Dan Tawfik - The robustness and innovability of protein folds

I'll give an overview of our hypotheses regarding protein evolvability, how evolvability relates to structural plasticity, to a protein's architecture and to the matrix of interactions, the interaction network if you like, that maintain its configuration.

When we say proteins evolve, what do we actually mean?

Evolvability - the ability to change along evolutionary time.

Evolvability relates to the capacity to accommodate sequence changes over time, as well as to adopt new functions - the latter is also driven by sequence changes.

Evolvability has therefore two components that are interlinked:

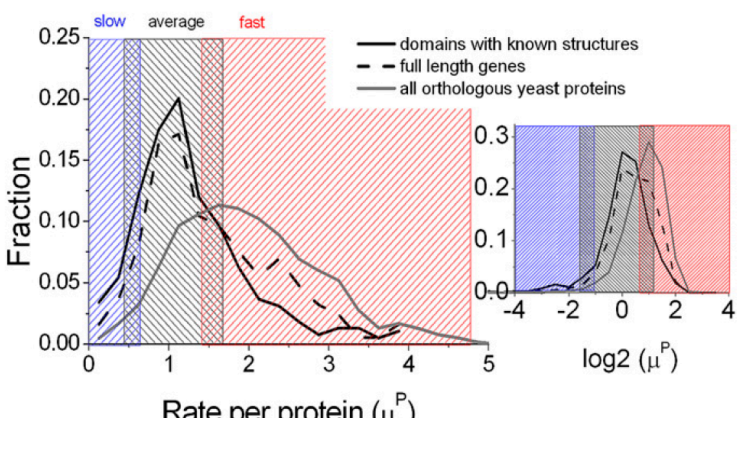
- **Robustness** - the ability to preserve a phenotype in the face of genotype changes.
- Protein robustness of proteins is defined as the ability to tolerate mutations whilst maintaining the original structure and function, and thus have the sequence change over evolutionary time (drift) at a relatively fast rate. Alternative terms**: genetic robustness**, **designability** or **neutrality**.
- **Innovability** is the ability to acquire new functions. Mutations are rare, and their combinations are extremely rare [34]. Innovability therefore relates to the ability of relatively few sequences changes to induce large changes in function and /or structure of a protein.

  Robustness and innovability are seemingly contradictory, but in fact they are complementary – robustness is a prerequisite for innovability.

How can we measure the two properties?

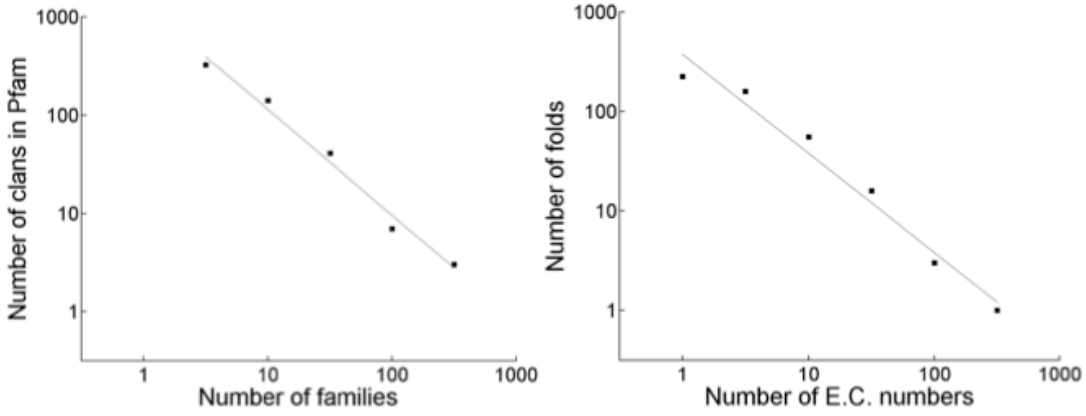
• **Robustness** is manifested in **evolutionary rates**.

