

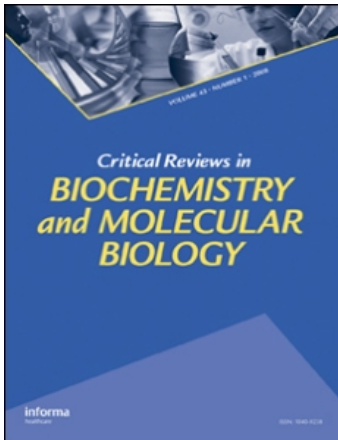
This article was downloaded by: [University of Washington]

On: 5 September 2008

Access details: Access Details: [subscription number 789063415]

Publisher Informa Healthcare

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Critical Reviews in Biochemistry and Molecular Biology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713609207>

### Genetic Constraints on Protein Evolution

Manel Camps <sup>a</sup>; Asael Herman <sup>b</sup>; Ern Loh <sup>b</sup>; Lawrence A. Loeb <sup>b</sup>

<sup>a</sup> Department of Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, California, USA <sup>b</sup> Joseph Goststein Memorial Cancer Research Laboratory, Department of Pathology, University of Washington, Seattle, Washington, USA

Online Publication Date: 01 September 2007

**To cite this Article** Camps, Manel, Herman, Asael, Loh, Ern and Loeb, Lawrence A. (2007) 'Genetic Constraints on Protein Evolution', *Critical Reviews in Biochemistry and Molecular Biology*, 42:5, 313 — 326

**To link to this Article:** DOI: 10.1080/10409230701597642

**URL:** <http://dx.doi.org/10.1080/10409230701597642>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Genetic Constraints on Protein Evolution

## Manel Camps

Department of Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, California, USA

## Asael Herman, Ern Loh, and Lawrence A. Loeb

Joseph Gosttstein Memorial Cancer Research Laboratory, Department of Pathology, University of Washington, Seattle, Washington, USA

**ABSTRACT** Evolution requires the generation and optimization of new traits (“adaptation”) and involves the selection of mutations that improve cellular function. These mutations were assumed to arise by selection of neutral mutations present at all times in the population. Here we review recent evidence that indicates that deleterious mutations are more frequent in the population than previously recognized and that these mutations play a significant role in protein evolution through continuous positive selection. Positively selected mutations include adaptive mutations, *i.e.* mutations that directly affect enzymatic function, and compensatory mutations, which suppress the pleiotropic effects of adaptive mutations. Compensatory mutations are by far the most frequent of the two and would allow potentially adaptive but deleterious mutations to persist long enough in the population to be positively selected during episodes of adaptation. Compensatory mutations are, by definition, context-dependent and thus constrain the paths available for evolution. This provides a mechanistic basis for the examples of highly constrained evolutionary landscapes and parallel evolution reported in natural and experimental populations. The present review article describes these recent advances in the field of protein evolution and discusses their implications for understanding the genetic basis of disease and for protein engineering *in vitro*.

**KEYWORDS** adaptation, evolvability, enzyme evolution, epistasis, positive selection

## INTRODUCTION

Understanding evolution at the molecular level is a central goal in modern biology. The rates of protein sequence evolution provide information about the contribution of individual proteins to fitness, the location of functional sites within these proteins, and are relevant to understanding genetic diversity in disease. The present article reviews recent advances in the field of protein evolution and describes their implications for protein engineering and for understanding the genetic basis of disease.

Mutations occur at an approximately constant rate. They can be fixed stochastically (by random drift), eliminated by negative or “purifying” selection, or they can be positively selected. The present discussion will focus exclusively on point mutations, specifically missense mutations, and will not consider other genetic mechanisms contributing to protein evolution such as duplications/shuffling and horizontal transfer. This discussion will also focus on natural evolution, and will only consider “*in vitro*” evolution as a model for natural evolution. Natural evolution (with the exception of viral *quasispecies*)

Address correspondence to Manel Camps, Department of Environmental Toxicology, University of California at Santa Cruz, 434 Physical Sciences Building, Santa Cruz, CA 95064, USA.  
E-mail: camps@etox.ucsc.edu

differs from evolution *in vitro* in that mutations are so rare that selection can only survey one mutant at a time. Therefore, in natural evolution fitness valleys limit the possible trajectories available for evolution.

## THE CHALLENGE OF MOLECULAR ADAPTATION

Biological functions have been optimized for the conditions under which they evolved. When these conditions change, existing activities need to be fine-tuned or new traits need to arise. This process, known as “adaptation,” occurs through the selection of mutations in genes controlling biological activities. Some sort of gain of function is usually involved, which is a challenge, considering that mutations occur randomly and that the sequence space is astronomically large.

The potential of biological activities or of specific genes for molecular adaptation is of critical evolutionary and biotechnological importance. This property, known as “evolvability”, is at least partially inherent to proteins. In general, evolvability increases with mutation robustness (the ability of proteins to tolerate mutations), because it allows a wider exploration of sequence space. Robustness may be intrinsic to a given protein, through locally (Bershtein *et al.*, 2006) or globally stabilizing mutations (Baroni *et al.*, 2004; Huang & Palzkill, 1997). It can also be provided extrinsically, through chaperones or other interacting proteins (Ellis, 2007). Promiscuity, the recognition of alternative substrates or catalysis of alternative chemical reactions, is likely a major mechanism of functional diversification because enhancing preexisting activities requires far fewer amino acid changes than generating new ones. Moreover, enhancement of a pre-existing activity generally maintains the original activity, which limits the costs of evolving a new trait on fitness (reviewed in (Khersonsky *et al.*, 2006)). Finally, modularity, *i.e.*, the presence of functionally independent motifs, also contributes to increased evolvability. Given that evolvability appears to be an intrinsic property of proteins, it could be subject to selective pressures. Protein designs that promote evolvability could be favored by evolution to facilitate adaptation to changing environments (Blazquez *et al.*, 2000).

Similarly, if the rate of mutagenesis is limiting for the generation of genetic diversity, it could be fine-tuned to the needs for adaptation of particular organisms (Lenski *et al.*, 2006; O’Loughlin *et al.*, 2006). In general, high

genomic mutation rates are disfavored because they increase the load of deleterious mutations. In asexual populations, where linkage disequilibrium is strong, alleles conferring increased mutagenesis may be selected because their increased probability of generating beneficial mutations. Indeed, adaptation has been documented to result in the selection of “mutator strains” both in culture (Mao *et al.*, 1997) and *in vivo* (Bjedov *et al.*, 2003). This selection, however, is very inefficient, as it operates at the population level rather than at the individual level. The effective selection for mutator alleles is further limited by competition between clones bearing different beneficial mutations, a phenomenon observed in experimental microbial populations and referred to as “clonal interference” (Arjan *et al.*, 1999). Thus, evolvability may be a selectable trait in asexual populations in situations in which beneficial mutations are extremely infrequent: in the presence of bottlenecks (Levin *et al.*, 2000), or when a population is initially well adapted (Arjan *et al.*, 1999). Alternatively, evolvability may be simply incidental, except in situations that increase the ratio of beneficial versus deleterious mutations such as antigenic variation in loci under intense immune pressure (Sniegowski & Murphy, 2006).

## INCONSISTENCIES OF THE “NEARLY NEUTRAL” THEORY

For a long time, the general assumption has been that adaptive mutations are selected from a pool of neutral or nearly neutral mutations, *i.e.*, mutations with little or no effect on fitness. The minimal effect of these mutations on the fitness of the organism would allow sequence space to drift enough to alter protein function so its performance can be adjusted to the demands of evolution. This view has been recently challenged by three different approaches: biophysical evidence that most missense mutations should affect protein function, phylogenetic evidence that the dispensability of proteins has little effect on rates of protein evolution, and the presence of signatures indicative of positive selection in the amino acid sequences of diverse genomes.

