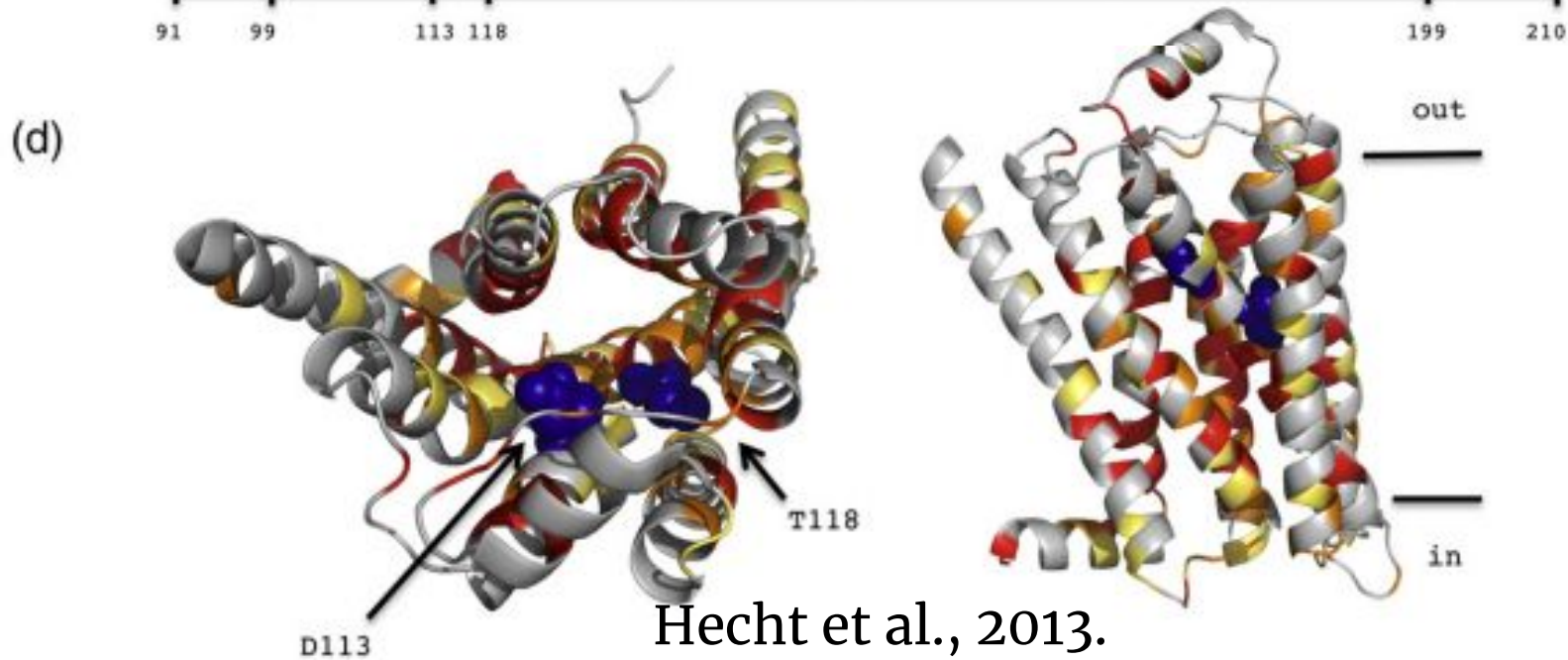
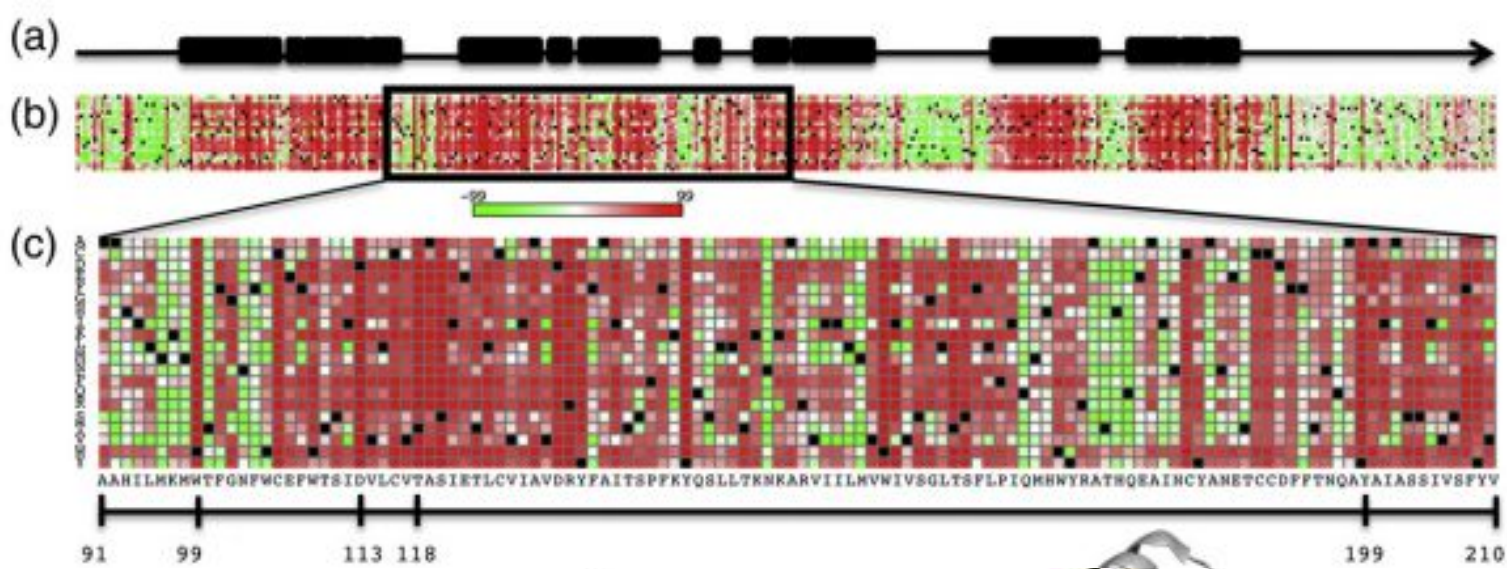