Examining alignments of protein families within a given phylogeny, say all vertebrate orthologs of a given protein, and calculating the average rate of amino acid exchanges per position. The variability in evolutionary rates is high, and proteins of one given species may show up to 100-fold different rates, and thus, very different degrees of robustness, for example, evolutionary rates (per protein) for yeast proteins:



- **Innovability can be measured by analyzing the diversity of superfamilies**. These comprise functionally and structurally related proteins, typically having the same fold and the same key catalytic residues, and that are likely to have all diverged from a common ancestor. Superfamilies are comprised of different families. Each family groups many different **orthologs** – proteins belonging to different species yet sharing the same structure and function. The sequence variability between **orthologs** represents robustness, while **divergence of paralogs**, *i.e*. evolutionary related proteins with different functions, represent **innovability**.

Plotted are the number of different enzymatic functions (~ the number different paralog families) observed in different folds, or superfamiles:



We observe, a power-law behaviour, the rich get richer: few folds/superfamilies gave rise to most functional diversity; most folds/superfamilies are 'frozen' in one function.

How do these two properties – robustness, innovability, relate to the proteins architecture?

• Order in biological molecules:

Primary (sequence of bases, amino acids)

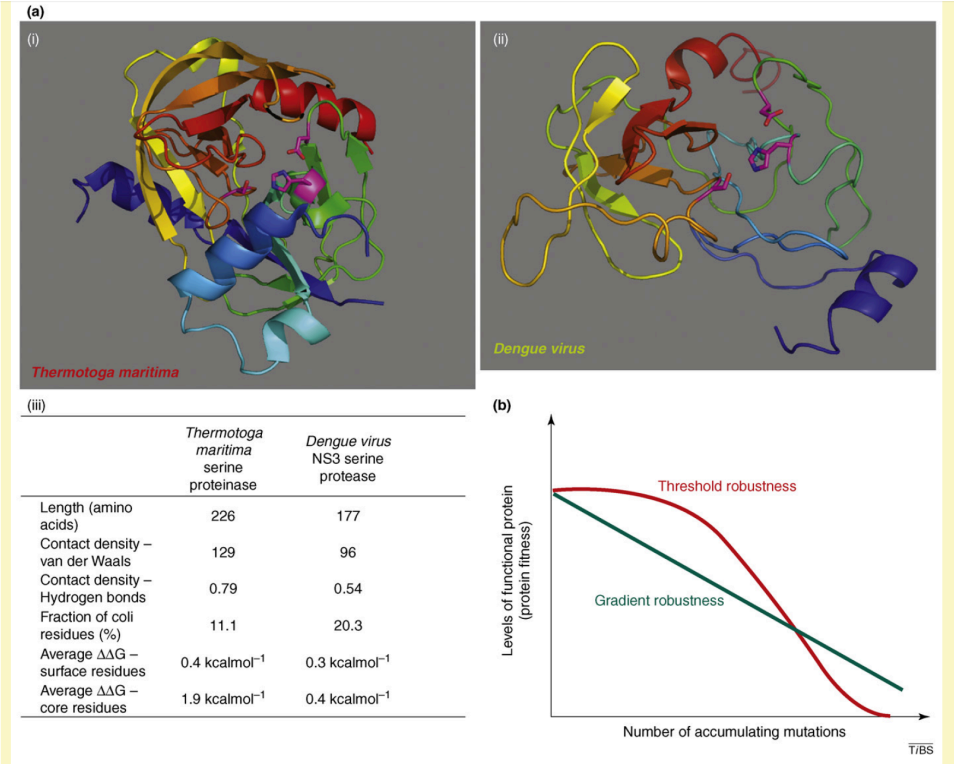
Secondary (local, short-range) packing – stem&loop, or protein helices

Tertiary – long-range, cooperative interactions dictating complex architectures

• The structural order-disorder paradox

Configurational stability is driven typically by a higher degree of structural order and compactness, and confers tolerance to mutations, i.e., higher robustness.

