

Fitness landscapes

Mapping genotype to phenotype

References

- Quantitative analyses of empirical fitness landscapes, Szendro et. al. **2012**
- Empirical fitness landscapes and the predictability of evolution, de Visser & Krug **2014**
- Evolutionary biochemistry: revealing the historical and physical causes of protein properties, Harms & Thornton **2013**
- Learning the pattern of epistasis linking genotype and phenotype in a protein, Poelwijk et. al. **2019**
- Capturing the mutational landscape of the beta-lactamase TEM-1, Jacquier et. al. **2013**
- A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape, Firnberg et. al. **2014**
- Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1, Figliuzzi et. al. **2015**
- Evolutionary information for specifying a protein fold, Socolich et. al. **2005**

The fitness landscape

Selection!

GPOP: abstract locus – alleles with fitness effects s_1, s_2, \dots

In practice

Locus



Gene



Protein (structure, function)

Allele



Nucleotide sequence



attcatgatctgcgccataccaaacgataaa...

Amino acid sequence

IHDLRHTNDK...

wild-type

IHDLRHTNDK... $s=0$

mutants ...

IHDLRHDNDK... $s=?$

IHDLCHTNDK... $s=?$

IADLRHTNTK... $s=?$

selection coefficient

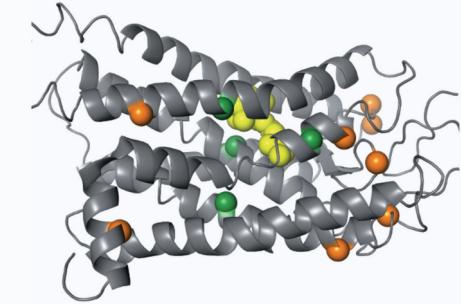
Amino acid sequence

$a_1, a_2 \dots a_L$

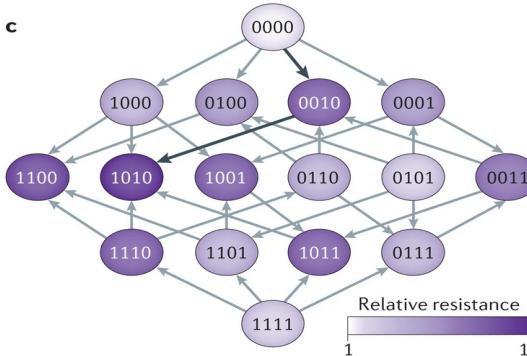
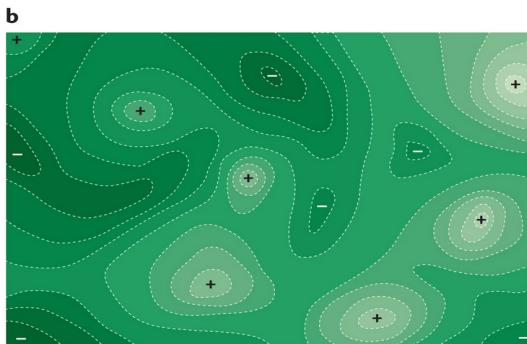
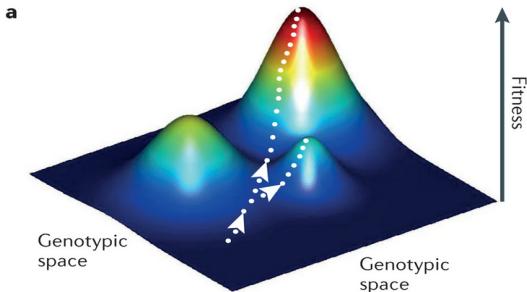
$$s = \phi(a_1, a_2 \dots a_L)$$



Fitness landscape



The fitness landscape: history



20th century: Concepts and models

- **S. Wright** 1932
 - concept of fitness landscape
 - visualization in a continuous space
- **J. Maynard Smith** 1970 — mutational pathways through functional sequences

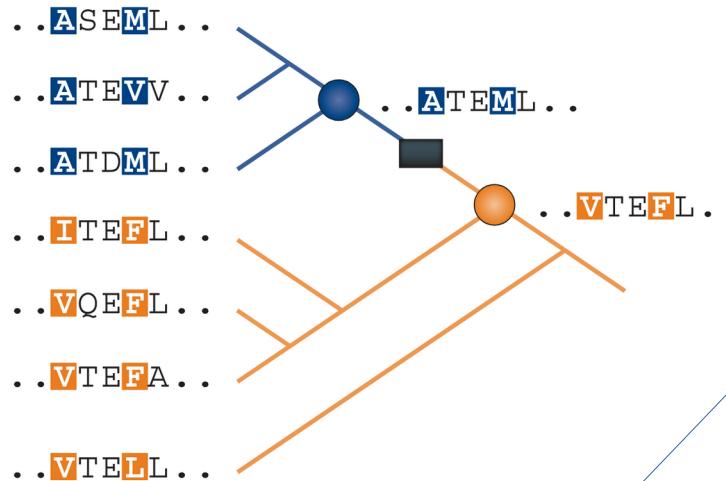
word – wore – gore – gone – gene
- **S Kauffman et. al.** 1987 — study of abstract fitness landscapes: HoC, NK, ...

21st century: Experimental data, computational models

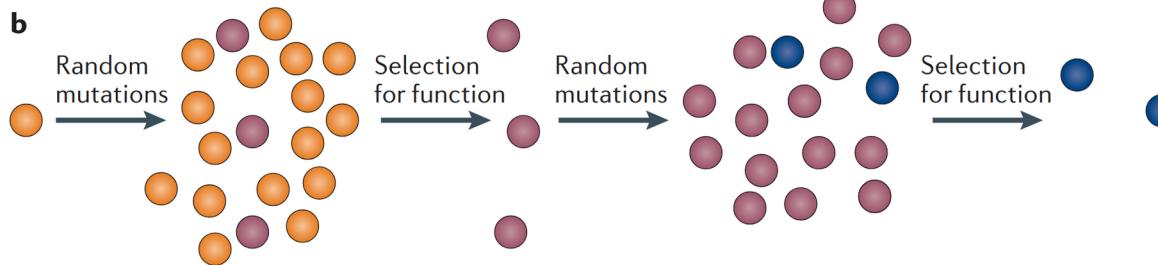
- **Pathway reconstruction:** all combinations of k mutations ($k < 13$) between two sequences
→ fitness for 2^k genotypes ($\sim 10\,000$)
- **Deep mutational scan:** single (& double) mutants around a given wild-type
- **Computational models & learning:** infer fitness landscape from sequences!