# Особенности молекулярной ЭВОЛЮЦИИ

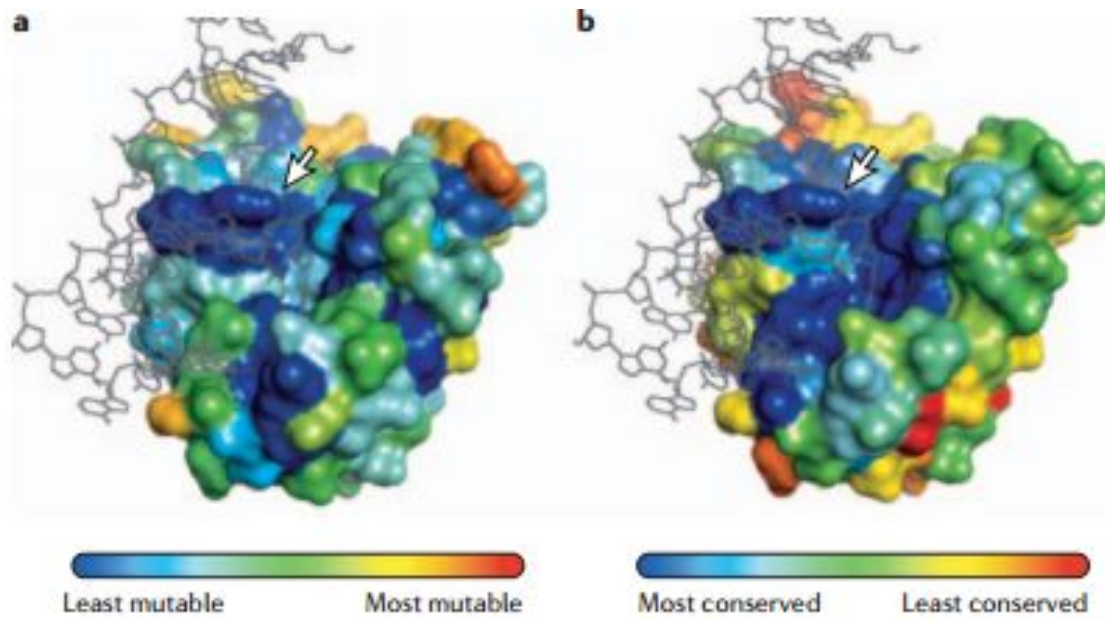
2. Функционально менее важные молекулы или их части эволюционируют (накапливая эволюционные замены) быстрее, чем более важные

3. Мутационные замены, приводящие к меньшим нарушениям структуры и функции молекулы (консервативные замены), в ходе эволюции происходят чаще тех, которые вызывают существенное нарушение структуры и функции этой



Necht et al., 2013.

Ландшафт мутабельности 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)