Higher **contact density** (average number of contacts per residue) i.e., intense interactions networks, result in higher robustness.



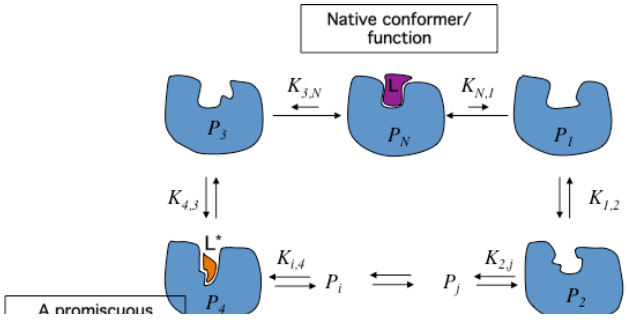
	Thermotoga maritima serine proteinase	*Dengue virus* NS3 serine protease
Length (amino acids)	226	177
Contact density – van der Waals	129	96
Contact density – Hydrogen bonds	0.79	0.54
Fraction of coli residues (%)	11.1	20.3
Average ΔΔG – surface residues	0.4 kcalmol^{-1}	0.3 kcalmol^{-1}
Average ΔΔG – core residues	1.9 kcalmol^{-1}	0.4 kcalmol^{-1}

However, this type of robustness (*threshold robustness*) is transient - once the margin of excess stability is sacrificed (typically within few mutations), the next mutations result in a large decline in the fraction or loss of soluble, functional protein molecules. Indeed, loosely packed structures, as seen in many viral proteins (right panel; figure below), may also confer higher tolerance to mutations by virtue of the destabilizing effects of mutations being weaker than in tightly packed proteins (*gradient robustness*).

Moreover, completely disordered regions exhibit the highest evolutionary rates (*i.e.*, higher robustness), and drift to the extreme of preserving only sequence compositions or patterns.

What about innovability? Mutations that promote new functions tend to be more destabilizing than those underlining drift. Consequently, higher degree of order (contact density, etc'), and thereby excess stability, promote the acquisition of new functions.

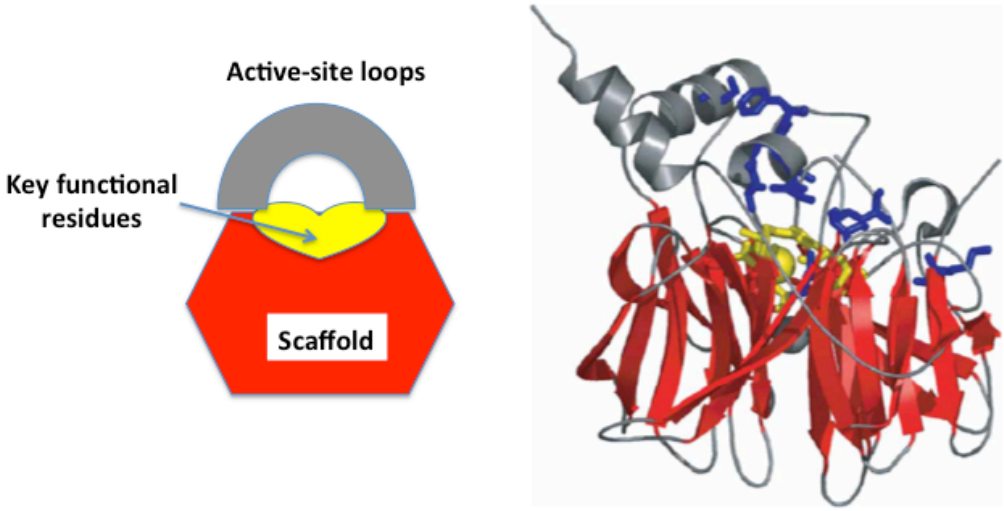
On the other hand, increased stability coincides with reduced conformational plasticity, and the acquisition of new functions depends on conformational plasticity.