Mutational pathways

Homologs / ancestral reconstruction



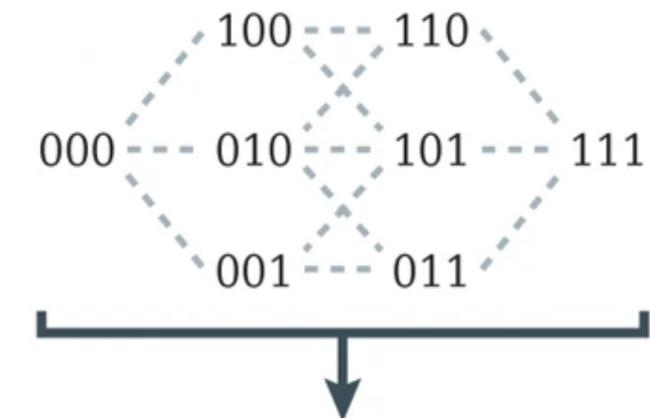
Directed evolution



Two sequences of interest

...NINGKFRTYETAYKAATQHNGSVQM...
...NIGGKF_YTYETAVKAATQHNGSFQM...

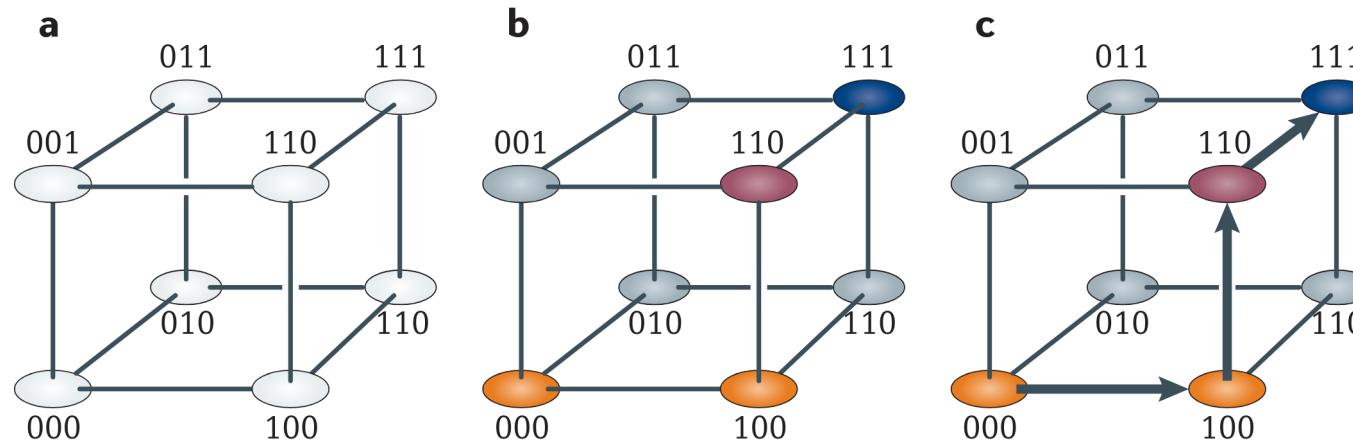
k mutations → 2^k intermediates



Measure fitness or proxy

Mutational pathways

k mutations $\rightarrow 2^k$ intermediates \rightarrow dimension k hypercube!

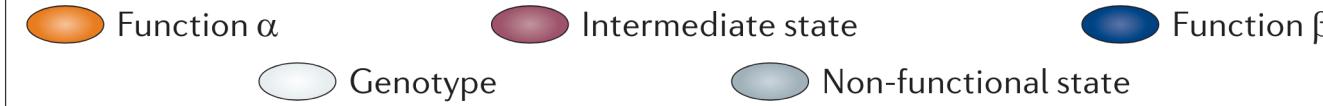


Biochemistry

Genotype space

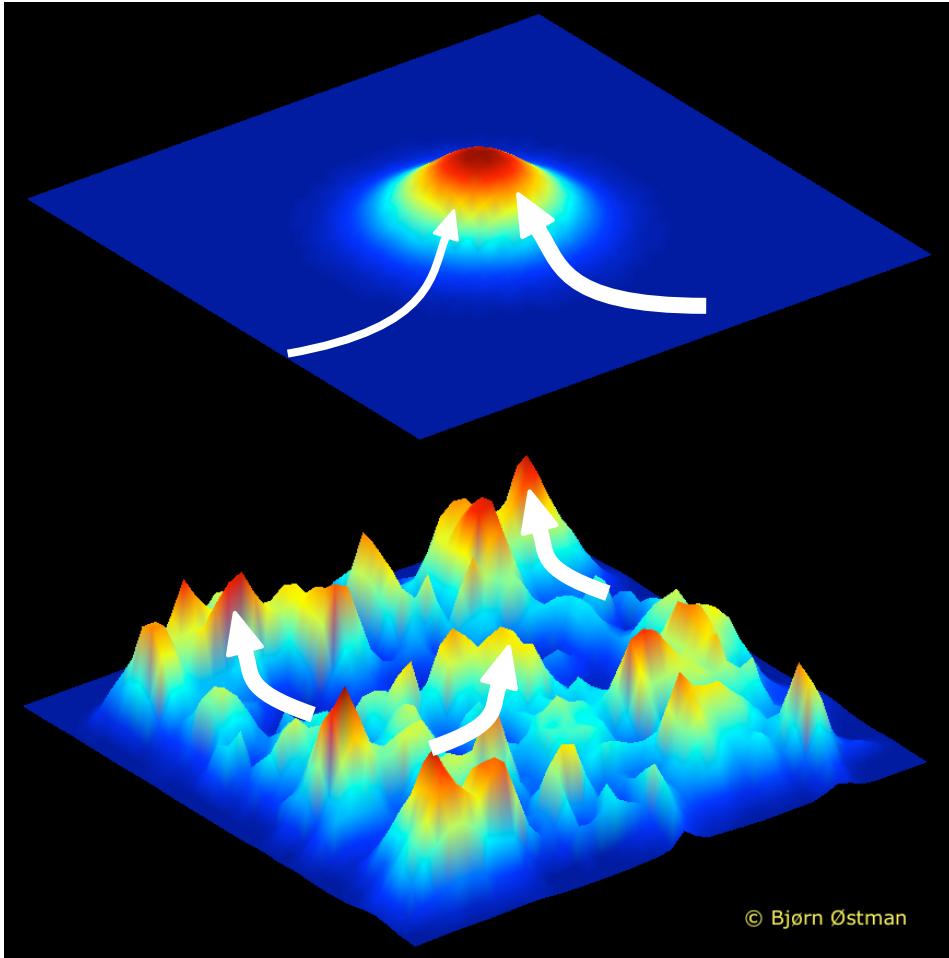
Evolution

Genotype–phenotype space \rightarrow Trajectory through space



Paths from one function to other?