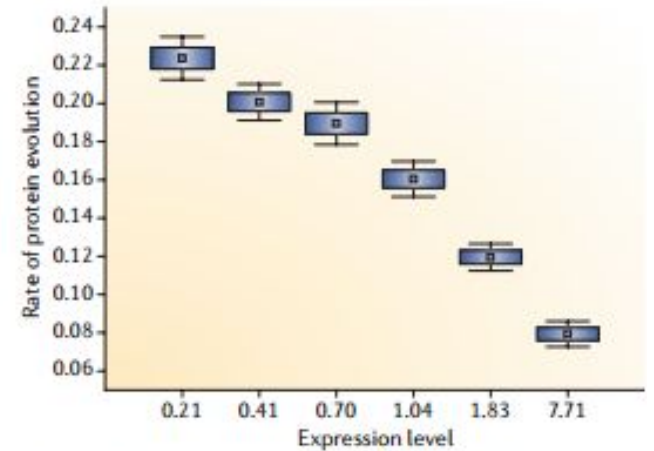
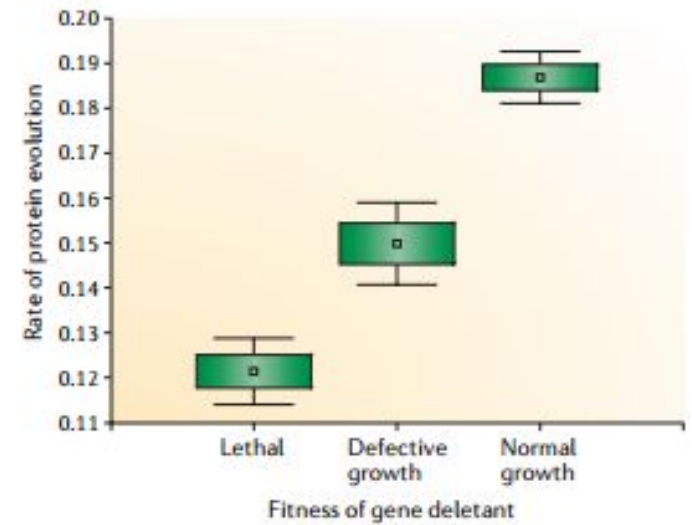
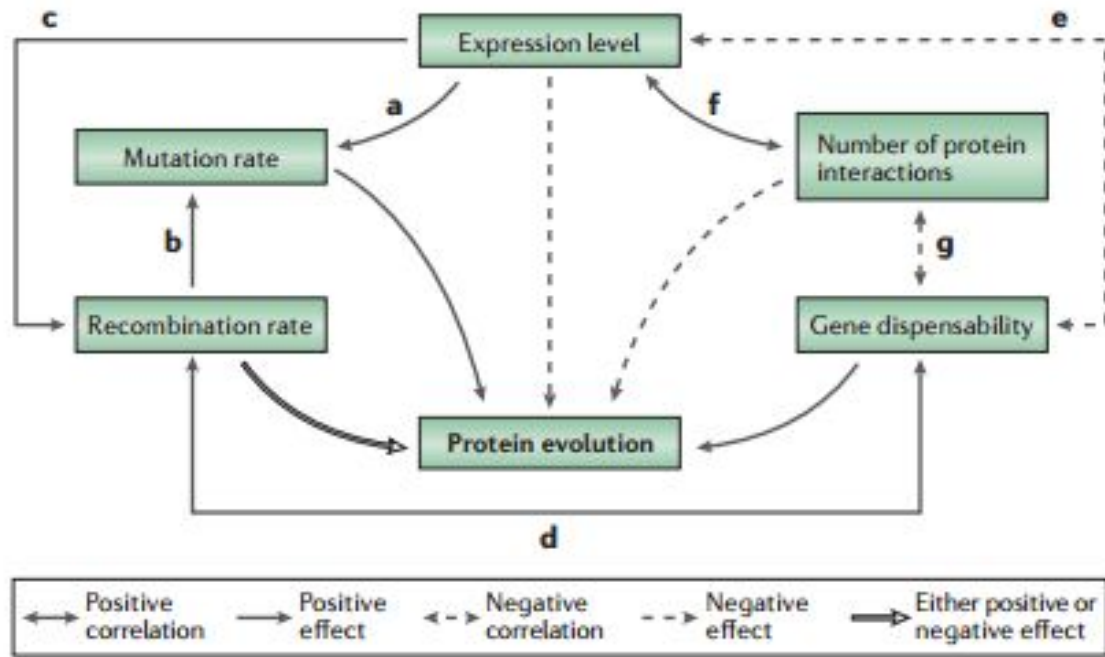
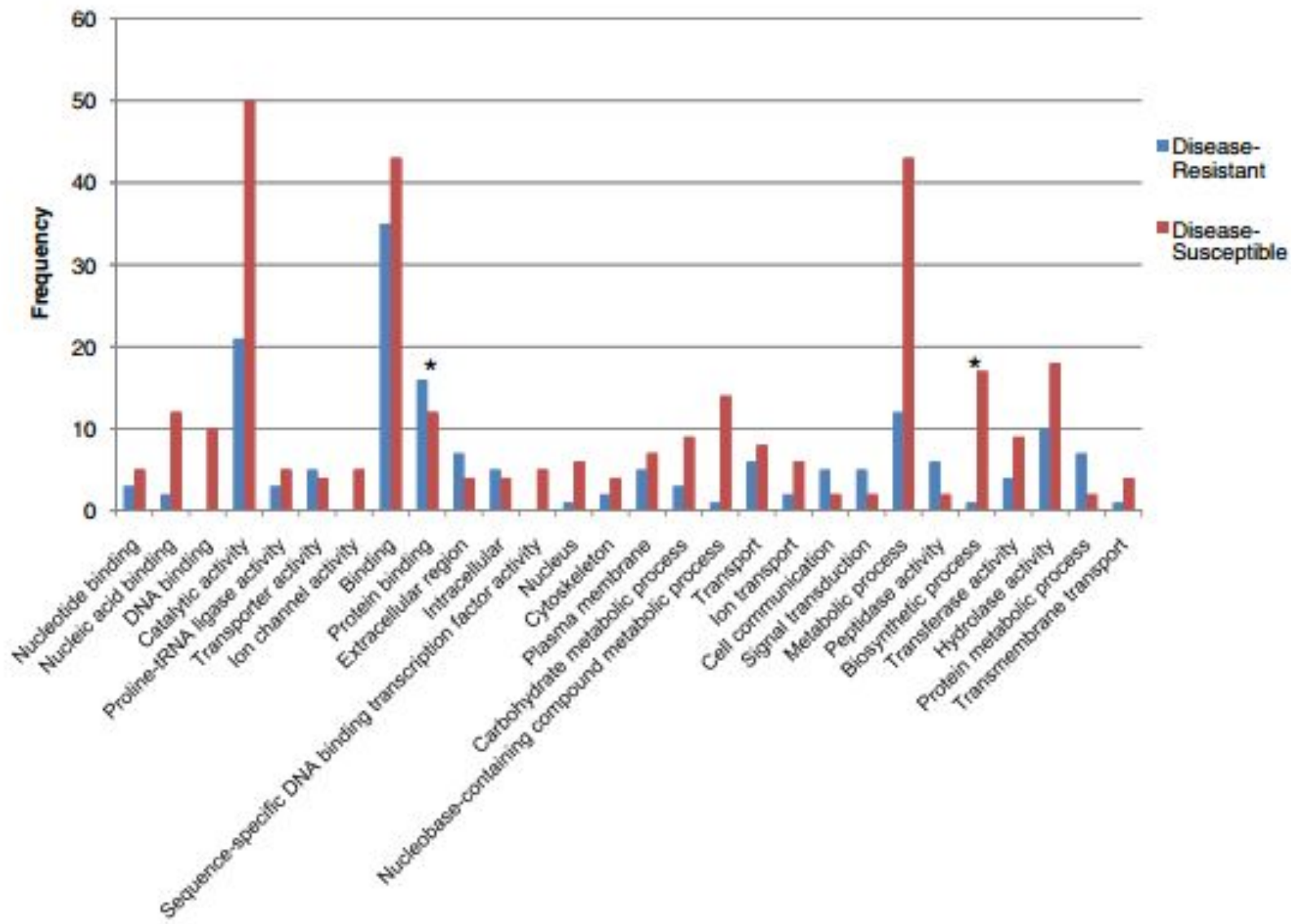


Figure 3 | **Gene-expression level and rate of protein evolution.** Gene-expression level (measured as mRNA abundance on a rich medium<sup>137</sup>) correlates strongly and negatively with the rate of protein evolution in yeast ( $R^2 = 0.29$  for individual genes). Evolutionary rate (non-synonymous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



- (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.