Overall, it seems that higher contact density and structural order, seems to be correlated with highly-evolvable structures and folds – i.e., with both robustness and/or innovability. However, confusingly, structural disorder was also correlated with the very same properties – *i.e.*' high evolutionary rates, or robustness, as well as the acquisition of multiple functions, *i.e.*, innovability.

Do these conflicting factors coincide?

Although considered as one 'pack', ordered, globular folds show a different degree of structural order-disorder, and a different degree of structural separation, and may thus independently promote robustness and innovability.



Polarity therefore relates to the degree of modularity within the same fold (since the term "modularity" is routinely used to describe 'cut and paste' within multi-domain proteins, we opted for the term "polarity" instead).

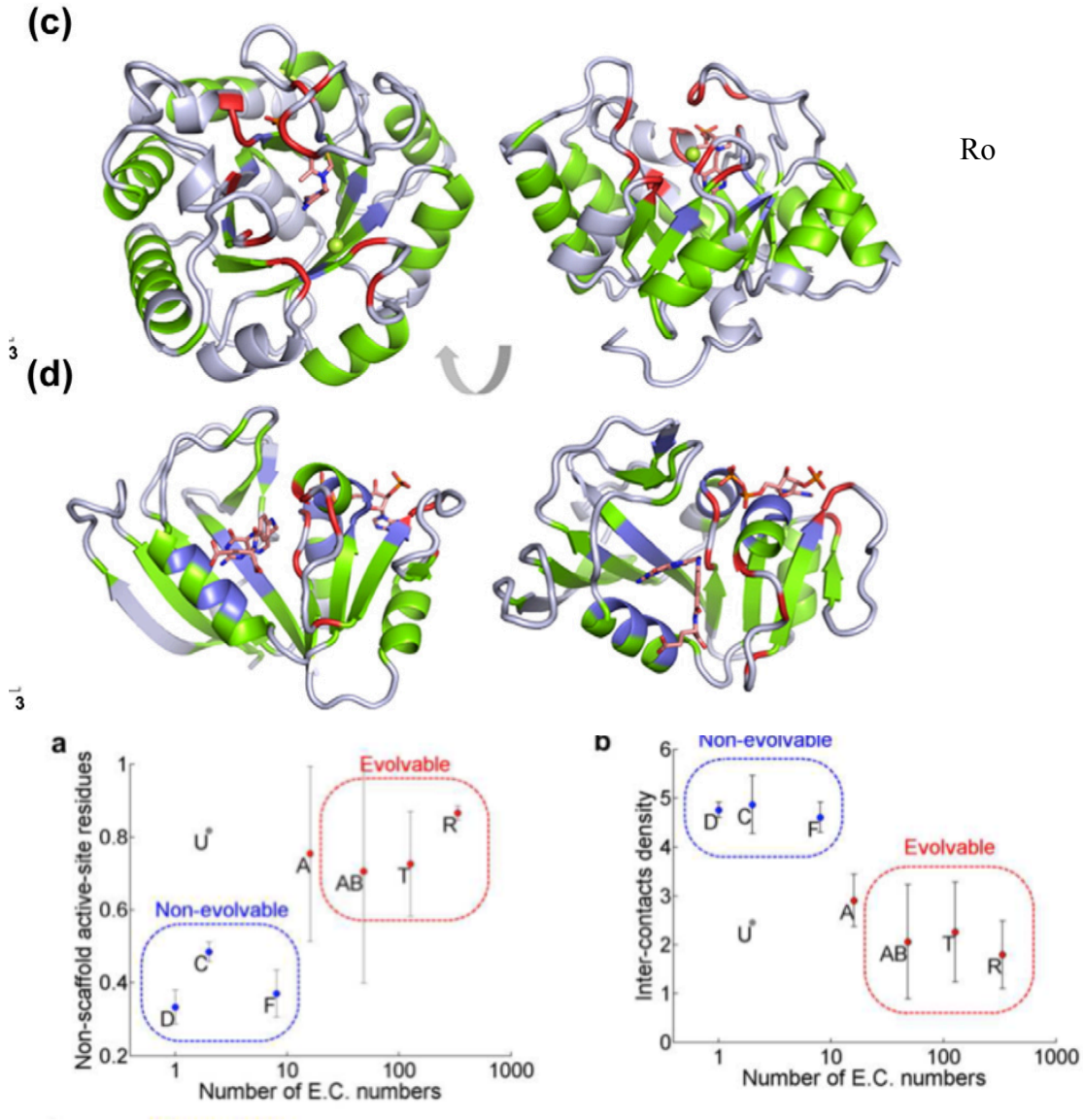
The polarity of a given structure can be assessed by two simple measures:

(i) The degree of overlap between the two residue classes: What is the fraction of active-site residues that are part of the scaffold?

(*ii*) What is the level of connectivity between scaffold and active-site residues: The average number of contacts between the active site and scaffold residues?

Scaffold – the array of secondary-structure elements that underlines a given fold.

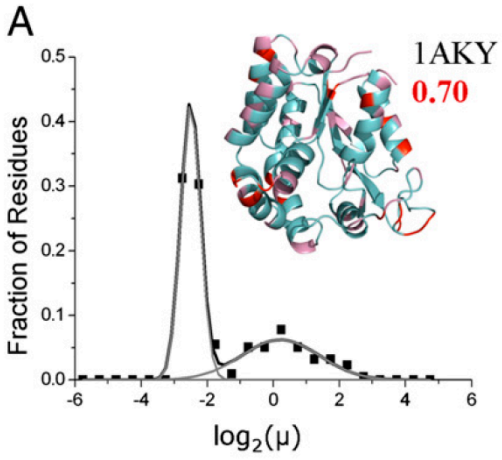
Scaffold–active-site polarity correlates with innovability



Ro

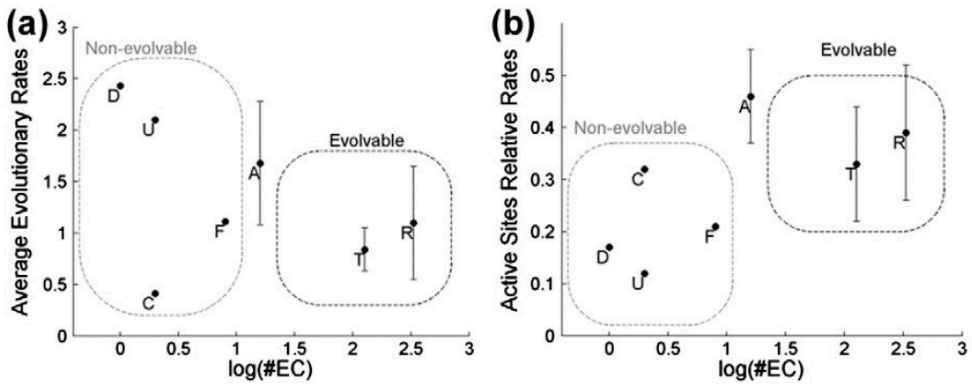
Robustness and innovability also correlate

Evolutionary rates can be obtained per position, not only per protein:

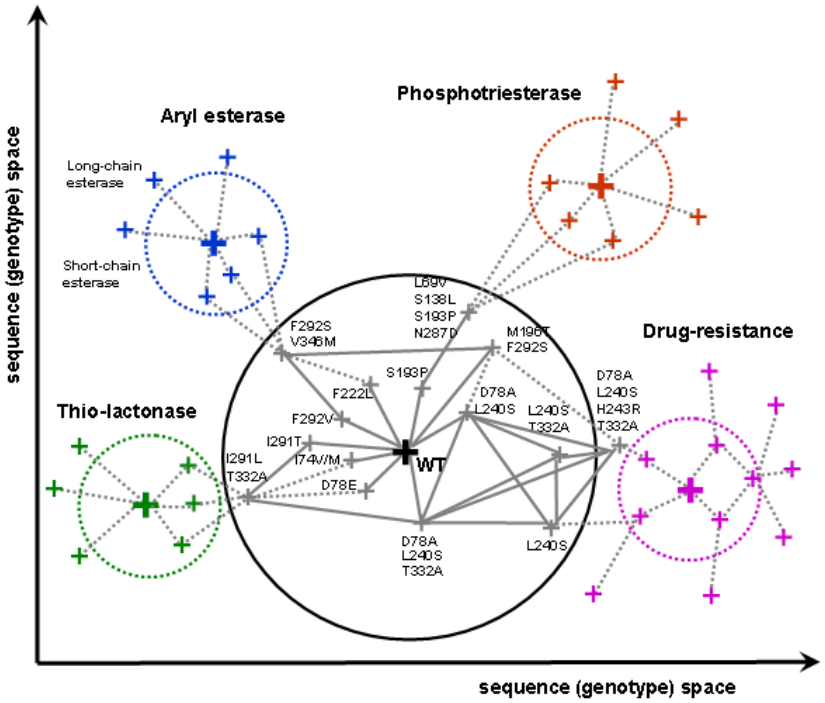


The overall evolutionary rates (per protein; *left panel*) do not correlate with

evolvability, but positional rates for the active-site residues do (right panel).



Higher tolerance for mutations, i.e., faster neutral drift, promotes the acquisition of new functions (notions of neutral networks, quasi-species, Maynard-Smith's mutational walk, etc').



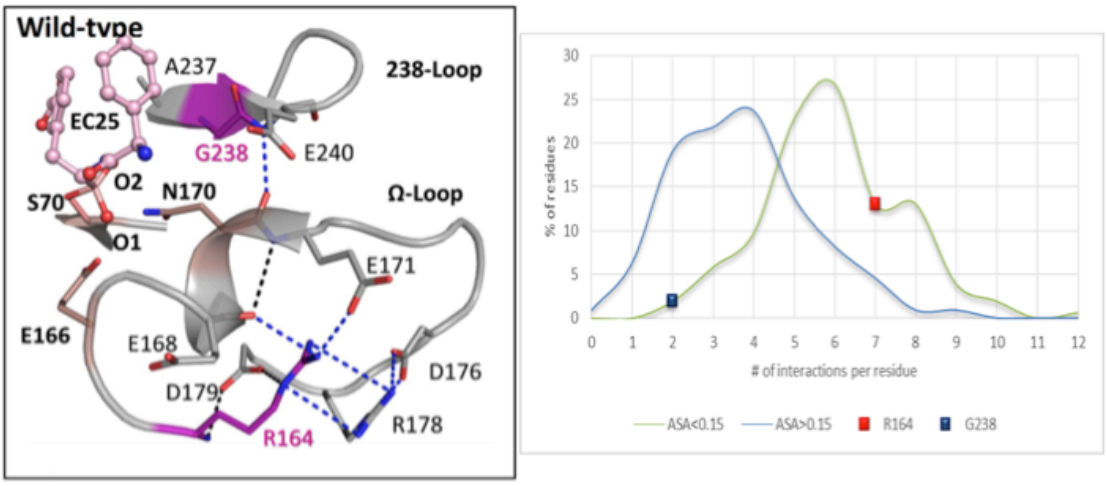
From macro- to micro-evolution?

What dictates the robustness and innovability of different residues within the same protein?

