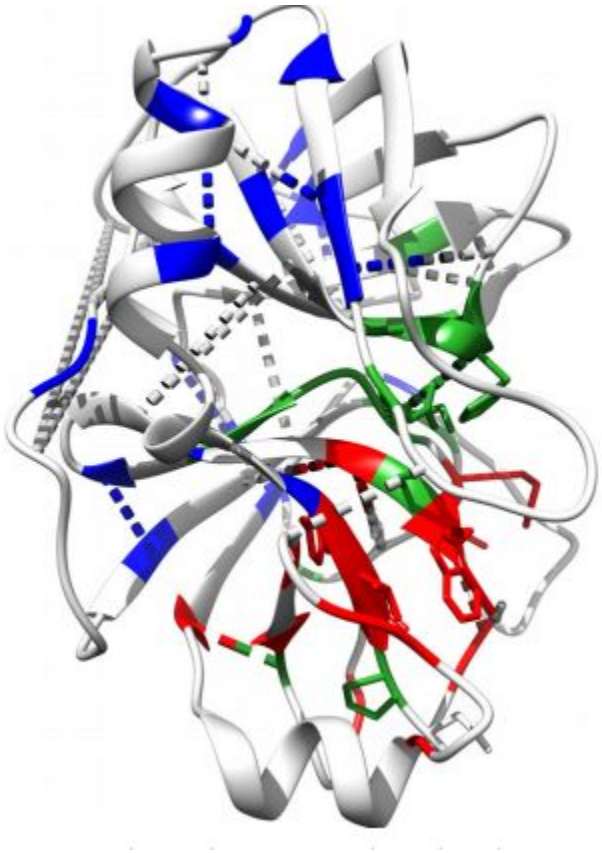
Amino acid	One-letter symbol	Precursor metabolites	Energetic cost		
			~P	H	Total, ~P
Ala	A	pyr	1.0	5.3	11.7
Cys	C	3pg	7.3	8.7	24.7
Asp	D	oaa	1.3	5.7	12.7
Glu	E	αkg	2.7	6.3	15.3
Phe	F	2 pep, eryP	13.3	19.3	52.0
Gly	G	3pg	2.3	4.7	11.7
His	H	penP	20.3	9.0	38.3
Ile	I	pyr, oaa	4.3	14.0	32.3
Lys	K	oaa, pyr	4.3	13.0	30.3
Leu	L	2 pyr, acCoA	2.7	12.3	27.3
Met	M	oaa, Cys, -pyr	9.7	12.3	34.3
Asn	N	oaa	3.3	5.7	14.7
Pro	P	αkg	3.7	8.3	20.3
Gln	Q	αkg	3.7	6.3	16.3
Arg	R	αkg	10.7	8.3	27.3
Ser	S	3pg	2.3	4.7	11.7
Thr	T	oaa	3.3	7.7	18.7
Val	V	2 pyr	2.0	10.7	23.3
Trp	W	2 pep, eryP, P RPP, -pyr	27.7	23.3	74.3
Tyr	Y	eryP, 2 pep	13.3	18.3	50.0

У высоко экспрессированных белков E. coli отбираются те аминокислоты, чья энергетическая цена синтеза меньше

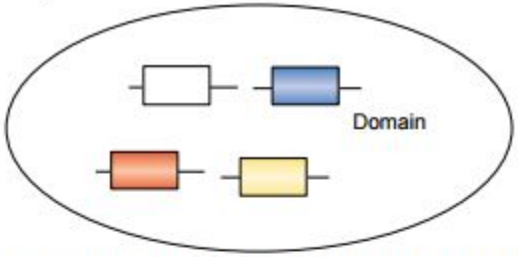
Структура белка и его эволюция



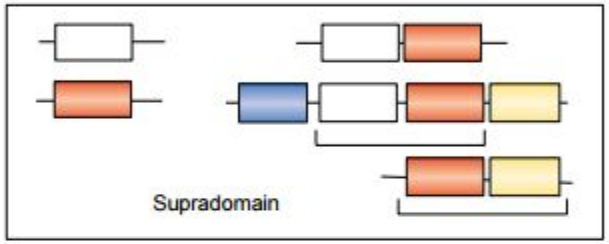




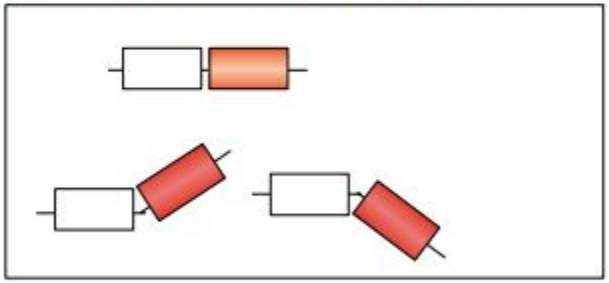
The repertoire of domain superfamilies...