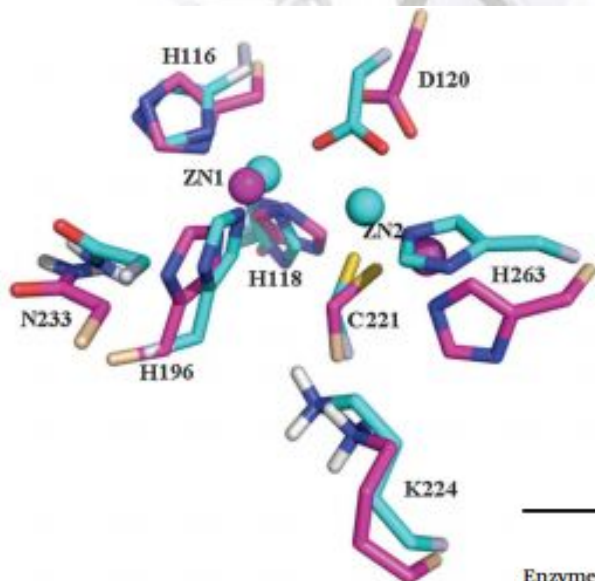
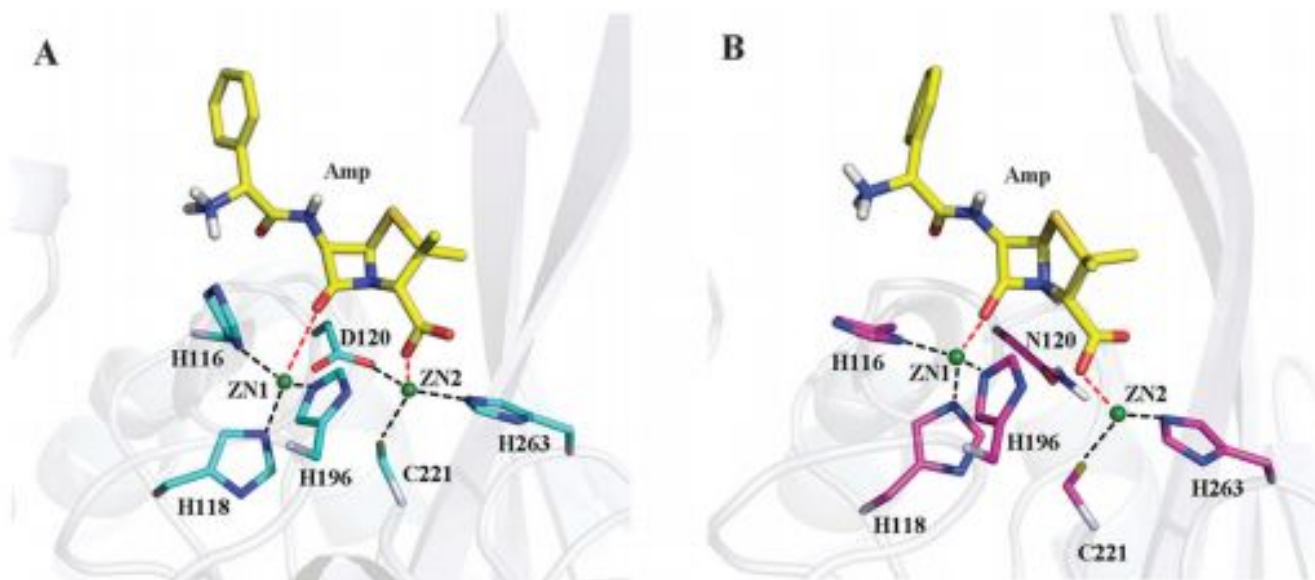




Yates, Sternberg, 2013.

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

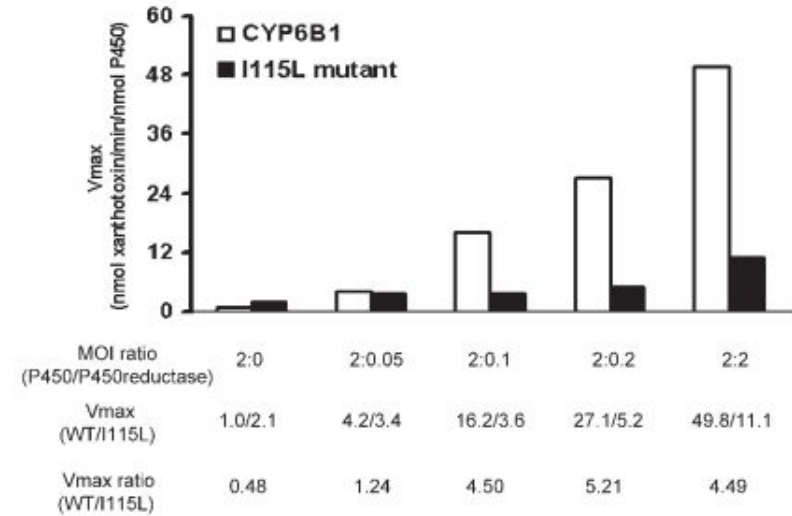
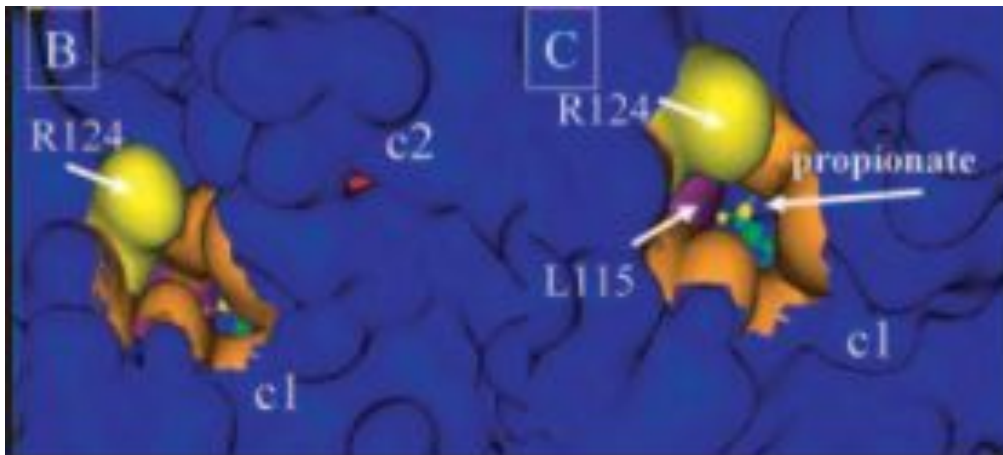
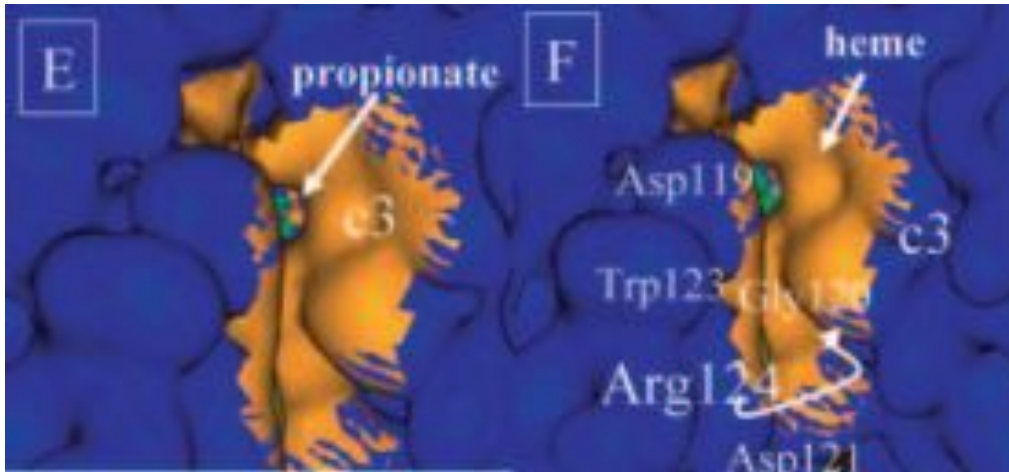
# Мутации в активном центре



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.