### Most Missense Mutations Affect Protein Function

Proteins appear to have little thermodynamic stability. Direct measurements of  $\Delta\Delta G$ , the difference in free energy between mutant and wild-type forms of an

enzyme, reveal that most proteins can only tolerate stability losses of between 3 and 10 kcal/mol (DePristo *et al.*, 2005), which is the energy of one or two hydrogen bonds. Too much stability may also decrease activity, as proteins “breathe” and require mechanical flexibility for catalysis (James & Tawfik, 2003). Such a tradeoff between stability and activity has been demonstrated in *in vitro* studies on the evolution of thermostability (Akanuma *et al.*, 1998) and drug resistance (Wang *et al.*, 2002) and may well be a more general phenomenon (DePristo *et al.*, 2005). Given that the effect of amino acid substitutions on  $\Delta G$  is in the order of 0.5 to 5 kcal/mol, a significant fraction of missense mutations would be expected to affect enzyme function (DePristo *et al.*, 2005). Screening and selection from libraries with random mutations provides a direct experimental approach to determine the probability that a random amino acid substitution inactivates the enzymatic activity of a given protein. This approach, however, is not sensitive to moderate or subtle effects on activity and therefore these deleterious mutation estimates represent only a lower limit. Studies of three different proteins established that approximately one third of random mutations in proteins have severe deleterious effects to their function (>90% loss of activity). Two of these are monomers, human 3-methyladenine DNA glycosylase (AAG, 298 amino acids long) (Guo *et al.*, 2004), and the 430 amino acids of the *E. coli* Pol I Klenow fragment (Loh *et al.*, 2007). The third is the *E. coli* lac repressor, a tetramer of four 360 amino acid polypeptides (Markiewicz *et al.*, 1994). The fact that these three studies independently estimated a probability of inactivation of  $\sim 33\%$  suggests that this may be an intrinsic property of individual proteins based on general principles of protein folding and solubility. This depends, however, on the specific threshold for inactivation (Bershtein *et al.*, 2006) and may vary depending on the size of the proteins, as smaller proteins have more surface exposed (Axe *et al.*, 1998; Rennell *et al.*, 1991).

Chaperones are proteins that assist protein folding and thus would be expected to suppress the deleterious effects of destabilizing mutations. Consistent with this prediction, the chaperone HSP90 was shown to buffer inherent genetic variation in *D. melanogaster* (Rutherford & Lindquist, 1998), and *A. thaliana* (Queitsch *et al.*, 2002). This means that a significant number of variants present in a population are nearly neutral under basal conditions but only with folding assistance. This observation agrees with the biophysical data discussed

above, indicating that most missense mutations should have an impact on fitness in the absence of chaperone activity (or under conditions where this activity may become limiting).

## Dispensability of Proteins Is Not a Major Determinant of Rates of Protein Evolution

Dispensability has only a marginal effect on rates of protein evolution in *S. cerevisiae* (Wall *et al.*, 2005) and in rodents (Hurst & Smith, 1999). Pleiotropy or “fitness density,” *i.e.*, the diversity of a protein’s functional interactions, which is another indicator of how dispensable a protein is, also appears to have only a limited effect on evolutionary rates (Hahn *et al.*, 2004; Jordan *et al.*, 2005; Salathe *et al.*, 2006), although with some conflicting results (Fraser *et al.*, 2002). These results are at odds with the “near neutrality” scenario, which predicts that, by setting the rate of purifying selection, the dispensability of proteins should determine how fast they evolve.

Particularly surprising is the apparent lack of correlation between the number of functional interactions of a given protein and its rate of evolution. Overlaying three-dimensional structural information on top of protein interaction-networks revealed two distinct types of hubs, depending on the number of binding surfaces they share: single-interface and multi-interface hubs (Kim *et al.*, 2006). Single-interface hubs are the more transient of the two, due to competition for ligands. Flexibility is also built in the linear motifs mediating these interactions, typically 3 to 10 amino acids in length, of which only 2 or 3 are critical for function. These linear motifs are therefore susceptible to easy inactivation, regeneration or modulation (Neduva & Russell, 2005). The transient nature and plasticity of these protein-protein interactions would allow selective pressures to operate on integrated functional complexes rather than on exact binary interactions. This is consistent with the rapid rate of change observed in eukaryotic interactomes, on the order of 100 to 1000 interactions per million years (Beltrao & Serrano, 2007), and would explain the limited correlation between pleiotropy and mutation rate (Beltrao & Serrano, 2007; Neduva & Russell, 2005). Multi-interface hubs, on the other hand, are more integrated into the network of the cell, more conserved, and more likely to be essential, more in accord with initial expectations and possibly explaining initial conflicting reports (Kim *et al.*, 2006).

## Phylogenetic Signatures of Positive Selection

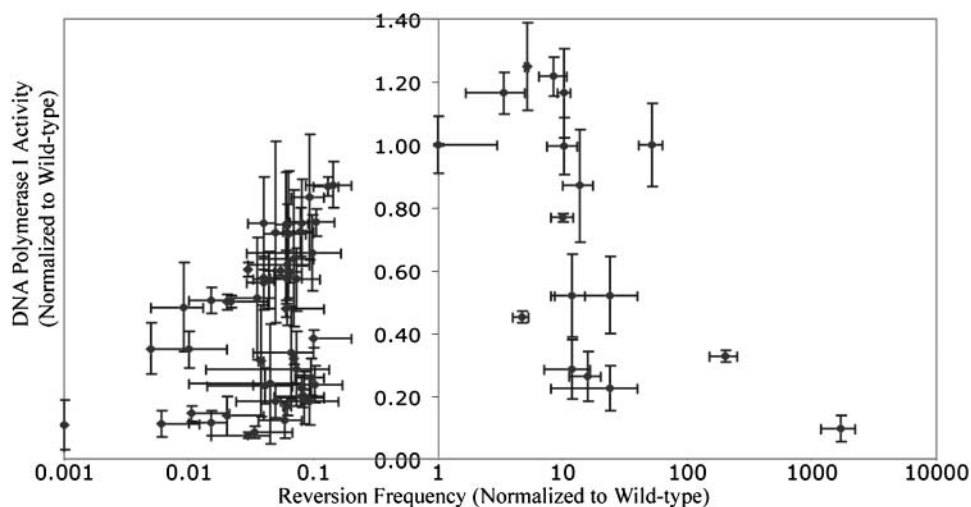
Phylogenetic and population genetic studies have revealed a sizable role of positive selection in shaping evolution. In phylogenetic studies, an increase in the ratio of nonsynonymous to synonymous substitutions (dN/dS) is indicative of positive selection, although this test has very low sensitivity and doesn't take into account possible contributions of negative selection. The most sensitive tests quantify amino acid divergence by normalizing the ratio of non-synonymous to synonymous mutations within a given species to the dN/dS ratio compared to a different but closely related species. An excess of amino acid divergence (suggestive of positive selection) was found comparing *D. megalogaster* and *D. simulans*, comparing different *D. simulans* strains, and comparing human and old world monkeys (Fay *et al.*, 2002). In *D. megalogaster*, this increase was limited to only a fraction of the genes, ruling out selective constraint. Using a more sophisticated variation of this test, Smith and Eyre-Walker (2002) estimated that in *Drosophila* 45% of amino acid substitutions have been fixed by positive selection. A different, highly sensitive test for positive selection relies on comparing changes in codons involving more than one position between closely related species. The increased presence of certain types of non-synonymous codons in the same lineage is evidence of positive selection. This telltale clumping of codons was found comparing the genomes of the

rat and the mouse, and comparing 12 pairs of bacterial genomes (Bazykin *et al.*, 2004).