Smooth vs rugged landscapes



Simple landscape

- single fitness peak
- evolution: uphill walk to peak

Rugged landscape

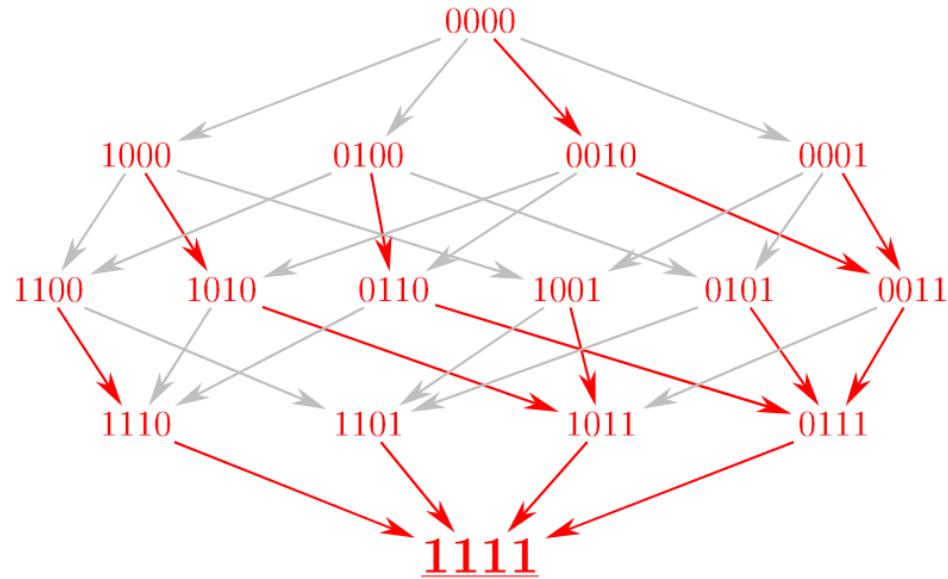
- multiple fitness peaks
- evolution: result depends on initial sequence and randomness

Why rugged?

Ruggedness caused by **epistasis**:
non-independence of mutations!

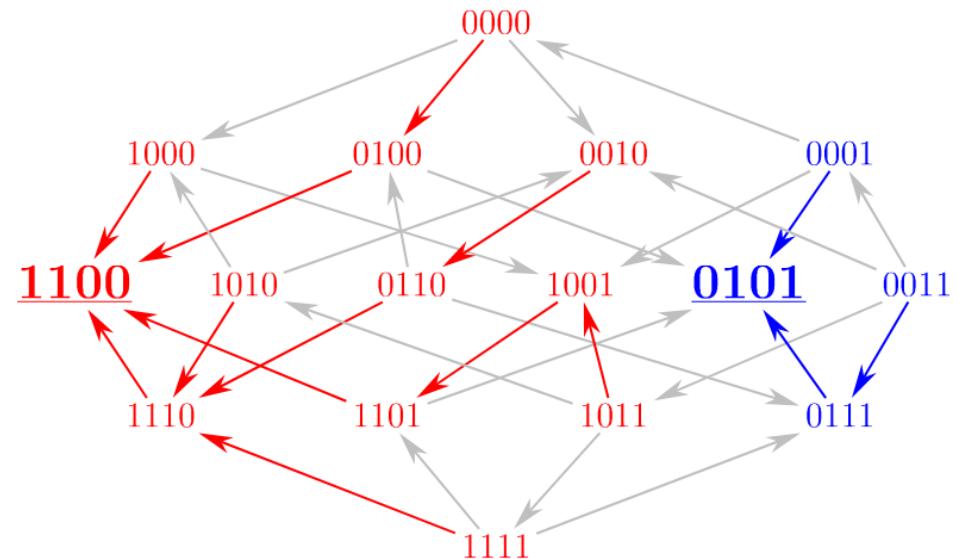
Empirical data

Example landscapes with four mutations



One fitness maximum

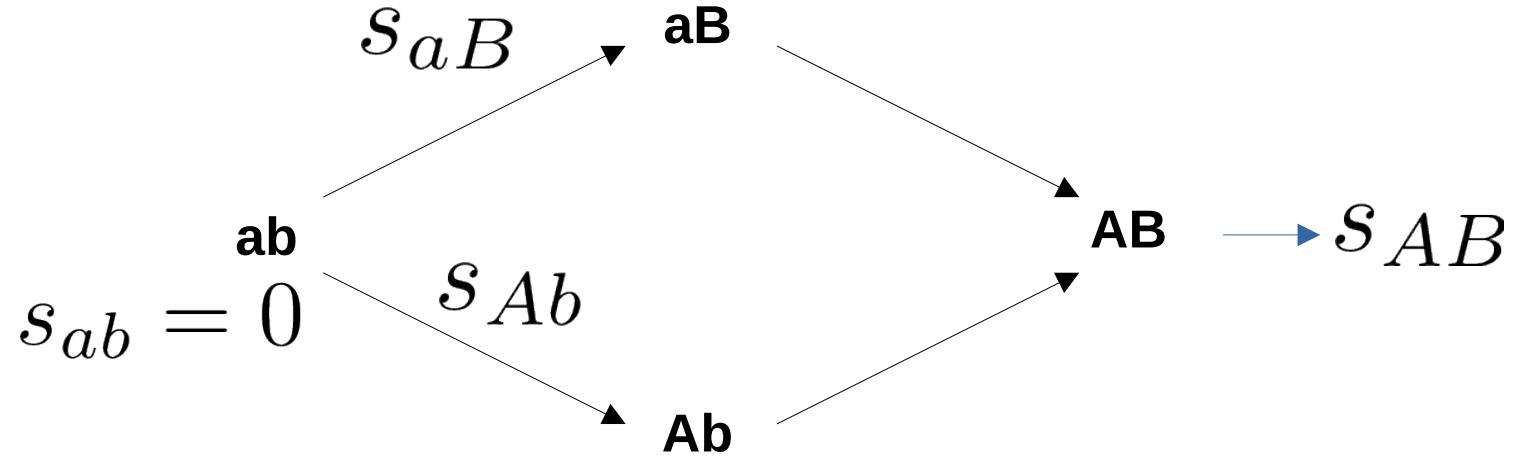
Beneficial mutations in a metabolic pathway of *Methylobacterium extorquens*
[Chou et. al. 2011]



Two fitness maxima

Four mutations conferring pyrimethamine resistance in Malaria (dihydrofolate reductase)
[Lozovsky et. al. 2009]

Why ruggedness? Epistasis



No epistasis: additivity of effects $s_{AB} = s_{aB} + s_{Ab}$

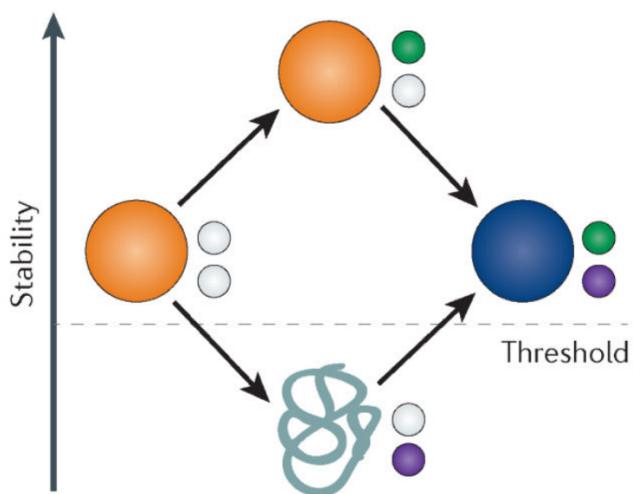
Epistasis: any deviation from additivity $s_{AB} \neq s_{aB} + s_{Ab}$