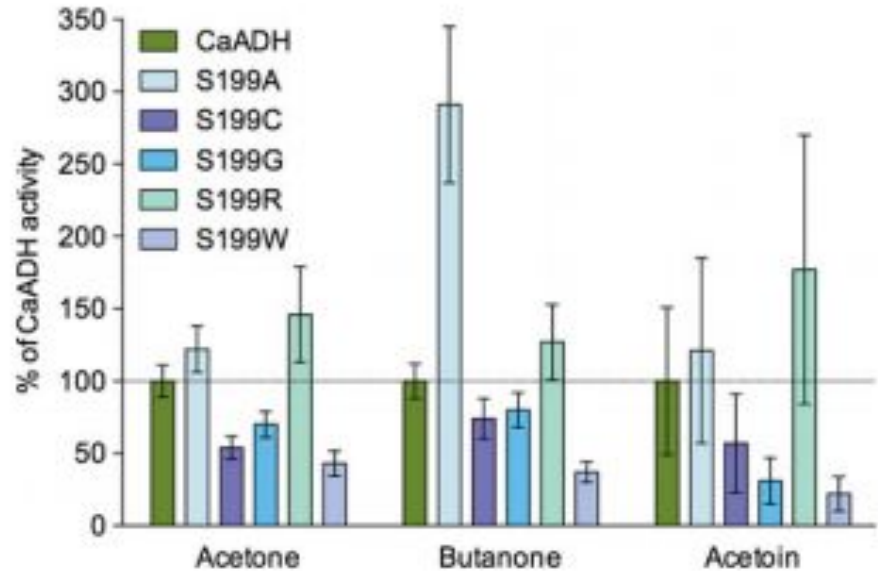
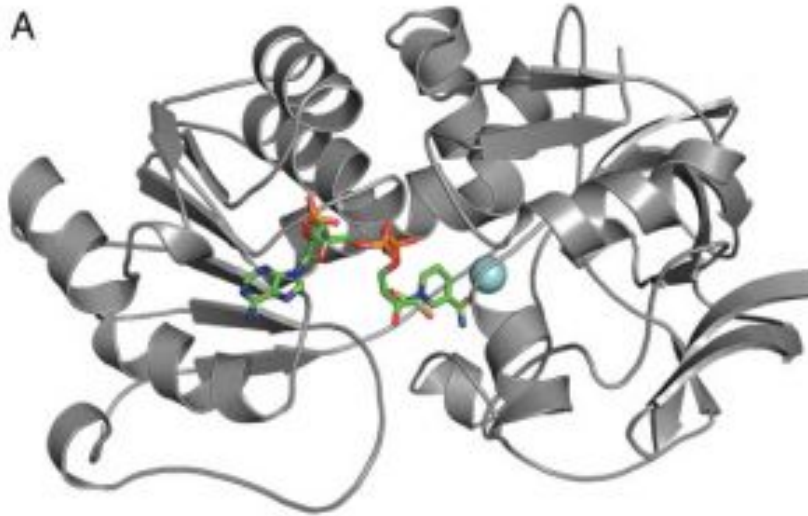
Enzymes harbored in <i>E. coli</i> BL21 (DE3)	Antibiotics ( $\mu\text{g ml}^{-1}$ )				
	Penicillin G	Ampicillin	Cefuroxime	Ceftizoxime	Meropenem
Wild-type NDM-1	512	> 512	> 512	64	16
D120N mutant	8	4	1	1	< 0.5

# Мутации, меняющие доступ к активному центру



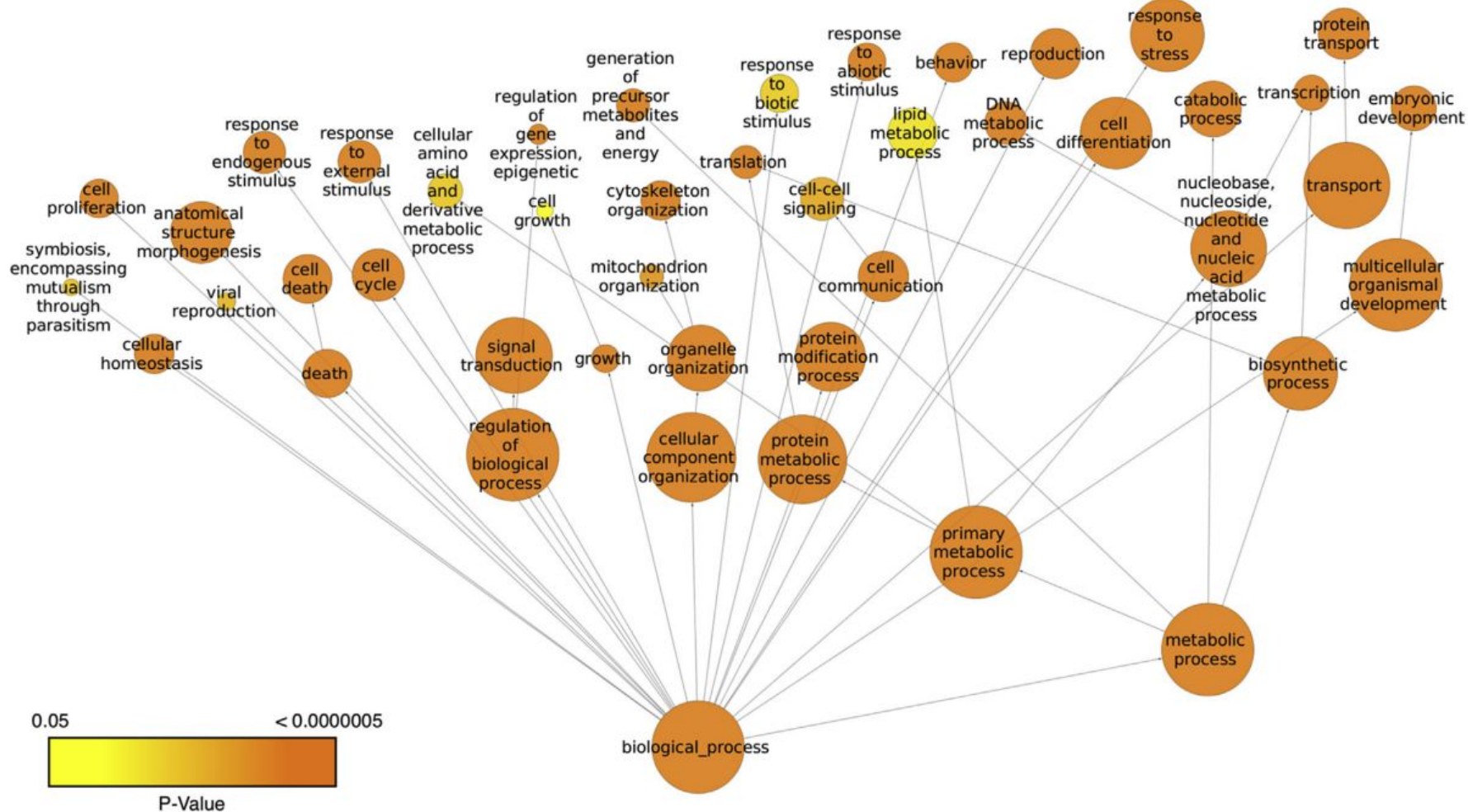
Замена изолейцина на лейцин в CYP6B1 *Papilio polyxenes* ограничила доступ субстрата к активному центру, что привело к невозможности метаболизировать фуранокумарин.