## COMPENSATORY MUTATIONS AS KEY PLAYERS IN EVOLUTION

In sum, biophysical, phylogenetic and whole-genomic analyses point to a critical role of positive selection in shaping protein evolution. Positively selected mutations, however, are not necessarily adaptive, *i.e.*, do not necessarily improve protein function. They can maintain function in response to challenges, upholding the *status quo*. Examples include mutations selected by competing for limited resources or as part of "arms races" between hosts and pathogens (Orr, 2005).

Compensatory mutations would also fall into the category of "nonadaptive mutations" that are subject to positive selection. Adaptive mutations frequently have deleterious, pleiotropic effects, and these are suppressed by compensatory mutations. *E. coli* DNA polymerase I (Pol I) serves as an example to illustrate the pleiotropic effects of adaptive mutations and the need for compensatory mutations. Pol I is a highly accurate polymerase (1 error in  $10^5$  nucleotides) involved in DNA repair and in Okazaki fragment processing (reviewed in Camps & Loeb, 2005). The fidelity of a panel of Pol I active mutants and their level of activity was determined. Figure 1 shows that mutations in Pol I frequently increase or decrease polymerase fidelity relative to the wild type, and



**FIGURE 1** Fidelity versus overall activity of active mutants of *E. coli* DNA Polymerase I. 69 active mutants of *E. coli* DNA polymerase I were tested for fidelity and activity. Replication fidelity (X-axis) was established as the reversion frequency of a  $\beta$ -lactamase gene encoding an ochre stop codon. Polymerase activity (Y-axis) was determined by incorporation of radio labeled dTMP. Values have been normalized to wild-type values, with wild-type controls occupying the position 1,1. Error bars represent standard error  $p < 0.05$  (Loh *et al.*, 2007), with permission).

also that this change in fidelity comes at a cost in overall activity (Loh *et al.*, 2007), with permission). Thus, substantially changing the level of fidelity of Pol I would be expected to have pleiotropic effects, calling for compensatory mutations to restore the level of activity.

A remarkable convergence of directed evolution, population genetics, experimental adaptation, and whole genome comparison studies suggests that compensatory mutations may constitute the bulk of positively selected mutations. This proposition rests on at least 4 lines of argument:

- a) A high prevalence of intragenic suppressor mutations has been found in multiple genetic screens.
- b) The fact that deleterious mutations are sometimes context-dependent. Mutations that are deleterious in one protein may not be in a different ortholog, presumably due to the presence of additional, compensatory mutations.
- c) The presence of telltale signatures of compensation: “mutation bursts” due to the rapid fixation of multiple mutations during adaptation.
- d) Evidence that thermodynamic stability is limiting for evolution. Compensatory mutations increase the pool of mutants available for selection during adaptation.

### High Prevalence of Suppressor Mutations

In genetically tractable organisms, mutations that modify the phenotypic effects of a given mutation (known as “suppressor mutations”) have been used to reveal functional interactions both within and between proteins.

A significant fraction of suppressor mutations occurs between interacting residues. The presence of one mutation was found to significantly increase the chances of another mutation at a structurally interacting site. In the case of ionic interactions, the increase was almost four-fold (Choi *et al.*, 2005). This is strong evidence of a role of compensatory mutations in shaping protein evolution.

An attempt to quantify the prevalence of compensatory mutations that become fixed during periods of adaptation has recently been reported (Poon *et al.*, 2005). This study assumed that compensatory mutations are roughly equivalent to intragenic suppressor mutations in genetic screens and estimated that for

each deleterious mutation there are, on average, 12 compensatory mutations. This suggests that suppressing the deleterious effects of single, missense mutations often requires multiple mutations in the same protein, possibly because the deleterious effects are pleiotropic (DePristo *et al.*, 2005).

### Context-Dependent Deleterious Effects

Mutations that are pathogenic in one species are often fixed in another, suggesting that the deleterious effects of missense mutations have been compensated by other mutations (Kondrashov *et al.*, 2002; Kulathinal *et al.*, 2004; Weinreich & Chao, 2005); (see also discussion on sign epistasis below). This phenomenon, known as “compensated pathogenic deviation,” appears to be widespread. For example, Kondrashov *et al.* compared 32 mammalian proteins with amino acid sites producing pathogenic deviation in humans. Of these pathogenic mutations, 10% are present in at least one nonhuman mammal. A comparison of three complete dipteran genomes yielded similar results (Kulathinal *et al.*, 2004).

### Signatures of Compensation

Compensation needs to occur shortly after the deleterious mutation appears if it is to succeed in retaining the deleterious mutation in the population. Indeed, different lines of evidence show that compensatory mutations become fixed rapidly, creating signature “mutation bursts” in the sequence.

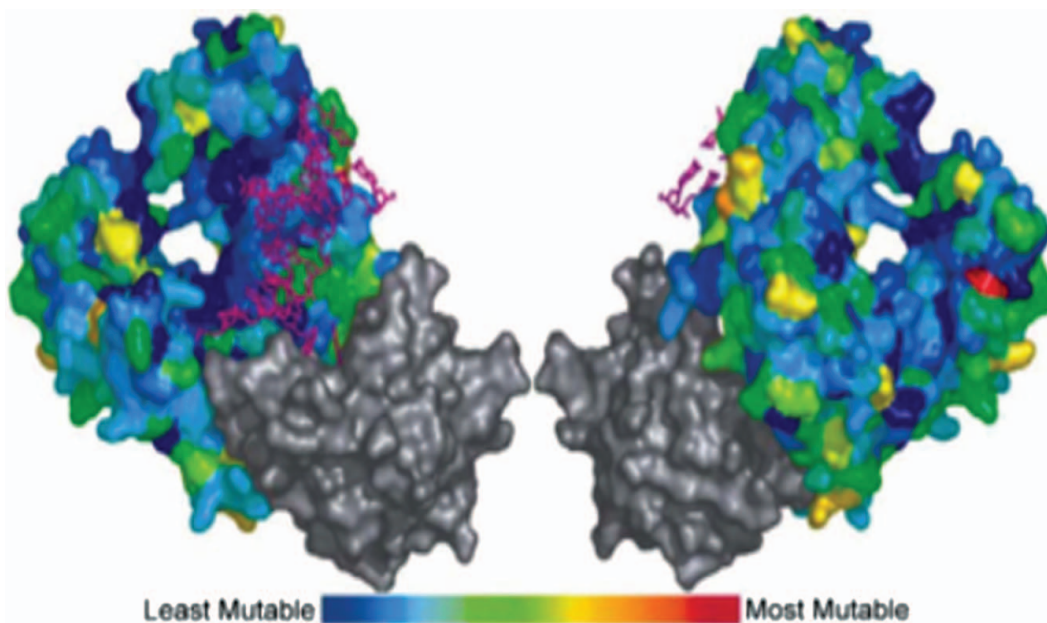
When closely related-species are compared, the bias of non-synonymous codons for individual species indicates that the changes occurred rapidly and successively (Bazykin *et al.*, 2004). Kondrashov’s study looking for the fixation of mutations that are deleterious in humans in other mammalian genomes found that compensating pathogenic variation stays constant over long phylogenetic distances (~10%) (Kondrashov *et al.*, 2002). Similarly, the rates of co-evolution between interacting residues are comparable regardless of whether mouse-rat-human or human-human-dog ortholog trios are used (Choi *et al.*, 2005). Thus, three different tests of positive selection independently found that positively-selected mutations occurred within a much shorter time frame than the evolutionary time separating these species.