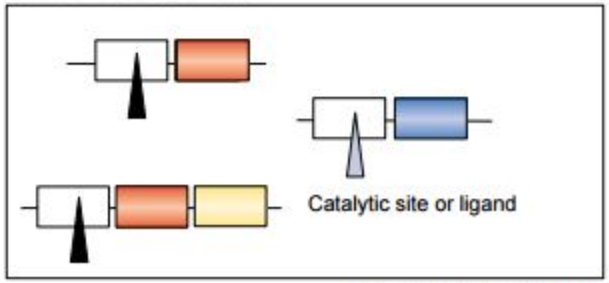
...duplicates and recombines to form single and multi-domain proteins.



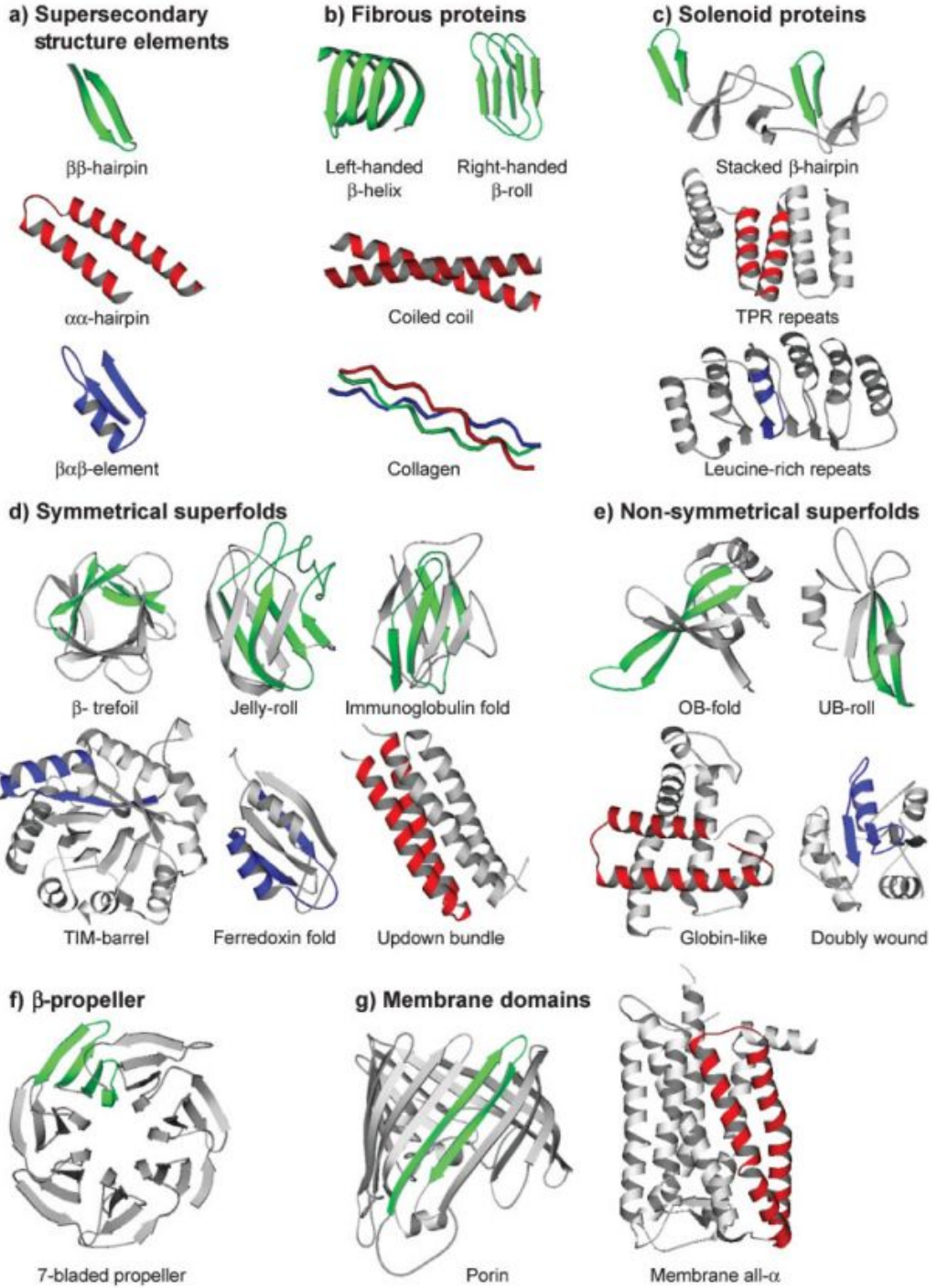
The same combination can adopt different geometries...