- Diminishing returns
- Sign epistasis

Molecular determinants of epistasis

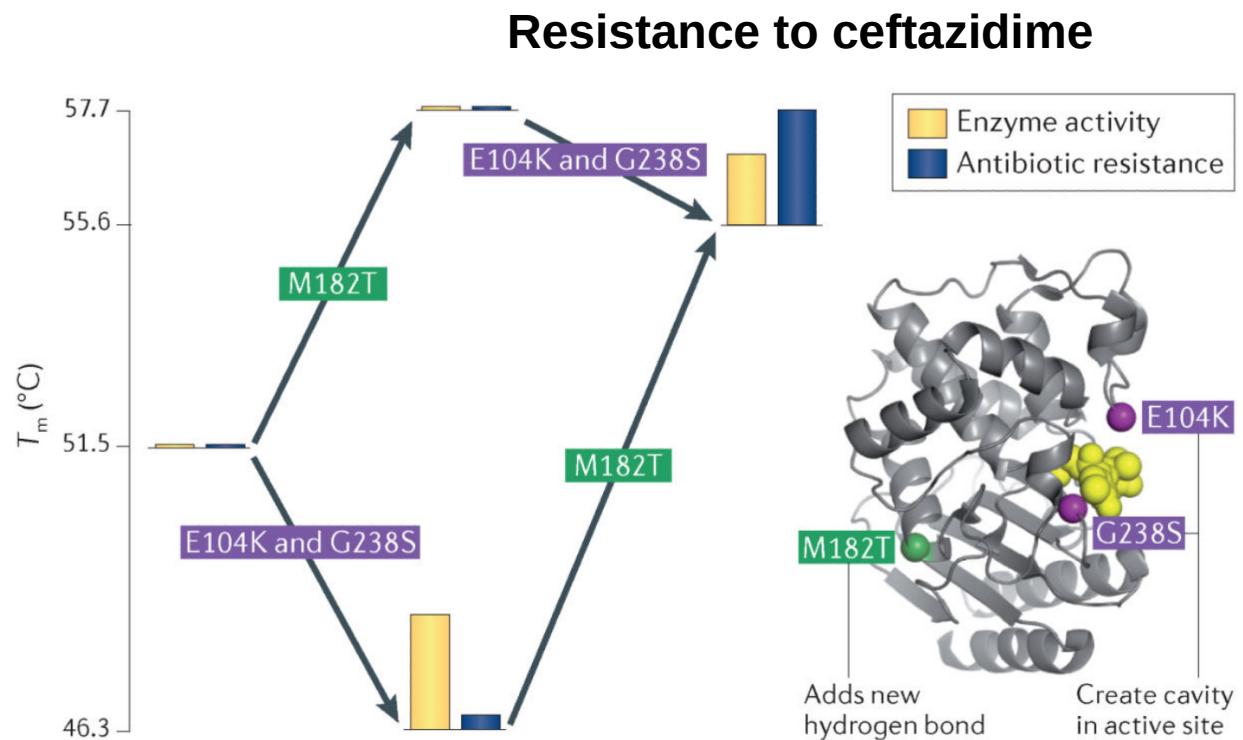
Permissive mutation

- No individual fitness effect
- Enables other beneficial mutations



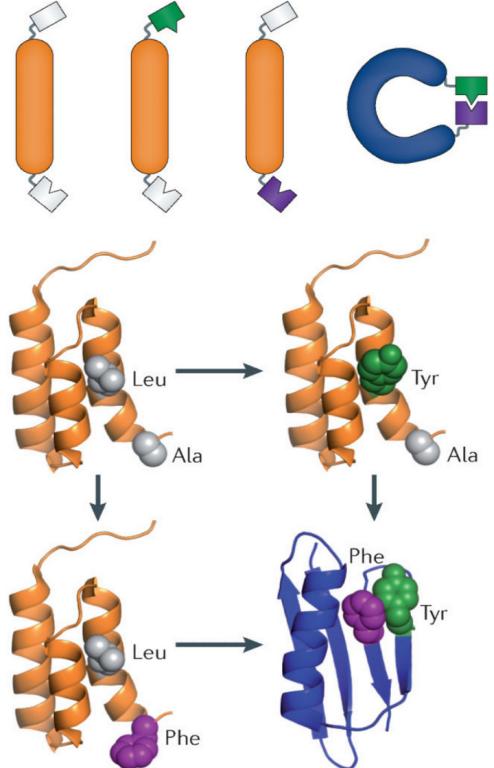
Non-specific, permissive epistasis

- Permissive mutation increases protein stability
- Beneficial mutation decreases protein stability