Speciation involves extensive adaptation. Pagel *et al.* (2006) used a correlation between path lengths from root to tip of phylogenetic trees and the number of speciation events occurring along that path to estimate mutational bursts in DNA driven by speciation. They found that ~22% of substitutional changes fall into this category. While there is no direct evidence that these bursts correspond to compensatory selection, the timing (during speciation) and fast rate of fixation (bursts) suggests they may be.

## Thermodynamic Stability is Limiting for Evolution

Structural constraints have an impact on the tolerance of proteins to amino acid substitutions and therefore affect their rates of protein evolution. Several studies have shown that principles of protein folding and solubility constrain protein evolution in similar manners. In general, residues located on the surface of globular proteins are more tolerant of amino acid substitutions (Guo *et al.*, 2004; Loh *et al.*, 2007; Suckow *et al.*, 1996). The hydrophobic protein core tends to be less tolerant of mutations due to the need for stable packing

of atoms, to the limited availability of stabilizing bonds, and to the increased likelihood that these residues are involved in catalysis. The impact of relative location on the substitutability of individual amino acids is illustrated in Figure 2 and Table 1. Figure 2 shows the substitutability profile of the *E. coli* Pol I Klenow fragment presented as a color gradient, with blue indicating lowest and red indicating highest tolerance for substitutions (Loh *et al.*, 2007), with permission). The most highly substitutable residues locate predominantly on the surface of the protein. Secondary structure also has an effect on mutation tolerance. Table 1 presents the average substitutability indexes for different structural motifs of Pol I and also of the enzyme 3-methyladenine DNA glycosylase (AAG). Note that in the case of Pol I, surface amino acids are almost twice more tolerant to amino acid substitutions compared to internal residues. Consistent with this observation, surface residues were found to evolve approximately twice as fast as core ones (Goldman *et al.*, 1998). Secondary structure also affects the rate of evolution. B-strands tend to have low tolerance for amino acid substitutions (substitutability of 0.52 for AAG in Table 1) due to the requirement for secondary structure folding and for tertiary structure



**FIGURE 2** Substitutability of the polymerase domain of *E. coli* DNA Polymerase I. A random library of *E. coli* DNA Polymerase I was selected for activity using functional complementation of JS200 cells, a strain encoding a temperature-sensitive DNA polymerase I gene (*recA718 PolIA12*). For each position, a substitutability index was calculated based on the number of tolerated mutations. The crystal structure of the Pol I Klenow fragment (pdb file 1KLN) was colored according to a gradient, going from blue (lowest) to red (highest) tolerance for mutations. The 5' exonuclease domain of the Klenow fragment, which did not undergo mutagenesis, is colored in grey. (Modified from Loh *et al.*, 2007).

**TABLE 1** Substitutability indices of Pol I\* and AAG†.

Protein	Pol I		AAG	
	Substitutability Index	Normalized to average	Substitutability Index	Normalized to average
Average	0.68	1	1.38	1
Evolutionarily conserved	0.31	0.46	0.75	0.54
$\alpha$ elices	0.75	1.1	1.19	0.81
$\beta$ strands	0.61	0.90	0.73	0.53
Loops and Turns	0.81	1.2	1.57	1.1
Surface Residues	0.88	1.3	NA	NA
Internal Residues	0.46	0.69	NA	NA

Substitutability indices corresponding to specific regions of proteins. The number of mutations in each region after selection *in vivo* was divided by the total number of mutants sequenced. For the purpose of comparing the two proteins, substitutability index was then normalized to the average of each protein. For each protein, the left column represents the actual substitutability index and the right column the value normalized to the average substitutability index of the protein.

\* (Loh *et al.*, 2007).

† (Guo *et al.*, 2004).

interactions. Pol I appears to be an exception ( $\beta$ -strand substitutability of 0.9), but this can be attributed to the fact that Pol I  $\beta$ -strands, located in the palm subdomain, are highly exposed to solvent. The substitutability of mobile regions of the protein, such as loops and turns, depends largely on whether catalytically active residues are present or whether they adopt a specific secondary or tertiary structure. Both studies found significant disparities between evolutionary conservation and mean substitutability, presumably because artificial selection in culture of proteins expressed from multicopy plasmids likely doesn't accurately recapitulate the stringencies of natural selection (Guo *et al.*, 2004; Loh *et al.*, 2007).

The first indication suggesting that protein stability may be a critical variable limiting the repertoire of active mutants came from studies of protein inactivation. In these studies, misfolding was found to be the main cause of protein inactivation rather than loss of catalytic activity (Loeb *et al.*, 1989; Pakula *et al.*, 1986). The role of thermodynamic stability in limiting mutation tolerance has been confirmed in later studies using thermostable proteins and modeling *in silico*. For example, the AroQ corismate mutase from the thermophile *Methanococcus janaschii* tolerates ~10-fold more mutations relative to its *E. coli* counterpart (Besenmatter *et al.*, 2007). Parisi and Echave (2005) modeled protein evolution under stability constraints. At each step, this sequential *in silico* model introduced mutations and selected against structural perturbation. Strikingly, the resulting amino

acid conservation patterns resemble those found in natural proteins (Parisi & Echave, 2005). The implication of these studies is that thermodynamic stability requirements may severely restrict the evolvability of a given protein. Direct, experimental proof of this concept has been provided by Bloom *et al.* (2006). This elegant work demonstrated that thermostable variants of cytochrome p450 are more likely to evolve new or improved functions relative to the (marginally thermostable) wild-type (Bloom *et al.*, 2006).

Overall, these four different lines of evidence suggest that positive selection as a force driving evolution had previously been underestimated. Further, while mutations that generate new traits during adaptation (adaptive mutations) are the ones driving positive selection, adaptation involves a large number of concomitant compensatory mutations. These compensatory mutations appear to have been positively selected to suppress pleiotropic deleterious effects of adaptive mutations. The deleterious effects of adaptive mutations likely stem from the narrow window of thermodynamic stability of proteins, and the nonspecific, pleiotropic nature of the effects frequently calls for multiple suppressor mutations, often within the same gene product.

## SIGN EPISTASIS CONSTRAINS EVOLUTIONARY PATHWAYS

The effects of individual mutations may depend on the genetic context in which they occur. This



**TABLE 2** Epistatic effects of  $\beta$ -lactamase mutations on aztreonam resistance.\*

$\beta$ -lactamase mutant	Fold aztreonam resistance (wild type = 1)
Deletion	0.6
Wild-type	1.00
E104K	2.7
R164H	2.2
G267R	0.8
E104K R164H	43.7
E104K G267R	2.9
R164H G267R	2.7
E104K R164H G267R	67.8

\*Modified from Camps *et al.*, 2003.

phenomenon, known as epistasis, is illustrated in Table 2. This table presents several mutants of *E. coli*  $\beta$ -lactamase and the levels of resistance they confer to aztreonam, a monobactam antibiotic that is not the preferred substrate for  $\beta$ -lactamase. These mutants have different combinations of the following three mutations: E104K, R164H, and G267R. Mutations E104K and R164H by themselves increase resistance to aztreonam by  $\sim 2.5$ -fold. In combination, though, their effect on resistance is 40-fold. The third mutation, G267R has no effect on its own or in the presence of E104K or R164H, but it doubles the level of aztreonam resistance in the presence of both E104K and R164H (Camps *et al.*, 2003). In this case, all three mutations have epistatic effects because their effect on aztreonam resistance varies depending on the presence of the other two mutations.