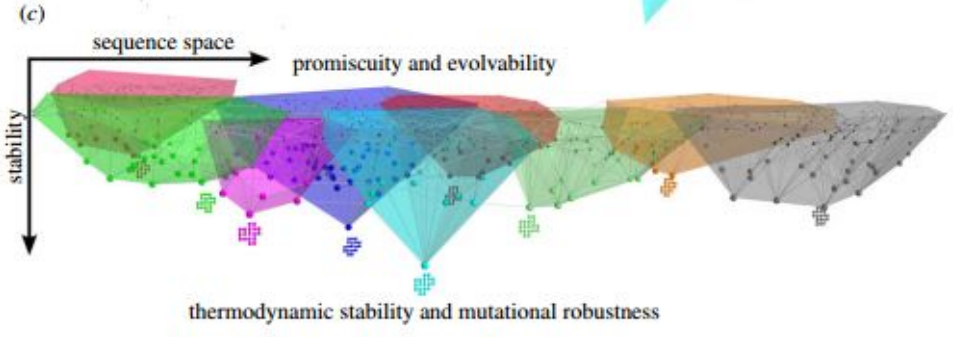
...and/or different functions.

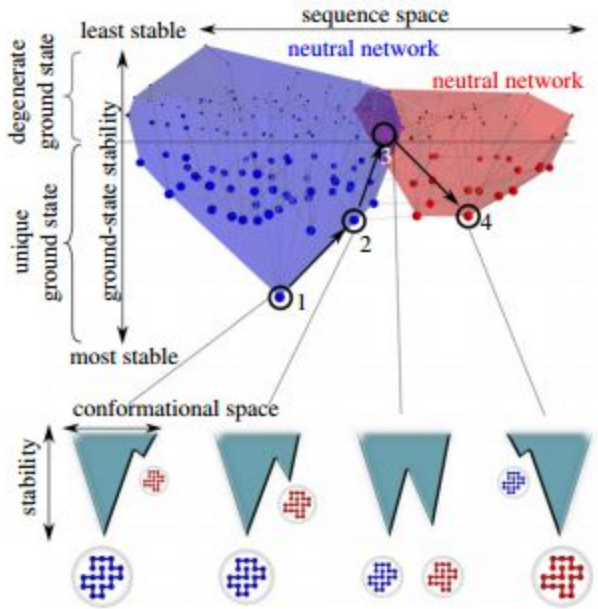


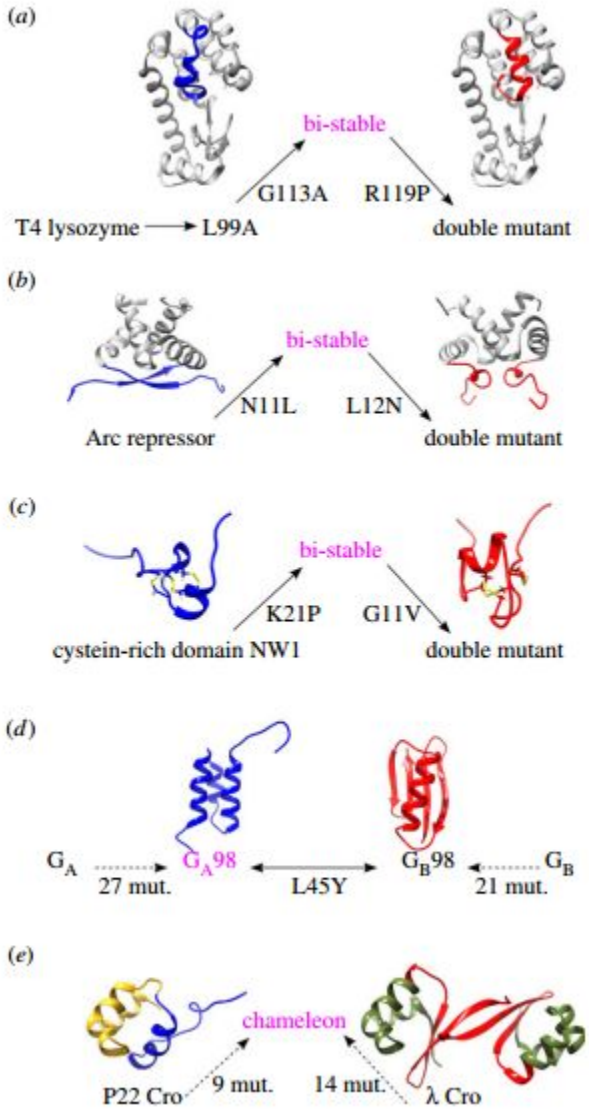
Current Opinion in Structural Biology



Структура белка и его эволюция







Variable	Surrogate	Comments
Expression levels	Codon usage bias	It measures translation levels indirectly, usually requires knowledge of highly expressed genes or optimal codons, and significant codon usage bias; strong association [2,5,8].
	mRNA abundance	Noisy data; strong association [4,6]
	Protein cellular concentration	Little data available; unknown accuracy; strong association [13]
Expression breadth	Gene expression in different tissues	Only applicable to multicellular differentiated organisms; noisy data (EST or microarrays); strong association [16], which vanishes when expression levels are controlled for [15]
Essentiality	Absence of growth after knockout	Some methodologies are error-prone (e.g. transposon mutagenesis); it only measures growth in one-by-one knockouts and in one set of nearly optimal conditions; either no association [17,18] or it vanishes when expression levels are controlled for [5]
Dispensability	Decrease in growth rate after knockout	The same experimental problems as essentiality; either weak association or the association vanishes when expression levels are controlled for [13,19–21]
Density of contact functions	Connectiveness in protein–protein interactions network (PPIN)	Noisy data except in the rare and smaller curated data sets; the association is weak in yeasts [22–24] and becomes even weaker when expression levels are controlled for [13], suggesting it could be artifactual [25]; in *H. pylori*, the association in not significant [26]
	Closeness in PPIN	Weakly correlated ($\rho < 0.16$) even before expression levels are controlled for [24]
	Betweenness in PPIN	Weakly correlated ($\rho < 0.18$) even before expression levels are controlled for [24]