# Мутации в сайте связывания с коферментом



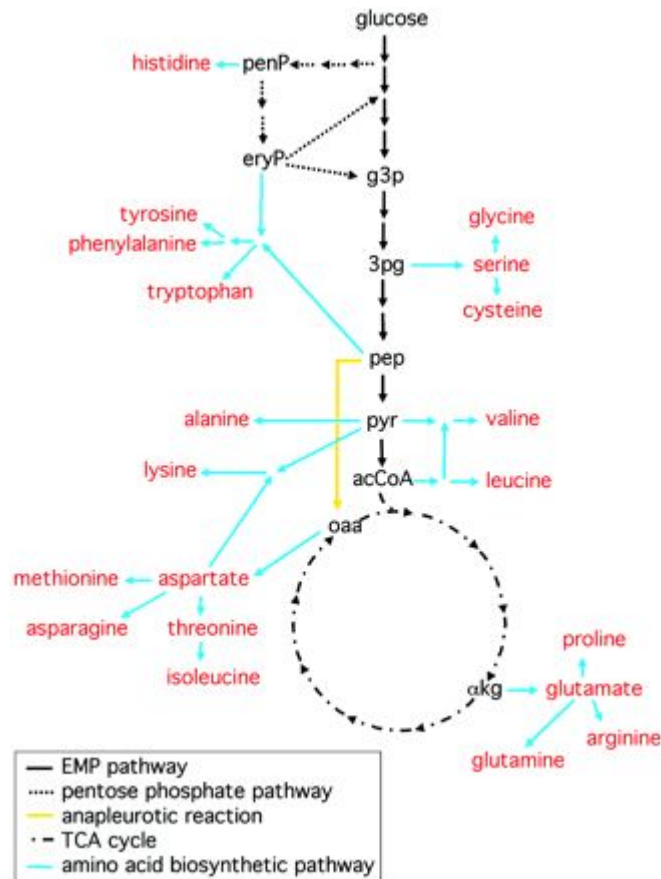
Изменение в месте связывания с НАДФН, расположенному в удалении от активного центра, в алкогольдегидрогеназе *Clostridium autoethanogenum* привело к изменению в специфичности к субстрату. В некоторых случаях менялась специфичность к коферменту (НАДН вместо НАДФН)