TEM-1, a beta-lactamase mediating antibiotics resistance. The earliest mutations dictate discrete, unbridgeable mutational paths that lead to the enzyme's adaptation

towards the 3rd-generation antibiotics cefotaxime. Two mutations are known, at positions 164 and 238, each leading to a different adaptive peak. The path underlined by 238 leads to a proficient, physiologically fit new protein, and organism, whilst the path initiated by mutations at 164 is a *cul-de-sacs*.

164 is highly networked residue:



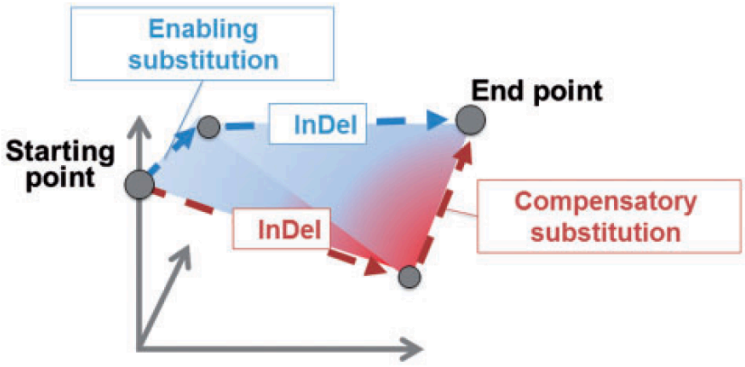
As can be seen in the distribution of the number of interactions per-residue, 164 is unusually networked, even for a core residue. Shown are distributions for core residues (accessible surface area, ASA≤0.15) and surface residues (ASA>0.15). Both R164 (red dot) and G238 (blue dot) are assigned as core residues, with ASA values of 0.10 and 0.11 respectively.

Accordingly, position 238 shows ~10-fold faster evolutionary rate than 164 (238 evolves 3 time slower than the protein's average, and R164 30-fold slower).

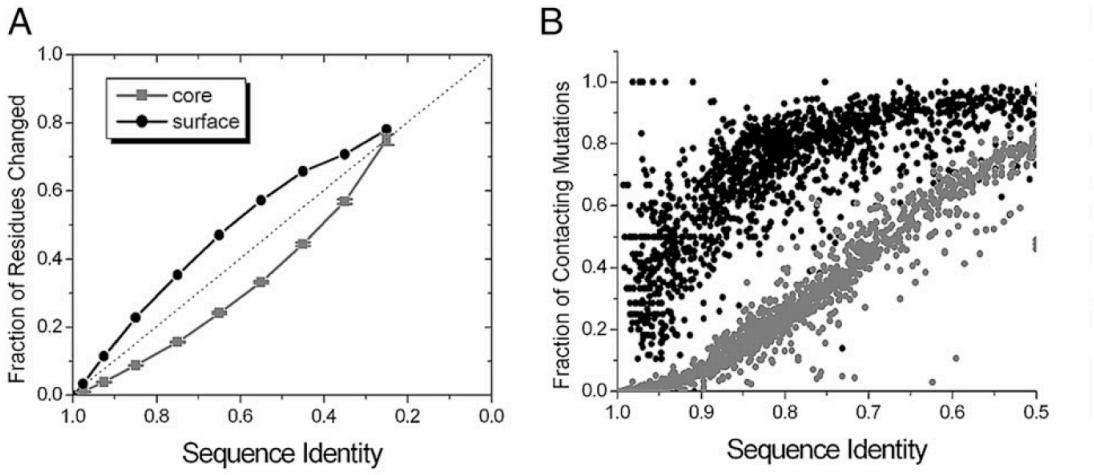
Epistasis and evolvability

Why do highly networked residues evolve slowly?

Basically, because exchanges in such residues are **highly correlated, or epistatic**, i.e., depend on sequence exchanges in other positions.



For example, exchanges in core residues occur very slowly, and seem to be dependent on exchanges in surface residues. Proteins in which the surface is highly constrained (e.g. due to function) the core does not evolve either – 'mutual freeze'.