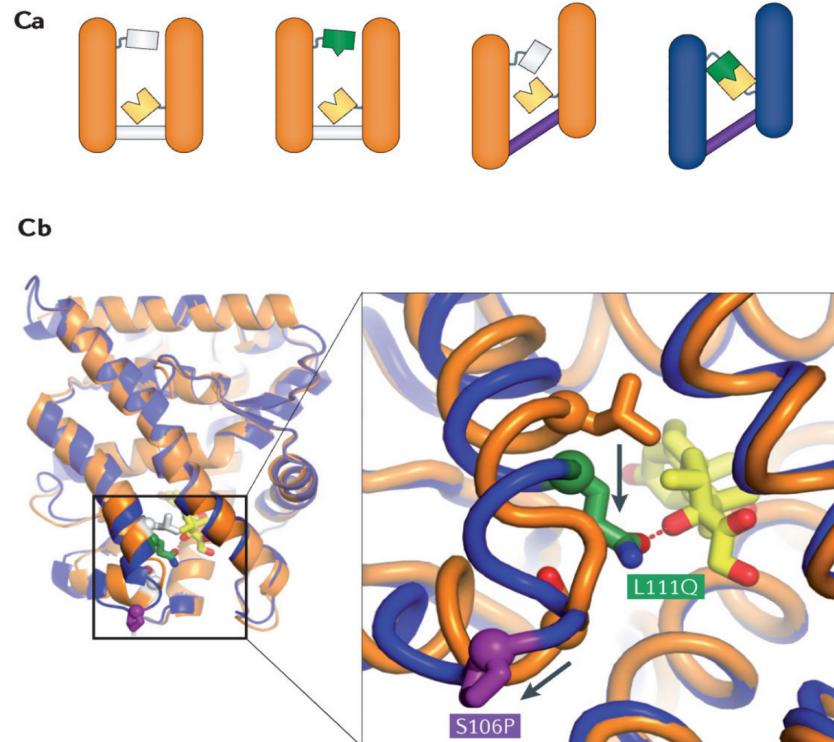


Molecular determinants of epistasis

“Specific” epistasis: combination of mutations have a collective effect



Direct physical interaction

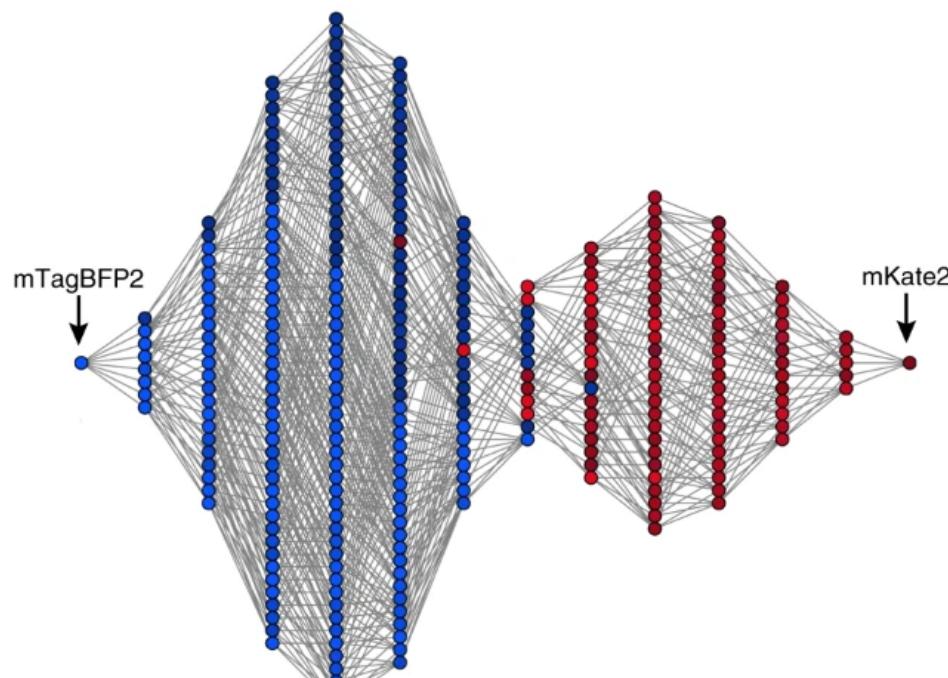


Indirect interaction

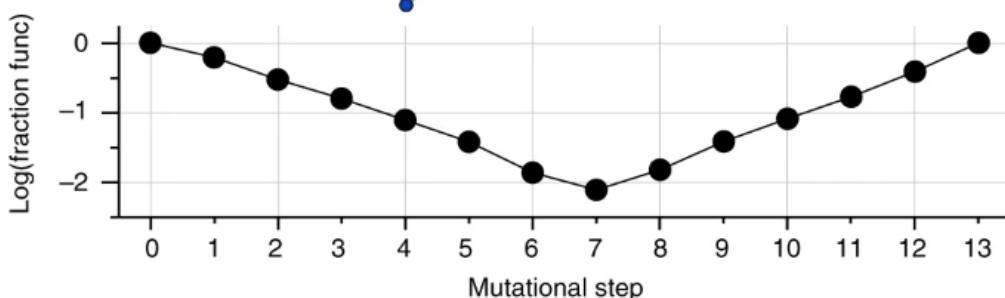
Green: potential interaction with yellow
Purple: conformational change enables interaction

Epistasis constrains pathways

a



b



Two variants of the *Entacmaea quadricolor* fluorescent protein

- Blue or red fluorescence
- 13 mutations between the two
- $2^{13} = 8096$ phenotypes!

Predictability of evolution?

Small population

- Weak mutation regime
- beneficial mutations fix if they survive drift
- fixation probability \sim fitness advantage

Low predictability

Intermediate population

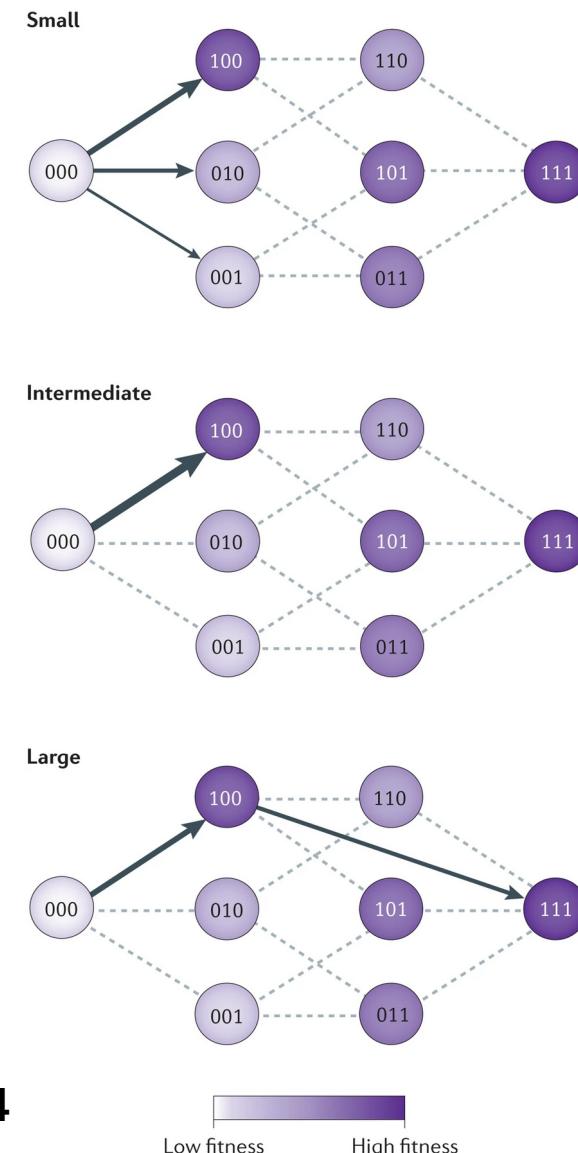
- Many mutations
- beneficial mutations compete: *clonal interference*
- most beneficial mutation likely to fix first