There are two types of epistasis: magnitude epistasis and sign epistasis. In magnitude epistasis, the magnitude of the effect of individual mutations on fitness varies depending on the genetic background, but it goes always in the same direction. In the example presented above, E104K and R164H would fall into this category, as they always have a positive effect on aztreonam resistance. In sign epistasis, not only the magnitude, but the sign of the effect (*i.e.*, positive, negative or neutral) changes depending on the genetic context (Weinreich & Chao, 2005). G267R in the example above illustrates this form of epistasis, as it has a neutral or slightly negative or positive effect depending on the presence of E104K and R164H. Sign epistasis limits the number of mutational trajectories available to selection because some paths to an optimum contain fitness decreases (Weinreich & Chao, 2005). This was shown experi-

mentally in the model enzyme  $\beta$ -lactamase. This study investigated all possible mutational pathways leading to five point mutations controlling resistance to cefotaxime. Strikingly, of the 120 possible direct mutational trajectories linking these alleles, only 18 were found to be accessible to selection (Weinreich *et al.*, 2006). These constraints on the mutational trajectories matched the structure of sign epistasis of the five mutations.

In this study, the mechanistic basis of sign epistasis was traced back to one compensatory mutation, M182T. Alone, this mutation modestly reduced cefotaxime hydrolysis. M182T, however, suppressed the reduced thermodynamic stability associated with G238S, the mutation increasing cefotaxime hydrolysis. Thus, M182T has a dramatically different effect on cefotaxime resistance depending on the presence or absence of G238S. As illustrated by the M182T mutation, compensatory mutations are expected to exhibit frequent sign epistasis because they are selected to suppress effects of other mutations, which makes them context-dependent.

Sign epistasis associated with compensatory mutations arising during periods of adaptation has the following two implications: a) It “locks in” deleterious mutations, precluding reversion to wild-type sequence; b) It constrains possible trajectories to an optimum, dramatically limiting genetic diversity resulting from adaptation.

## Reduced Reversion to Wild-Type Sequence

The presence of compensatory mutations rapidly creates selective valleys that preclude reversion to the ancestral, wild-type sequence. In phage, fixation of compensatory mutations was estimated to be twice as likely as reversion to wild-type sequence (Poon & Chao, 2005). In experimental microbial cultures, mutants selected under drug pressure often exhibit reduced fitness. Examples include HIV resistance to protease inhibitors (Borman *et al.*, 1996), streptomycin resistance in *E. coli* or *Salmonella* (Maisnier-Patin *et al.*, 2002; Schrag *et al.*, 1997), isoniazid or rifampicin resistance in mycobacteria (Gillespie, 2001), fucidin resistance in *Staphylococcus aureus* (Nagaev *et al.*, 2001), and resistance to fluconazole in *Saccharomyces cerevisiae* (Anderson *et al.*, 2003). In all these cases, growth in the absence of selective pressure resulted in a partial compensation of the fitness defect but not in reversion, indicating the presence of additional (presumably compensatory) mutations that

created an adaptive valley before the original mutation had a chance to revert.

## Constrained Evolutionary Trajectories

As discussed above, sign epistasis associated with compensatory mutations severely limits the number of evolutionary pathways available for selection. This has two consequences: it restricts the diversity of mutants coming out of positive selections and it increases the reproducibility of adaptation, at least under identical conditions and with a large population.

Selections for drug resistance typically yield a very limited number of mutants. For example, only 8 extended-spectrum mutants of  $\beta$ -lactamase (out of more than 90 known mutants from clinical isolates) were obtained in a selection *in vitro* under conditions resembling natural selection, and many of them shared mutations (Barlow & Hall, 2003). Another example of limited allelic representation following positive selection is an experiment replacing the thermostable adenylate kinase of *Geobacillus stearothermophilus* (a thermophilic organism) with adenylate kinase from *Bacillus subtilis* (a mesophile) to monitor adaptation to growth at high temperature at the level of a single gene. Only six alleles exhibiting increased thermal stability were observed, representing less than 1% of the total possible (Counago *et al.*, 2006). In both cases, the observed limited allelic representation likely reflects restrictions in the pathways available for selection, as each allele involves more than one mutation and a much larger number of mutants producing the desired effects is known. Multiple selective pressures, simultaneous or sequential, further restrict the outcome of positive selections. In the case of  $\beta$ -lactamase, this scenario would arise with alternating exposure to different  $\beta$ -lactam antibiotics in the clinic. Selections using a single antibiotic typically result in the isolation of resistant mutants that are different from those isolated in the clinic (Blazquez *et al.*, 2000; Orenca *et al.*, 2001). Exposing a culture of *E. coli* to amoxicillin and ceftazidime, however, resulted in the isolation of only naturally occurring  $\beta$ -lactamase mutants (Blazquez *et al.*, 2000), suggesting that “fluctuating selection” likely contributed to restricting the allelic repertoire observed in clinical isolates. Fluctuating selection should favor “generalist” mutations, *i.e.*, mutations that increase resistance to multiple antibiotics. More intriguingly, it would also select mutations with strong positive epistatic effects for resistance to one of

the antibiotics, even if its effect *versus* other antibiotics bacteria are regularly exposed to is neutral or detrimental (Blazquez *et al.*, 1998). This could be an example of a protein whose evolvability is shaped by selective pressures.

Parallel evolution, *i.e.*, the independent occurrence of the same substitution in two independent lineages has been observed in natural and experimental populations of insects, bacteria and phage. It is commonly seen in selections for drug resistance under conditions that mimic natural evolution (Anderson *et al.*, 2003; Barlow & Hall, 2003). Similarly, during adaptation of adenylate kinase to thermostability, the same three major mutants were observed in two independent experimental runs (Counago *et al.*, 2006). However, the most striking examples of parallel evolution come from adaptation studies in two closely related phages of *E. coli*,  $\Phi$ X174 and S13. Adaptation of  $\Phi$ X174 to higher temperature and to a new host (*Salmonella*) resulted in 50% identical changes between independent lineages (Wichman *et al.*, 1999) and long-term adaptation to culture in the laboratory resulted in 40% identical independent changes (Wichman *et al.*, 2005). In contrast, no such reproducibility has been observed in more complex situations such as the adaptation of *E. coli* to growth in liquid culture (Woods *et al.*, 2006) or to growth in glycerol (Herring *et al.*, 2006), or in the development of human cancer (Thomas *et al.*, 2007). This lack of reproducibility is likely due to the plasticity built into networks of functional interactions.

## CONCLUSIONS

A recent convergence of directed evolution, population genetics, genomics analysis and experimental adaptation experiments have challenged traditional notions about protein evolution. Positive selection has been recognized as an important driver of evolution. A large portion of positively selected mutations appears to be compensatory, suppressing deleterious pleiotropic effects of adaptive mutations.