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

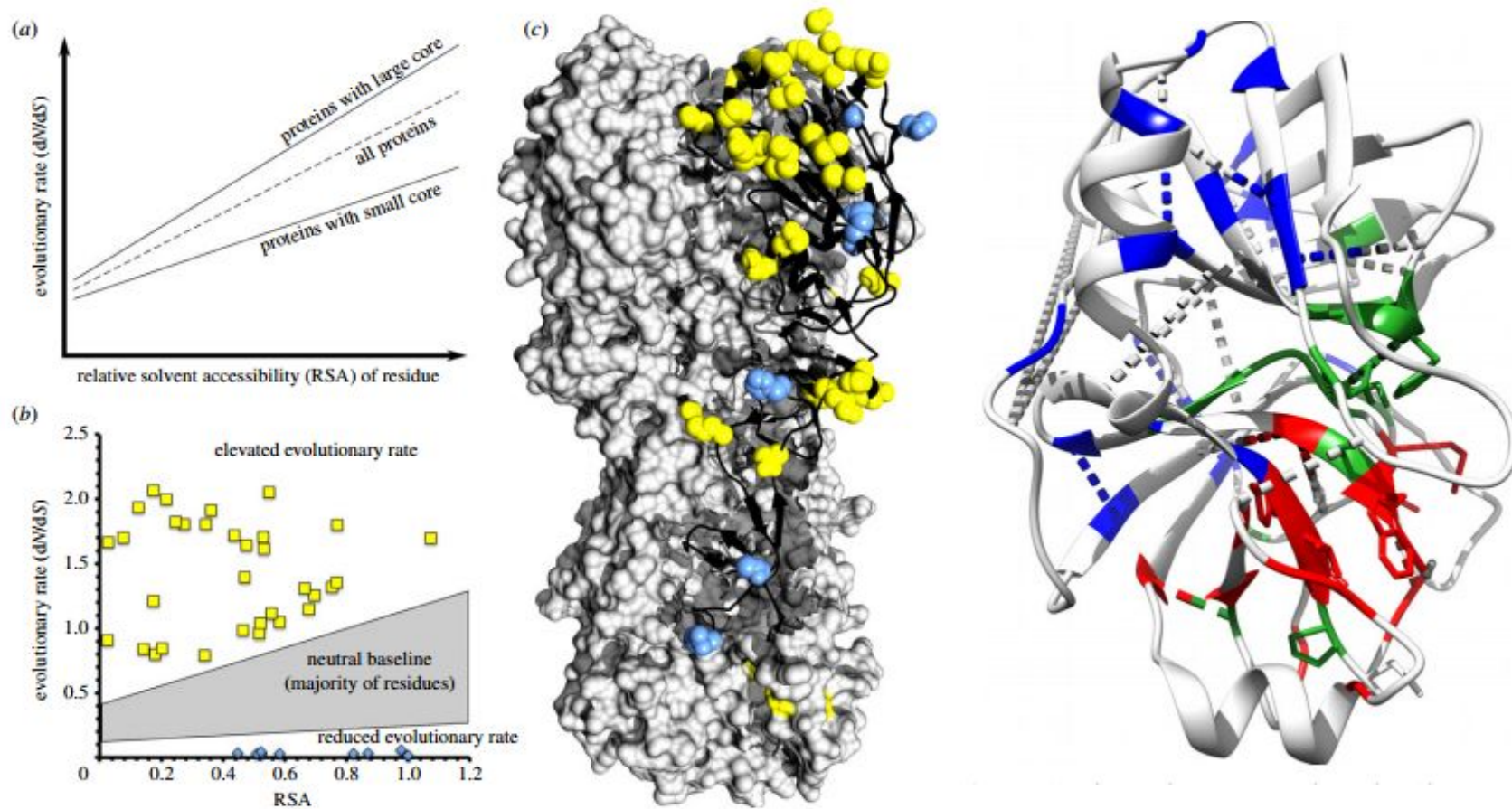
# Цена за аминокислоту



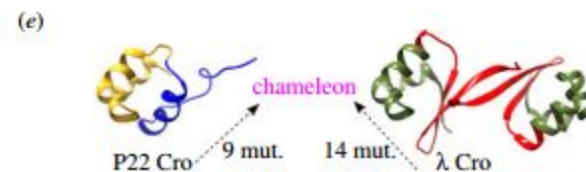
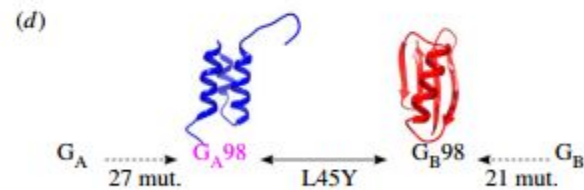
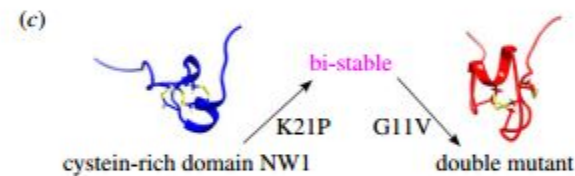
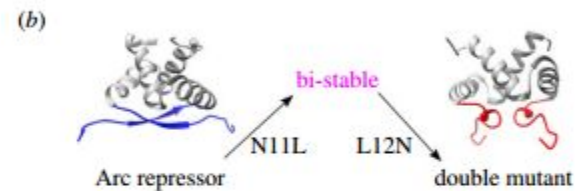
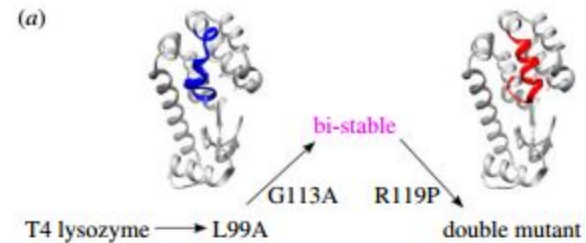
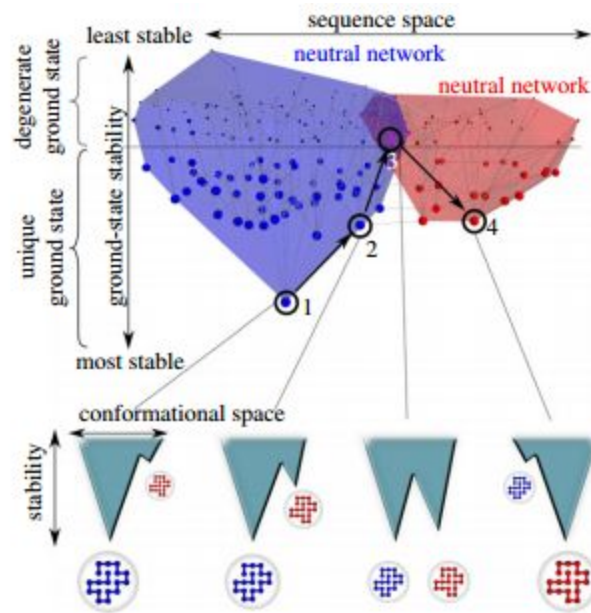
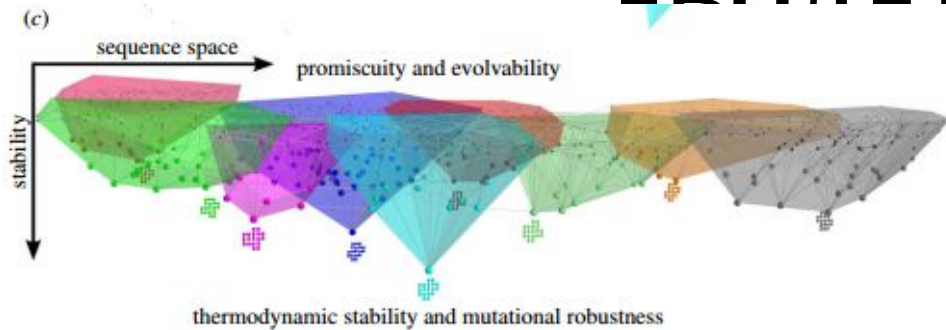
Amino acid	One-letter symbol	Precursor metabolites	Energetic cost		
			-P	H	Total, -P
Ala	A	pyr	1.0	5.3	11.7
Cys	C	3pg	7.3	8.7	24.7
Asp	D	oaa	1.3	5.7	12.7
Glu	E	αkg	2.7	6.3	15.3
Phe	F	2 pep, eryP	13.3	19.3	52.0
Gly	G	3pg	2.3	4.7	11.7
His	H	penP	20.3	9.0	38.3
Ile	I	pyr, oaa	4.3	14.0	32.3
Lys	K	oaa, pyr	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