High predictability

Large population

- Strains with many beneficial mutations
- Cross fitness valleys

Low predictability

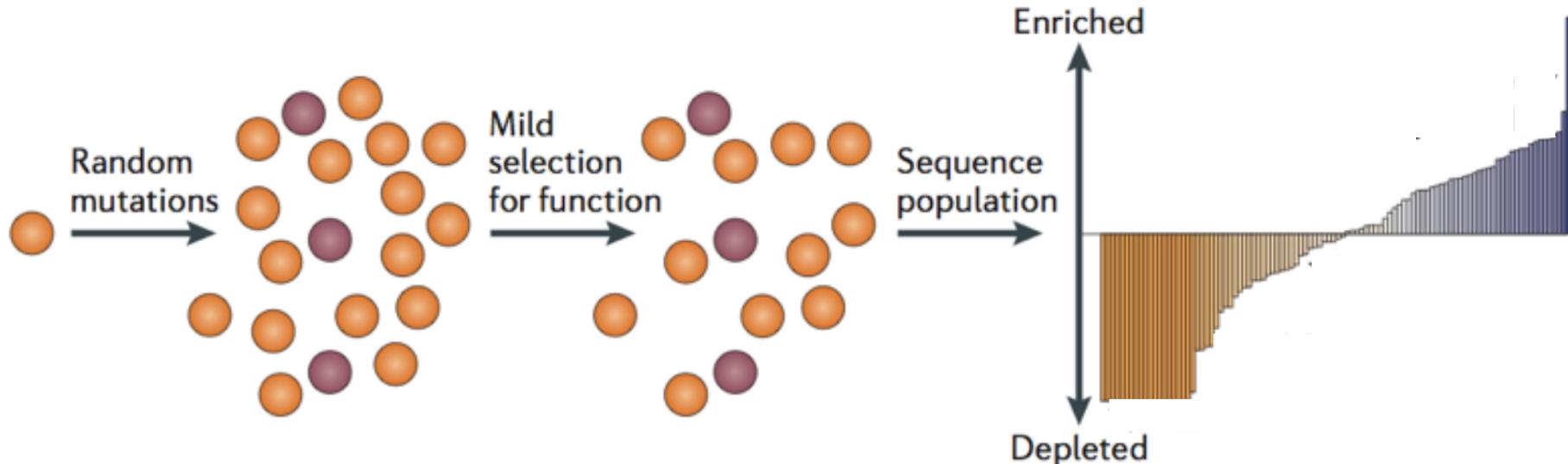


de Visser & Krug 2014

Deep mutational scans

Mutational pathway: steps between two sequences of interest

Mutational scan: explore landscape around one reference sequence



→ **Large library with few mutations** around reference

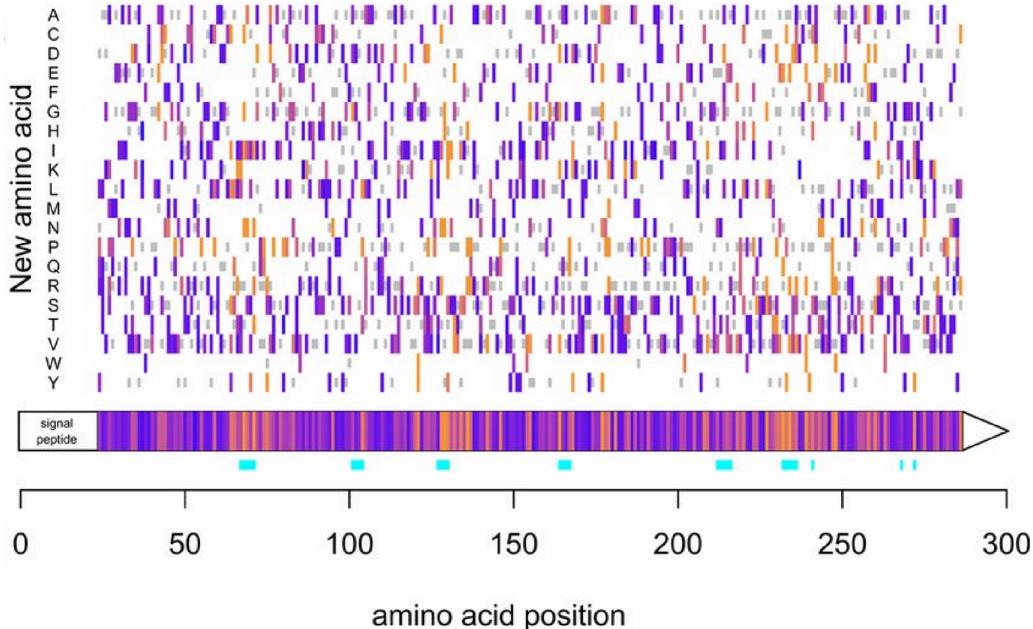
Deep mutational scans

PNAS

Capturing the mutational landscape of the beta-lactamase TEM-1

Hervé Jacquier^{a,b,c,1}, André Birgy^{a,b}, Hervé Le Nagard^{a,b,d,e}, Yves Mechulam^f, Emmanuelle Schmitt^f, Jérémie Glodt^{a,b}, Beatrice Bercot^{c,g}, Emmanuelle Petit^h, Julie Poulaïn^h, Guilène Barnaudⁱ, Pierre-Alexis Gros^{a,b,j}, and Olivier Tenaillon^{a,b,1}

Edited by Bruce R. Levin, Emory University, Atlanta, GA, and approved June 25, 2013 (received for review September 21, 2012)



→ most or all **single mutants** around one sequence

Phenotype: resistance to antibiotic

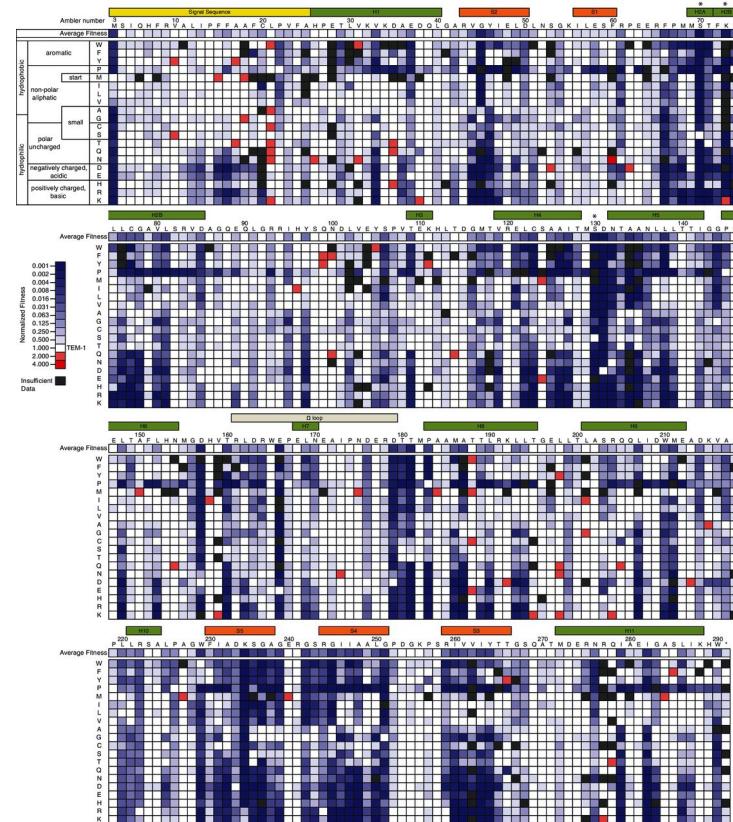
MOLECULAR BIOLOGY AND EVOLUTION

A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape

Elad Firnberg, Jason W. Labonte, Jeffrey J. Gray, Marc Ostermeier Author Notes

Molecular Biology and Evolution, Volume 31, Issue 6, June 2014, Pages 1581–1592, <https://doi.org/10.1093/molbev/msu081>