The realization that a significant fraction of adaptive mutations have detrimental, pleiotropic effects underscores the role of deleterious mutations in the generation of novel traits during evolution. While neutral or nearly neutral mutations should have a long lifespan within a given population, they should rarely generate novel properties because of smaller phenotypic effects. In other words, neutral mutations are neutral

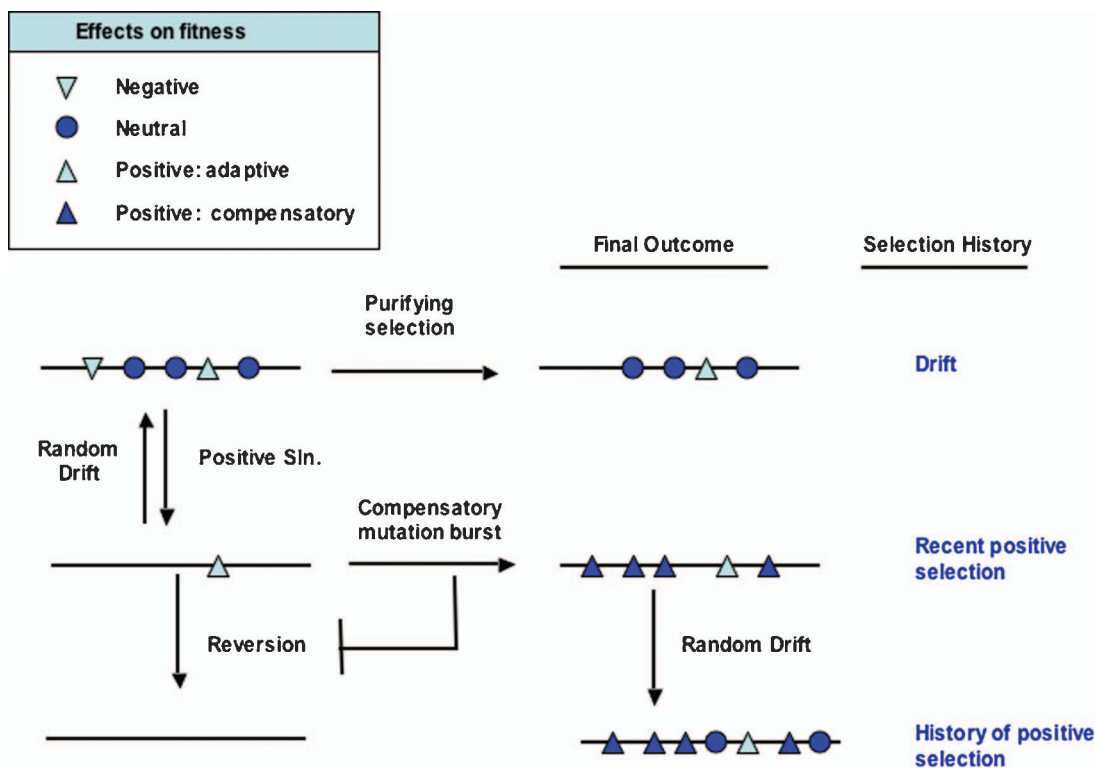
because they have little effect on function (Lenski *et al.*, 2006). On the other hand, pleiotropic mutations, while they may be short-lived in the population because of purifying selection, are more likely to modify protein function. The relative contributions to evolution of nearly neutral mutations versus pleiotropic ones may thus depend on a trade-off between the lifespan of missense mutations in the population and their functional impact. Compensatory mutations would play a key role in tipping the balance toward pleiotropic mutations by allowing them to persist longer in the population.

This emerging picture of protein evolution is summarized in Figure 3 and Table 3. Mutations that generate novel traits (adaptive mutations) often have pleiotropic effects. Deleterious effects of adaptive mutations are subsequently suppressed through the fixation of compensatory mutations. Typically, suppression occurs within the time frame of adaptation and often involves multiple compensatory mutations. These com-

pensatory mutations facilitate adaptation to novel environments by precluding reversion to wild-type once selective pressures are removed and by prolonging the lifespan of adaptive mutations in the population. The telltale signs of neutral, positive and negative selection, and the implications of each type of selection history for protein evolution are summarized in Table 3.

These new concepts in protein evolution have far-reaching implications:

- 1) The contribution of deleterious mutations to protein evolution has likely been underestimated. Indeed, most amino acid substitution prediction methods estimate that between 20% and 25% of nonsynonymous nucleotide sequence polymorphisms (SNPs) in the human genome significantly reduce protein function (Ng & Henikoff, 2006; Yue & Moulton, 2006). Consistent with the predicted deleterious nature of these SNPs, they tend to be rare, which suggests that they are undergoing purifying selection (reviewed in



**FIGURE 3** Effects of mutation fitness on protein evolution. The lifespan of mutations in the population depends on their effects on the fitness of the organism. Neutral mutations, having little effect on fitness, accumulate and persist for a long time in the population (although they may be eliminated *via* recombination). Deleterious mutations are rapidly eliminated by purifying selection. Adaptive mutations often have pleiotropic deleterious effects because they affect residues that are important for function. In this case, compensatory mutations rapidly arise, typically in bursts of multiple mutations, allowing adaptive mutations to persist longer in the population. The presence of compensatory mutations has an epistatic effect that generally precludes a return to the ancestral sequence. In a genome with a significant history of positive selection such as the human genome, the final contribution of positive and negative selection likely reflects a trade-off between persistence in the population (*neutral mutations*) and the need to generate novel traits (*adaptive mutations*). Some adaptive mutations with potential deleterious effects remain in the population at low frequencies because compensatory mutations delay purifying selection.

**TABLE 3** Selection history: signatures and implications for protein evolution

Selection	Type	Signature	Epistasis	Implications
Neutral		NS/S = 0.3	No	Long lifespan
Negative		NS/S < 0.3	Negative (moderate)*	Short lifespan Key functional residues
Positive		NS/S > 0.3 Increased aa divergence Codon clustering Increase in rare SNPs		
	Adaptive		Magnitude	Functional constraints Pleiotropic effects
	Compensatory	Mutation bursts	Sign, strong	Biophysical constraints Constrained trajectories Directionality Reproducibility

Summary of the effect of each type of selection: neutral, negative and positive on the ratio of non-synonymous to synonymous (NS/S) mutations and other telltale signatures of selection. Column 4 outlines the epistatic effects associated with each type of selection. Column 5 lists key concepts such as average lifespan of a given mutation in the population, the types of residues most frequently affected, and types of genetic constraints associated with each type of selection.

\*Bershtein *et al.*, 2006.

- Ng & Henikoff, 2006). This scenario agrees with the view that most spontaneous mutations are deleterious and that their removal by negative selection is delayed by the rapid fixation of compensatory mutations. The proposition that deleterious mutations facilitate adaptation provides an evolutionary explanation for their abundance in the human genome that is more general and therefore more plausible than overdominance, *i.e.*, a fitness advantage conferred by the deleterious variants in the heterozygous state.
- 2) The predicted abundance of deleterious SNPs, if confirmed, has profound implications for understanding genetic disease. It suggests that the contribution of deleterious mutations to “complex” (*i.e.*, non-mendelian) disease is more significant than anticipated and that the genetic component in these conditions may have been underestimated because they are caused by different alleles in different individuals.
  - 3) The newly recognized relevance of thermodynamic stability for function has obvious implications for protein modeling and engineering. It has dramatically improved our ability to identify critical functional residues in a protein sequence and to predict the functional impact of mutations (Ng & Henikoff, 2006; Stone & Sidow, 2005). Enzymatic stability considerations should become key for protein engineering. Consensus design based on phylogenetic comparison has already been used to improve enzyme stability (Forrer *et al.*, 2004; Watan-

abe *et al.*, 2006) and thermostable enzymes will likely be chosen as templates in preference to their mesophile counterparts in future directed evolution experiments because of their increased evolvability (Besenmatter *et al.*, 2007; Bloom *et al.*, 2006).