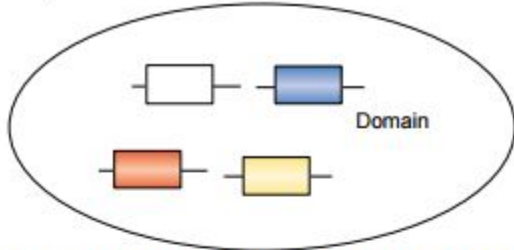


# Структура белка и его эволюция

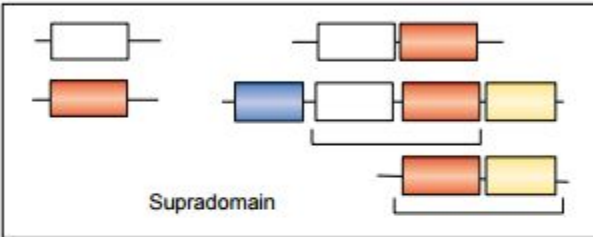




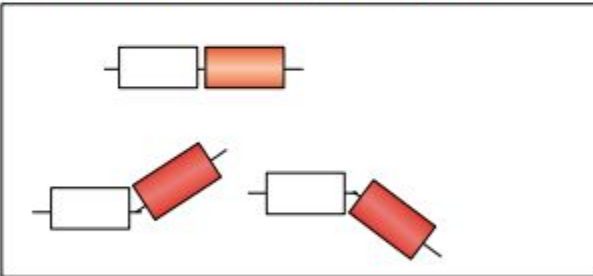
The repertoire of domain superfamilies...



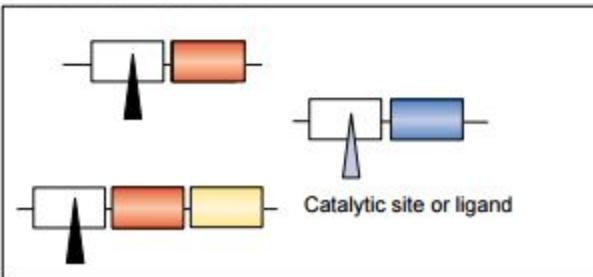
...duplicates and recombines to form single and multi-domain proteins.



The same combination can adopt different geometries...

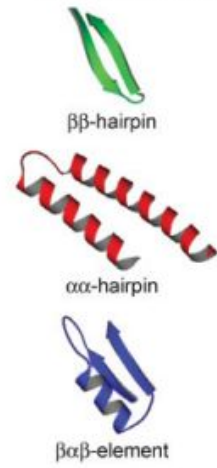


...and/or different functions.

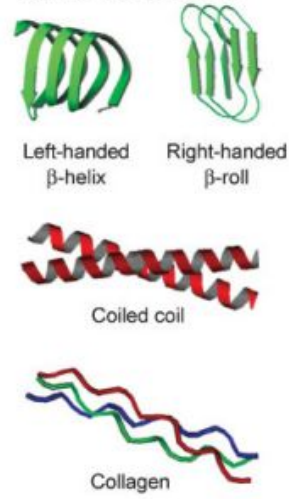


Current Opinion in Structural Biology

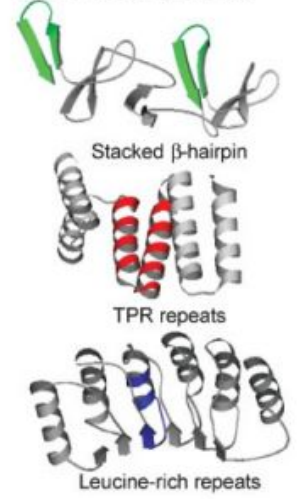
**a) Supersecondary structure elements**



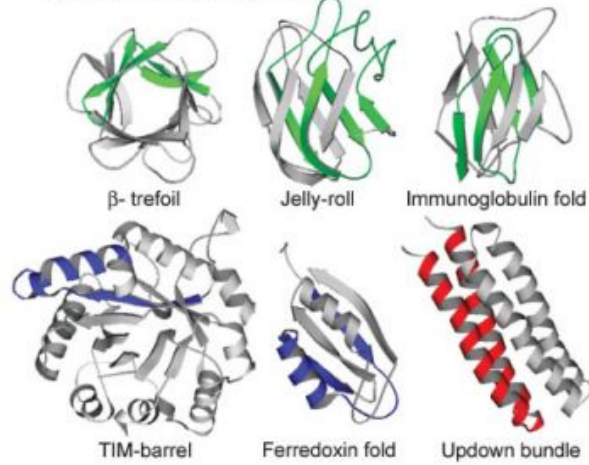
**b) Fibrous proteins**



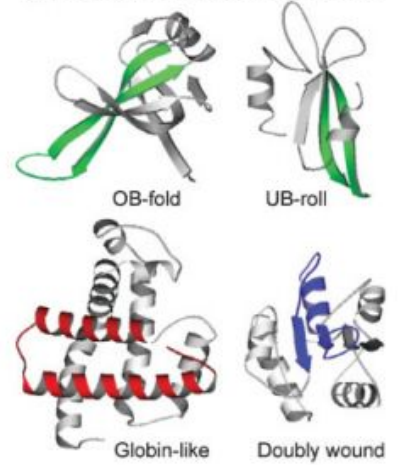
**c) Solenoid proteins**



**d) Symmetrical superfolds**



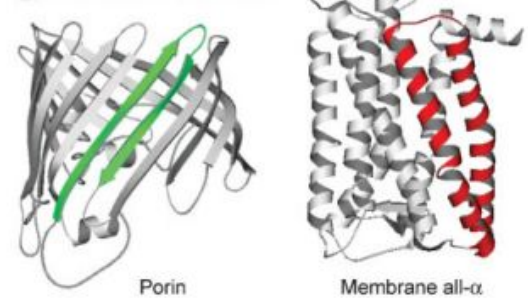
**e) Non-symmetrical superfolds**



**f)  $\beta$ -propeller**



**g) Membrane domains**



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 <i>H. pylori</i> , the association is 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 <i>Chlamydomonas</i> [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 than 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