Published: 23 February 2014



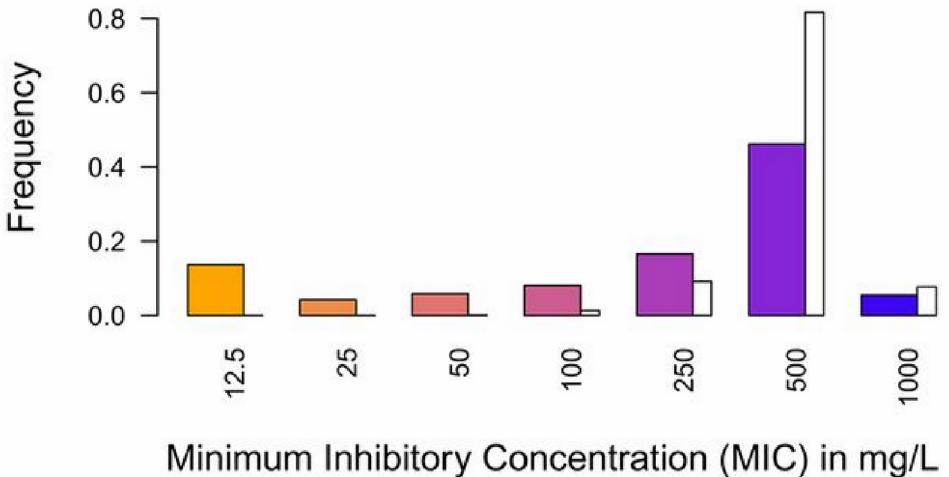
Distribution of fitness effects

PNAS

Capturing the mutational landscape of the beta-lactamase TEM-1

Hervé Jacquier^{a,b,c,1}, André Birgy^{a,b}, Hervé Le Nagard^{a,b,d,e}, Yves Mechulam^f, Emmanuelle Schmitt^f, Jérémie Glodt^{a,b}, Beatrice Bercot^{c,g}, Emmanuelle Petit^h, Julie Poulain^h, Guilène Barnaudⁱ, Pierre-Alexis Gros^{a,b,j}, and Olivier Tenaillon^{a,b,1}

Edited by Bruce R. Levin, Emory University, Atlanta, GA, and approved June 25, 2013 (received for review September 21, 2012)

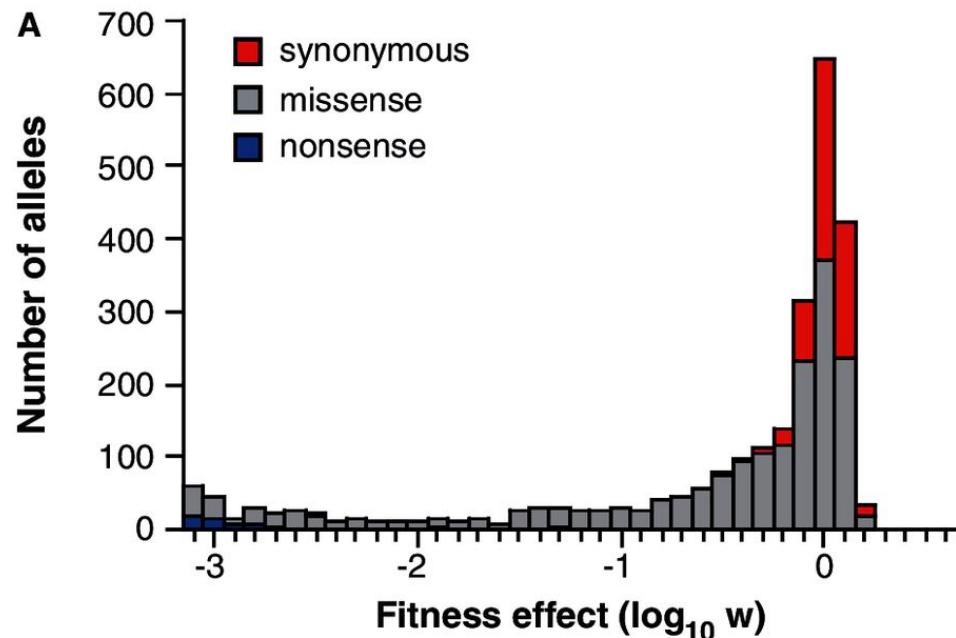


MOLECULAR BIOLOGY AND EVOLUTION
A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape

Elad Firlberg, Jason W. Labonte, Jeffrey J. Gray, Marc Ostermeier

Molecular Biology and Evolution, Volume 31, Issue 6, June 2014, Pages 1581–1592, <https://doi.org/10.1093/molbev/msu081>

Published: 23 February 2014



Prediction of fitness effects?

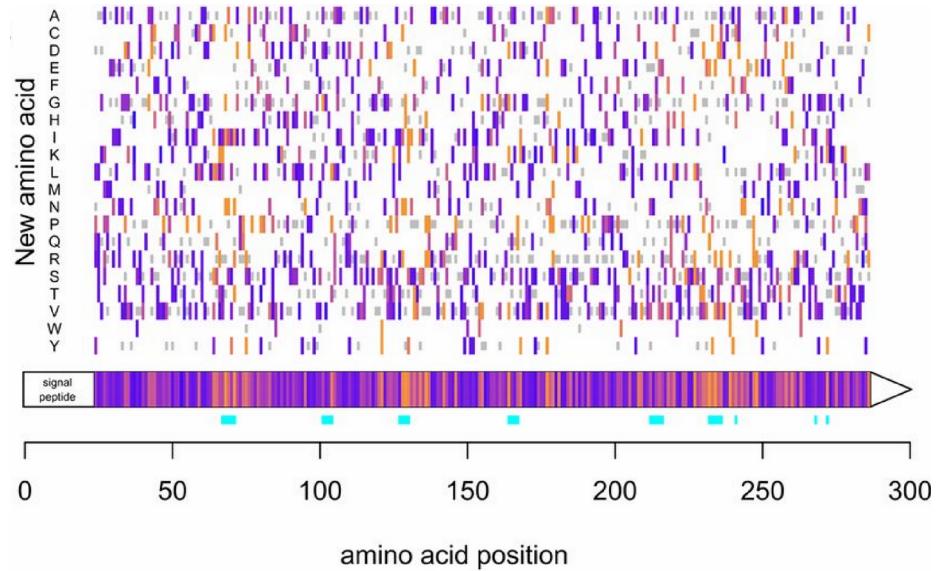
PNAS

Capturing the mutational landscape of the beta-lactamase TEM-1

Hervé Jacquier^{a,b,c,1}, André Birgy^{a,b}, Hervé Le Nagard^{a,b,d,e}, Yves Mechulam^f, Emmanuelle Schmitt^f, Jérémie Glodt^{a,b}, Beatrice Bercot^{c,g}, Emmanuelle Petit^h, Julie Poulaïn^h, Guilène Barnaudⁱ, Pierre-Alexis Gros^{a,b,i}, and Olivier Tenaillon^{a,b,1} Jérémie Glodt^{a,b}, Beatrice Bercot^{c,g}, Emmanuelle Petit^h, Julie Poulaïn^h, Guilène Barnaudⁱ, Pierre-Alexis Gros^{a,b,i}, and Olivier Tenaillon^{a,b,1} Olivier Tenaillon^{a,b,1}