- 4) Sign epistasis conferred by compensatory mutations confounds the information that can be gathered from phylogenetic comparisons. This explains the difficulty in obtaining meaningful insights into the structural determinants of thermal stability through sequence comparisons between enzymes naturally adapted to low, middle, and high temperature (Wintrode & Arnold, 2000) and likely limits the specificity of methods to detect deleterious SNPs that are based solely on phylogenetic comparisons (Ng & Henikoff, 2006).
- 5) Sign epistasis, by reducing the chances of going back to an ancestral sequence even if the original selective pressure driving adaptation is removed, gives directionality to evolution. By constraining the trajectories available for evolution, sign epistasis also restricts genetic variability. This suggests that in natural or experimental populations large parts of the sequence space are left unexplored. This has obvious implications for drug resistance. In order to make predictions about the likelihood of drug resistance based on evolving mutants *in vitro*, we need to understand the molecular pathways available in natural evolution (Hall, 2004). Sign epistasis also has implications for the design of

directed evolution experiments. Methods that minimize the likelihood of fitness valleys such as using high mutation loads (Orenca *et al.*, 2001) or shuffling orthologous sequences (Wang *et al.*, 2006) may lead to solutions not found in nature.

- 6) Extreme evolutionary constraints associated with sign epistasis have been proposed to focus genetic variability in a few residues, at least in small genomes such as phage or influenza (Wichman *et al.*, 2000). Given the limited number of residues that would *de facto* show variation in these organisms, these residues would control fitness at multiple levels, including host range, and susceptibility to coinfection and to the immune response. How general this phenomenon is remains to be established, but it would have profound implications for understanding the genetic determinants of virulence and of immunogenicity in pathogens.
- 7) Cancer progression has been recognized as a process akin to speciation and subject to the rules of ecology and population genetics (Merlo *et al.*, 2006). Large-scale sequencing of tumor samples is becoming a reality, and genes involved in adaptation to malignancy (CAN) genes have been identified (Sjoblom *et al.*, 2006). Currently efforts are directed at distinguishing “driver” (*i.e.*, adaptive) versus “passenger” (*i.e.*, hitchhiking, neutral mutations) mutations amongst the mutations identified in tumors (Greenman *et al.*, 2007). We predict that a significant fraction of mutations identified as “driver” based on evidence of positive selection will turn out to be compensatory. Consistent with this prediction, an analysis of more than 1000 somatic mutations in coding exons of 518 protein kinases found many driver mutations outside the kinase domains (Greenman *et al.*, 2007). We also predict that CAN genes with a gain-of-function will carry multiple mutations, one adaptive mutation and other compensatory mutations added subsequently.

Rapid progress is being made in the area of protein evolution. A clearer picture of protein evolution that integrates observations from a variety of biological disciplines should emerge in the not too distant future. It would be extremely useful to find ways to distinguish between adaptive and compensatory mutations. This may be based on biophysical predictions, on the presence of some sort of sequence signature, or on a combination of biophysical and phylogenetic methods. We

also look forward to extension of what we have learned at the level of single proteins and interactomes to higher levels of cellular organization.

## ACKNOWLEDGMENTS

The authors would like to thank Drs. David Haussler (University of California Santa Cruz, CA), Ann Blank (University of Washington, Seattle, WA), and Eddie Fox (St. Vincent’s University Hospital, Dublin, Ireland) for critical reading on the manuscript and for useful comments, and Cole Bower for help with the manuscript. Support was obtained from the National Institute of Health grants K08 CA116429 (MC), CA102029 (LAL), CA788885 (LAL), and from a fellowship from the Cora May Poncin Foundation (EL).

## REFERENCES

- Akanuma, S., Yamagishi, A., Tanaka, N., and Oshima, T. 1998. Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. *Protein Sci* 7:698–705.
- Anderson, J.B., Sirjusingh, C., Parsons, A.B., Boone, C., Wickens, C., Cowen, L.E., and Kohn, L.M. 2003. Mode of selection and experimental evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. *Genetics* 163:1287–1298.
- Arjan, J.A., Visser, M., Zeyl, C.W., Gerrish, P.J., Blanchard, J.L., and Lenski, R.E. 1999. Diminishing returns from mutation supply rate in asexual populations. *Science* 283:404–406.
- Axe, D.D., Foster, N.W., and Fersht, A.R. 1998. A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease. *Biochemistry* 37:7157–7166.
- Barlow, M. and Hall, B.G. 2003. Experimental prediction of the natural evolution of antibiotic resistance. *Genetics* 163:1237–1241.
- Baroni, T.E., Wang, T., Qian, H., Dearth, L.R., Truong, L.N., Zeng, J., Denes, A.E., Chen, S.W., and Brachmann, R.K. 2004. A global suppressor motif for p53 cancer mutants. *Proc Natl Acad Sci USA* 101:4930–4935.
- Bazykin, G.A., Kondrashov, F.A., Ogurtsov, A.Y., Sunyaev, S., and Kondrashov, A.S. 2004. Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. *Nature* 429:558–562.
- Beltrao, P., and Serrano, L. 2007. Specificity and Evolvability in Eukaryotic Protein Interaction Networks. *PLoS Comput Biol* 3:e25.
- Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D. S. 2006. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–932.
- Besenmatter, W., Kast, P., and Hilvert, D. 2007. Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins* 66:500–506.
- Bjedov, I., Tenailon, O., Gerard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., and Matic, I. 2003. Stress-induced mutagenesis in bacteria. *Science* 300:1404–1409.
- Blazquez, J., Negri, M.C., Morosini, M.I., Gomez-Gomez, J.M., and Baquero, F. 1998. A237T as a modulating mutation in naturally occurring extended-spectrum TEM-type beta-lactamases. *Antimicrob Agents Chemother* 42:1042–1044.
- Blazquez, J., Morosini, M.I., Negri, M.C., and Baquero, F. 2000. Selection of naturally occurring extended-spectrum TEM beta-lactamase variants by fluctuating beta-lactam pressure. *Antimicrob Agents Chemother* 44:2182–2184.

- Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103:5869–5874.
- Borman, A.M., Paulous, S., and Clavel, F. 1996. Resistance of human immunodeficiency virus type 1 to protease inhibitors: selection of resistance mutations in the presence and absence of the drug. *J Gen Virol* 77 (Pt 3):419–426.
- Camps, M., Naukkarinen, J., Johnson, B.P., and Loeb, L.A. 2003. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc Natl Acad Sci USA* 100:9727–9732.
- Camps, M. and Loeb, L.A. 2005. Critical role of R-loops in processing replication blocks. *Front Biosci* 10:689–698.
- Choi, S.S., Li, W. and Lahn, B.T. 2005. Robust signals of coevolution of interacting residues in mammalian proteomes identified by phylogeny-aided structural analysis. *Nat Genet* 37:1367–1371.
- Counago, R., Chen, S., and Shamoo, Y. 2006. In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol Cell* 22:441–449.
- DePristo, M.A., Weinreich, D.M., and Hartl, D.L. 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 6:678–687.
- Ellis, R.J. 2007. Protein misassembly: macromolecular crowding and molecular chaperones. *Adv Exp Med Biol* 594:1–13.
- Fay, J.C., Wyckoff, G.J., and Wu, C.I. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415:1024–1026.
- Ferrer, P., Binz, H.K., Stumpp, M.T., and Pluckthun, A. 2004. Consensus design of repeat proteins. *ChemBiochem* 5:183–189.
- Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C., and Feldman, M.W. 2002. Evolutionary rate in the protein interaction network. *Science* 296:750–752.
- Gillespie, S.H. 2001. Antibiotic resistance in the absence of selective pressure. *Int J Antimicrob Agents* 17:171–176.
- Goldman, N., Thorne, J.L., and Jones, D.T. 1998. Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics* 149:445–458.
- Greenman, C., Stephens, P., Smith, R., and other authors 2007. Patterns of somatic mutation in human cancer genomes. *Nature* 446:153–158.
- Guo, H.H., Choe, J., and Loeb, L.A. 2004. Protein tolerance to random amino acid change. *Proc Natl Acad Sci U S A* 101:9205–9210.
- Hahn, M.W., Conant, G.C., and Wagner, A. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? *J Mol Evol* 58:203–211.
- Hall, B.G. 2004. Predicting the evolution of antibiotic resistance genes. *Nat Rev Microbiol* 2:430–435.
- Herring, C.D., Raghunathan, A., Honisch, C., et al. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* 38:1406–1412.
- Huang, W. and Palzkill, T. 1997. A natural polymorphism in beta-lactamase is a global suppressor. *Proc Natl Acad Sci USA* 94:8801–8806.
- Hurst, L.D. and Smith, N.G. 1999. Do essential genes evolve slowly? *Curr Biol* 9:747–750.
- James, L.C. and Tawfik, D.S. 2003. Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem Sci* 28:361–368.
- Jordan, I.K., Kondrashov, F.A., Adzhubei, I.A., Wolf, Y.I., Koonin, E.V., Kondrashov, A.S., and Sunyaev, S. 2005. A universal trend of amino acid gain and loss in protein evolution. *Nature* 433:633–638.
- Khersonsky, O., Roodveldt, C., and Tawfik, D.S. 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10:498–508.
- Kim, P.M., Lu, L.J., Xia, Y., and Gerstein, M.B. 2006. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314:1938–1941.
- Kondrashov, A.S., Sunyaev, S., and Kondrashov, F.A. 2002. Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci USA* 99:14878–14883.
- Kulathinal, R.J., Bettencourt, B.R., and Hartl, D. L. 2004. Compensated deleterious mutations in insect genomes. *Science* 306:1553–1554.
- Lenski, R.E., Barrick, J.E., and Ofria, C. 2006. Balancing robustness and evolvability. *PLoS Biol* 4:e428.
- Levin, B.R., Perrot, V., and Walker, N. 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* 154:985–997.
- Loeb, D.D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E., and Hutchison, C.A., 3rd 1989. Complete mutagenesis of the HIV-1 protease. *Nature* 340:397–400.
- Loh, E., Choe, J., and Loeb, L.A. 2007. Highly tolerated amino acid substitutions increase the fidelity of *E. coli* DNA polymerase I. *J Biol Chem* 282:12201–12209.
- Maisnier-Patin, S., Berg, O.G., Liljas, L., and Andersson, D.I. 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. *Mol Microbiol* 46:355–366.
- Mao, E.F., Lane, L., Lee, J., and Miller, J.H. 1997. Proliferation of mutators in *A* cell population. *J Bacteriol* 179:417–422.
- Markiewicz, P., Kleina, L.G., Cruz, C., Ehret, S., and Miller, J.H. 1994. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as “spacers” which do not require a specific sequence. *J Mol Biol* 240:421–433.
- Merlo, L.M., Pepper, J.W., Reid, B.J., and Maley, C.C. 2006. Cancer as an evolutionary and ecological process. *Nat Rev Cancer* 6:924–935.
- Nagaev, I., Bjorkman, J., Andersson, D.I., and Hughes, D. 2001. Biological cost and compensatory evolution in fusidic acid-resistant *Staphylococcus aureus*. *Mol Microbiol* 40:433–439.
- Neduva, V. and Russell, R.B. 2005. Linear motifs: evolutionary interaction switches. *FEBS Lett* 579:3342–3345.
- Ng, P.C. and Henikoff, S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 7:61–80.
- O’Loughlin, T.L., Patrick, W.M., and Matsumura, I. 2006. Natural history as a predictor of protein evolvability. *Protein Eng Des Sel* 19:439–442.
- Orencia, M.C., Yoon, J.S., Ness, J.E., Stemmer, W.P., and Stevens, R.C. 2001. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat Struct Biol* 8:238–242.
- Orr, H.A. 2005. The genetic theory of adaptation: a brief history. *Nat Rev Genet* 6:119–127.
- Pagel, M., Venditti, C., and Meade, A. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* 314:119–121.
- Pakula, A.A., Young, V.B., and Sauer, R.T. 1986. Bacteriophage lambda cro mutations: effects on activity and intracellular degradation. *Proc Natl Acad Sci USA* 83:8829–8833.
- Parisi, G. and Echave, J. 2005. Generality of the structurally constrained protein evolution model: assessment on representatives of the four main fold classes. *Gene* 345:45–53.
- Poon, A. and Chao, L. 2005. The rate of compensatory mutation in the DNA bacteriophage phiX174. *Genetics* 170:989–999.
- Poon, A., Davis, B.H., and Chao, L. 2005. The coupon collector and the suppressor mutation: estimating the number of compensatory mutations by maximum likelihood. *Genetics* 170:1323–1332.
- Queitsch, C., Sangster, T.A., and Lindquist, S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* 417:618–624.
- Rennell, D., Bouvier, S.E., Hardy, L.W., and Poteete, A.R. 1991. Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 222:67–88.
- Rutherford, S.L. and Lindquist, S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–342.
- Salathe, M., Ackermann, M., and Bonhoeffer, S. 2006. The effect of multifunctionality on the rate of evolution in yeast. *Mol Biol Evol* 23:721–722.

- Schrag, S.J., Perrot, V., and Levin, B.R. 1997. Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. *Proc Biol Sci* 264:1287–1291.
- Sjoblom, T., Jones, S., Wood, L.D., and other authors 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* 314:268–274.
- Smith, N.G. and Eyre-Walker, A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.
- Sniegowski, P.D. and Murphy, H.A. 2006. Evolvability. *Curr Biol* 16:R831–834.
- Stone, E.A. and Sidow, A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res* 15:978–986.
- Suckow, J., Markiewicz, P., Kleina, L.G., Miller, J., Kisters-Woike, B., and Muller-Hill, B. 1996. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J Mol Biol* 261:509–523.
- Thomas, R.K., Baker, A.C., Debiasi, R.M., and other authors 2007. High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 39:347–351.
- Wall, D.P., Hirsh, A.E., Fraser, H.B., Kumm, J., Giaever, G., Eisen, M.B., and Feldman, M.W. 2005. Functional genomic analysis of the rates of protein evolution. *Proc Natl Acad Sci U S A* 102:5483–5488.
- Wang, T.W., Zhu, H., Ma, X.Y., Zhang, T., Ma, Y.S., and Wei, D.Z. 2006. Mutant library construction in directed molecular evolution: casting a wider net. *Mol Biotechnol* 34:55–68.
- Wang, X., Minasov, G., and Shoichet, B.K. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320:85–95.
- Watanabe, K., Ohkuri, T., Yokobori, S., and Yamagishi, A. 2006. Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J Mol Biol* 355:664–674.
- Weinreich, D.M. and Chao, L. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution Int J Org Evolution* 59:1175–1182.
- Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312:111–114.
- Wichman, H.A., Badgett, M.R., Scott, L.A., Boulianne, C.M., and Bull, J.J. 1999. Different trajectories of parallel evolution during viral adaptation. *Science* 285:422–424.
- Wichman, H.A., Scott, L.A., Yarber, C.D., and Bull, J.J. 2000. Experimental evolution recapitulates natural evolution. *Philos Trans R Soc Lond B Biol Sci* 355:1677–1684.
- Wichman, H.A., Millstein, J., and Bull, J.J. 2005. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170:19–31.
- Wintrobe, P.L. and Arnold, F.H. 2000. Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv Protein Chem* 55:161–225.
- Woods, R., Schneider, D., Winkworth, C.L., Riley, M.A., and Lenski, R.E. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc Natl Acad Sci U S A* 103:9107–9112.
- Yue, P. and Moul, J. 2006. Identification and analysis of deleterious human SNPs. *J Mol Biol* 356:1263–1274.

Editor: Susan M. Rosenberg