	Temporal connectivity of hubs in PPIN	Proteins that are party hubs (i.e. have many simultaneous connections) are more conserved than other hubs and even more than the generality of proteins; strong effect that concerns only the few proteins that are hubs [27]
Cost of biosynthesis	Cost of amino acid biosynthesis	Metabolic pathways vary and several organisms import rather than produce many of their amino acids, which complicates the computation of amino acid cost; the cost correlates negatively with expression [28] but does not correlate with substitution rates in bacteria [5] nor in *Chlamydomonas* [7]
	Protein length	Although smaller proteins evolve slower [15,29], the effect is weak and likely to disappear when expression is controlled for, because smaller proteins are also more highly expressed [30]
Functional category	Protein families, functional ontologies, localization	How should biologically pertinent categories be defined and delimited? Some surrogates correlate [17], others do not [5]; proteins implicated in multiple processes evolve slower, although the effect is extremely weak (<1% of variance explained) even without controlling for expression levels [31]
Modularity	A mix of PPIN, co-expression and comparative genomics	When PPIN party hubs are defined as intramodule hubs and date hubs as intermodule hubs (see above), the former evolve significantly slower that the latter [27]; when modularity is defined using a mixture of different variables including direct (physical) and non-direct functional interactions, there is a co-association of evolutionary rates within modules, which is weak but significant when controlled for expression levels and is independent of directly connecting pairs in PPIN [32]
Intronic structure	Number and length of introns	Exonic splicing enhancers lower substitution rates in exons [33]; although highly expressed genes have smaller introns, the number of introns is not significantly different [11]; the effect in protein conservation is unknown but if it was predominant dS should also be high and dN/dS should not increase with dN as it does; it is only applicable to genomes with significant numbers of introns