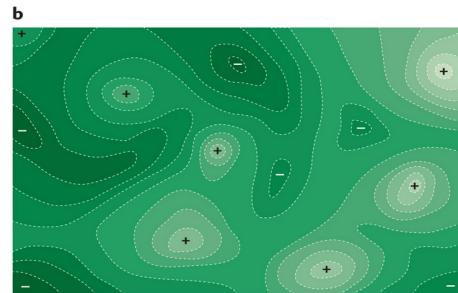
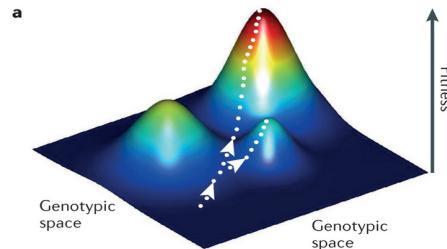
Edited by Bruce R. Levin, Emory University, Atlanta, GA, and approved June 25, 2013 (received for review September 21, 2012)

Edited by Bruce R. Levin, Emory University, Atlanta, GA, and approved June 25, 2013 (received for review September 21, 2012)



Predictor	R^2
BLOSUM62	0.16
Accessibility	0.22
$\Delta\Delta G$ Popmusic	0.19
$\Delta\Delta G$ foldX	0.15
BLOSUM62 + Accessibility	0.38 (0.43)
BLOSUM62 + $\Delta\Delta G$ Popmusic	0.28 (0.28)
BLOSUM62 + $\Delta\Delta G$ foldX	0.24 (0.24)
Accessibility + $\Delta\Delta G$ Popmusic	0.27 (0.27)
Accessibility + $\Delta\Delta G$ foldX	0.30 (0.32)
BLOSUM62 + Accessibility + $\Delta\Delta G$ Popmusic	0.40 (0.44)
BLOSUM62 + Accessibility + $\Delta\Delta G$ foldX	0.42 (0.46)

Modeling mutational landscapes



Genotype – phenotype map

$$A_1, \dots, A_L \xrightarrow{\text{sequence}} \varphi(A_1, \dots, A_L) \xrightarrow{\text{phenotype}}$$

20^L possible sequences to 20^L possible values

Reduce number of parameters?

No epistasis

$$\varphi(A_1 \dots A_L) = \sum_i \phi_i(A_i)$$

$20 \times L$ parameters

Equivalent to **sequence profiles**

Pairwise epistasis

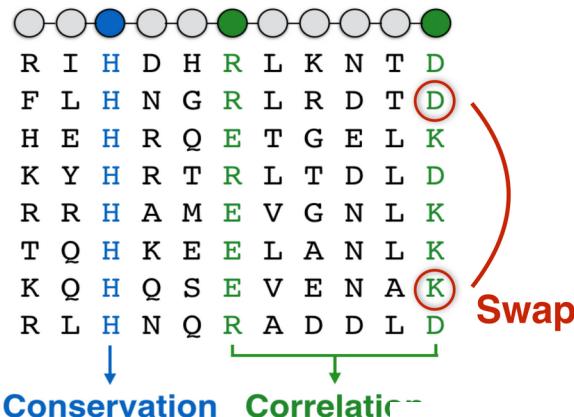
$$\varphi(A_1 \dots A_L) = \sum_i \phi_i(A_i) + \sum_{i < j} \phi_{ij}(A_i, A_j)$$

$O(L^2)$ parameters

Why pairwise epistasis?

[Socolich et. al. 2005]

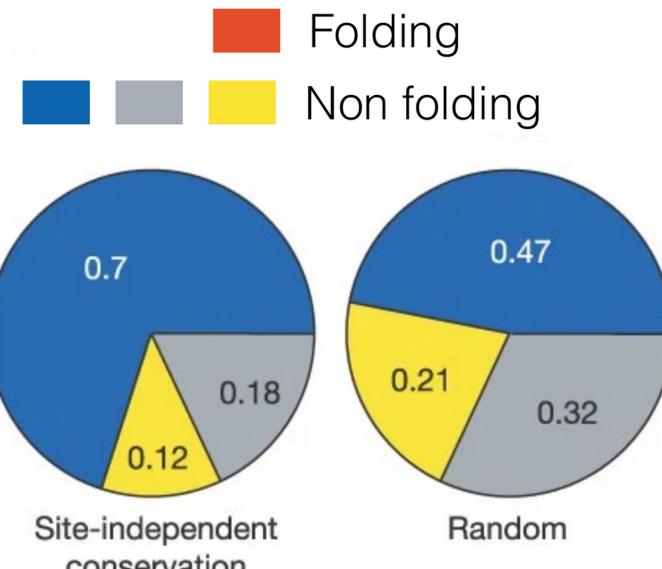
Protein family: WW domain



Experiment: scramble the alignment

- Randomly: global amino-acid frequencies constant
- By column: **conservation** constant
- By column + constraints: **correlation** constant

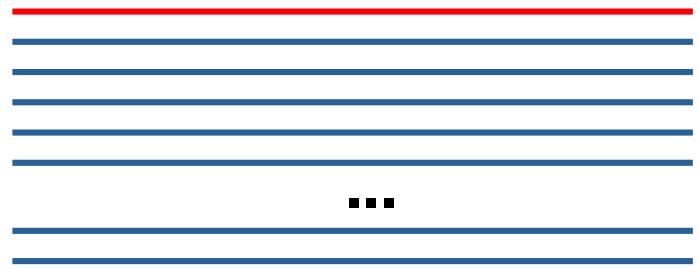
Express resulting proteins, measure **folding**



Pairwise statistics are (mostly) good enough!

Learning from protein families

Beta-lactamase2 family (PF13354)



Family of homologous proteins

- ~ 3000 **diverged** sequences (from different species)
- hundreds of mutations between pairs of sequences
- evolved a long time under similar constraints

Learn statistical landscape

Hypothesis:

$$P(A_1 \dots A_L) \sim e^{\varphi(A_1 \dots A_L)}$$

Some (meager) GPOP justification

Learn parameters of φ to fit statistics found in protein family

$$\varphi(A_1 \dots A_L) = \sum_i \phi_i(A_i) + \sum_{i < j} \phi_{ij}(A_i, A_j)$$



Score mutants

$$\Delta\varphi = \varphi(\text{mutant}) - \varphi(\text{wild-type})$$

Learning from protein families