Indeed, epistasis relates to the network properties of proteins, to their structural complexity, their architecture, whereby the latter is a consequence of functional demands.

In turn, the degree of epistasis dictates both robustness and innovability. So more polarized structures are expected to exhibit a lower degree of epistasis, and especially of exchanges in active-site residues depending on exchanges in scaffold ones.

Relevant references from our own work on protein robustness and evolvability

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D.S. (2006) Epistasis-robustness link shape the fitness landscape of a randomly drifting protein. *Nature*, **444**, 929-932.

Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D.S. (2007) The stability effects of protein mutations appear to be universally distributed. *Journal of Molecular Biology* **369**, 1318-1332.

Amitai, G. Devi, R. & Tawfik, D.S. (2007) Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP Journal* **1**, 67-78.

Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D.S. (2008) How protein stability and new functions tradeoff. *PLoS Computational Biology* **4**, e1000002.

Bershtein, S., Goldin, K. & Tawfik, D.S. (2008) Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of Molecular Biology* **379**: 1029-1044.

Bershtein, S. & Tawfik, D.S. (2008) Ohno's model revisited: Measuring the frequency of potentially adaptive mutations under various mutational drifts. *Molecular Biology and Evolution* 25: 2311-2318.

Tokuriki, N., Oldfield, C.J., Uversky, V.N., Berezovsky, I.N. & Tawfik, D.S. (2009) Do viral proteins possess unique biophysical features? *Trends in Biochemical Sciences* **34**: 53-59.

Goldsmith, M. & Tawfik, D.S. (2009) Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proceedings of the National Academy of the U.S.A.* **106**, 6197-6202.

Tokuriki, N. & Tawfik, D.S. (2009) GroEL/ES chaperonins promote genetic variation and accelerate enzyme evolution. *Nature* **459**, 668-673.

Tokuriki, N. & Tawfik, D.S. (2009) Protein dynamism and evolvability. *Science* **324**, 203-207.

Tokuriki, N. & Tawfik, D.S. (2009) Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology* **19**, 596-604.

Soskine, M. & Tawfik, D.S. (2010) Mutational effects and the evolution of new protein functions. *Nature Reviews Genetics* **11**, 572-582.

Tawfik, D.S. (2010) Messy biology and the origins of evolutionary innovations. *Nature Chemical Biology* **6**, 692-696.

Khersonsky, O. & Tawfik, D.S. (2010) Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annual Reviews of Biochemistry* ***79,*** 471-505.

Salverda, M.L.M., Dellus, E., Gorter, F.A., Debets, A.J. M., van der Oost, J, Hoekstra, R.F., Tawfik, D.S. & de Visser. J. A.G. M. (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS Genetics* **7**: e1001321.

Toth-Petroczy, A. & Tawfik, D.S. (2011) Slow protein evolutionary rates are dictated by surface-core association. *Proceedings of the National Academy of the U.S.A.* **108**: 11151-11156.

Tokuriki, N., Jackson, Afriat-Jornou, L., Wyganowski, K.T., Tang., R. & Tawfik, D. S. (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications*; **3**, 1257.

Toth-Petroczy, A. & Tawfik, D.S. (2013) Protein insertions and deletions enabled by neutral roaming in sequence space. *Molecular Biology & Evolution* **30**, 761-771.

Dellus-Gur, E., Toth-Petroczy, A., Elias, M. & Tawfik, D.S. (2013) What makes a protein fold amenable to functional innovation? Fold polarity and stability tradeoffs. *Journal of Molecular Biology* **425**, 2609-21.

Wellner, A., Raitses Gurevich, M & Tawfik, D.S. (2013) Mechanisms of Protein Sequence Divergence and Incompatibility. PLoS Genetics **9**, e1003665.

Toth-Petroczy, A. & Tawfik, D.S. (2014) The robustness and innovability of protein folds. *Current Opinion in Structural Biology*; in press.