

Mutational Reversions During Adaptive Protein Evolution

Mark A. DePristo,* Daniel L. Hartl,* and Daniel M. Weinreich*¹

*Department of Organismic and Evolutionary Biology, Harvard University

Adaptation is often regarded as the sequential fixation of individually, intrinsically beneficial mutations. Contrary to this expectation, we find a surprisingly large number of evolutionary trajectories on which natural selection first favors a mutation, then favors its removal, and later still favors its ultimate restoration during the course of antibiotic resistance evolution. The existence of reversion trajectories implies that natural selection may not follow the most parsimonious path separating two alleles, even during adaptation. Altogether, this discovery highlights the unusual and potentially circuitous routes natural selection can follow during adaptation.

Adaptation is often regarded as the sequential fixation by natural selection of individually, intrinsically beneficial mutations (Hartl and Clark 1997). This simple picture is complicated by the recent appreciation that mutations often exhibit sign epistasis (Weinreich et al. 2005), so that the effect of a mutation is beneficial on some genetic backgrounds but neutral or even deleterious on others. Consequently, the fixation of one mutation may alter both the number and even the identity of subsequent beneficial mutations (Weinreich et al. 2005).

In order to study adaptation in more detail, we recently determined the resistance of all combinations of 5 clinically relevant mutations in the TEM-1 allele of the β -lactamase gene (Weinreich et al. 2006) that jointly increase resistance of *Escherichia coli* to the antibiotic cefotaxime by $\sim 100,000$ -fold (Stemmer 1994). The fitness landscape (Wright 1932) corresponding to the resistance of these $2^5 = 32$ β -lactamase alleles displayed a striking degree of sign epistasis (Weinreich et al. 2005): in spite of their jointly beneficial effect, 4 of the 5 mutations reduced resistance in some combinations. Additionally, from this resistance data one can calculate the likelihood of evolution following a particular mutational trajectory—an ordering of the 5 mutations—from the initial TEM-1 allele to the high-resistance TEM* allele (see SOM for mathematical details). Disregarding mutational reversions there are $5! = 120$ possible mutational trajectories from TEM-1 to TEM*, but as a consequence of extensive sign epistasis (Weinreich et al. 2005), only 18 of these are selectively accessible (Weinreich et al. 2006).

We recently extended the initial trajectory analysis (Weinreich et al. 2006) to include not only forward mutational trajectories but also reversion trajectories (see Methods). Such reversion trajectories contain at least one site where a mutation is first fixed, then later reverted back to its initial state, though ultimately mutated again to arrive at the final TEM* allele. The class of acyclic reversion trajectories is immense: exhaustive computational enumeration identified 18,651,552,840 possible trajectories between 2 alleles separated by 5 mutations, ranging in length from 5 to 31 mutations (see Methods).

Applying this generalized analysis to the experimental TEM resistance data revealed an additional 9 selectively accessible trajectories involving mutational reversions: 7 with 1 reversion and 2 with 2 reversions (table 1). Although these 9 reversion trajectories (figure 1) are a small fraction of the ~ 18 billion possible reversion trajectories, they increase by 50% the number of accessible trajectories (table 1). Under a truncated selection scheme—the equal probability of fixation model—in which an allele either provides complete resistance or none at all, a ‘base’ assumption for bactericidal antibiotics (Pankey and Sabath 2004), then these reversions trajectories account for 14% of the probability density of realization by natural selection (see Methods). If instead fitness correlates with resistance levels—which may be true for even bactericidal antibiotics (Negri et al. 2000)—the likelihood of these trajectories falls to only 1% of the probability density (see Methods). Though the likelihood of these reversion trajectories may be modest in our dataset and clearly vary with fixation probability models, they nevertheless constitute a clearly biologically possible and intriguing phenomenon.

Two of the 5 mutations are involved in reversion trajectories. One is the non-coding mutation, g4205a, which increases gene expression by around 2-fold (Stemmer 1994) but reduces resistance when combined with alleles that likely result in thermodynamically unstable proteins. Second is M182T, a thermodynamically stabilizing mutation that is advantageous when combined with unstable alleles but can slightly reduce enzymatic performance (Wang et al. 2002). As is clear in figure 1, reversions of these 2 mutations occur near the end of the adaptive walk from TEM-1 to TEM*. This bias is qualitatively consistent with that fact that the number of potential reversions (and trajectories) increases rapidly with mutational distance from TEM-1 (see SOM). This behavior appears to be a general feature of reversion trajectories, as the opportunity for reversions grows striking quickly with the number of mutations that separate the initial and final alleles and the distance from the initial allele (see SOM).

Insofar as β -lactamase adaptation is representative of protein adaptation in general, the possibility of reversion trajectories has important implications for our view of molecular evolution. While potentially unlikely, the small but real possibility of reversion trajectories means that natural selection may not follow the most parsimonious path separating 2 alleles. Such reversion trajectories further indicate that the number of fixation events separating orthologous genes may exceed the number of sites at which they differ, an issue that also arises with the neutral multiple hits problem (Jukes and Cantor 1969). Moreover, selectively accessible reversion

¹ Present address: Department of Ecology and Evolutionary Biology and Center for Computational Molecular Biology, Brown University, Providence, RI 02912.

Key words: adaptation, reversion, evolution, protein, gene, mutation.

Email: mark@depristo.com.

Mol. Biol. Evol. 24(8):1608–1610. 2007

doi:10.1093/molbev/msm118

Advance Access publication June 7, 2007

Table 1
Reversion Trajectories and Probabilities

Length ^a	EVT ^b %	Equal ^c %	Trajectory ^d
7	0.04	3.13	E104K, <u>g4205a</u> , A42G, G238S, <u>a4205g</u> , M182T, g4205a
7	0.02	1.56	E104K, g4205a, G238S, A42G, <u>a4205g</u> , M182T, g4205a
7	0.05	3.13	E104K, A42G, <u>g4205a</u> , G238S, <u>a4205g</u> , M182T, g4205a
7	5.46 ⁰⁴	0.52	E104K, M182T, A42G, <u>T182M</u> , <u>g4205a</u> , G238S, M182T
7	0.01	1.04	E104K, M182T, A42G, <u>T182M</u> , G238S, M182T, g4205a
7	2.58 ⁰⁴	0.52	E104K, M182T, A42G, <u>g4205a</u> , <u>T182M</u> , G238S, M182T
7	0.58	2.78	G238S, A42G, <u>g4205a</u> , E104K, <u>a4205g</u> , M182T, g4205a
9	9.93 ⁰⁵	0.52	E104K, M182T, A42G, <u>T182M</u> , <u>g4205a</u> , G238S, <u>a4205g</u> , M182T, g4205a
9	4.69 ⁰⁵	0.52	E104K, M182T, A42G, <u>g4205a</u> , <u>T182M</u> , G238S, <u>a4205g</u> , M182T, g4205a
Sum	0.69	13.72	

^a Number of mutations along this path from the initial TEM-1 allele to the final TEM* allele.
^b Probability (out of 100%) of each reversion trajectory according to the correlated fixation model.
^c Probability (out of 100%) of each reversion trajectory according to the equal fixation probability model.
^d Mutations along the path, with reversions underlined.

trajectories hint that adaptive trajectories may even contain mutational detours (Horn et al. 1994; Poelwijk et al. 2006), the pathological situation in which some mutations are fixed only transiently along the trajectory. These effects may com-

plicate the comparison of experimentally-observed adaptive walks with theoretical expectations (Orr 2002, 2003). Altogether, the existence of reversion trajectories during antibiotic resistance evolution highlights the unusual paths natural

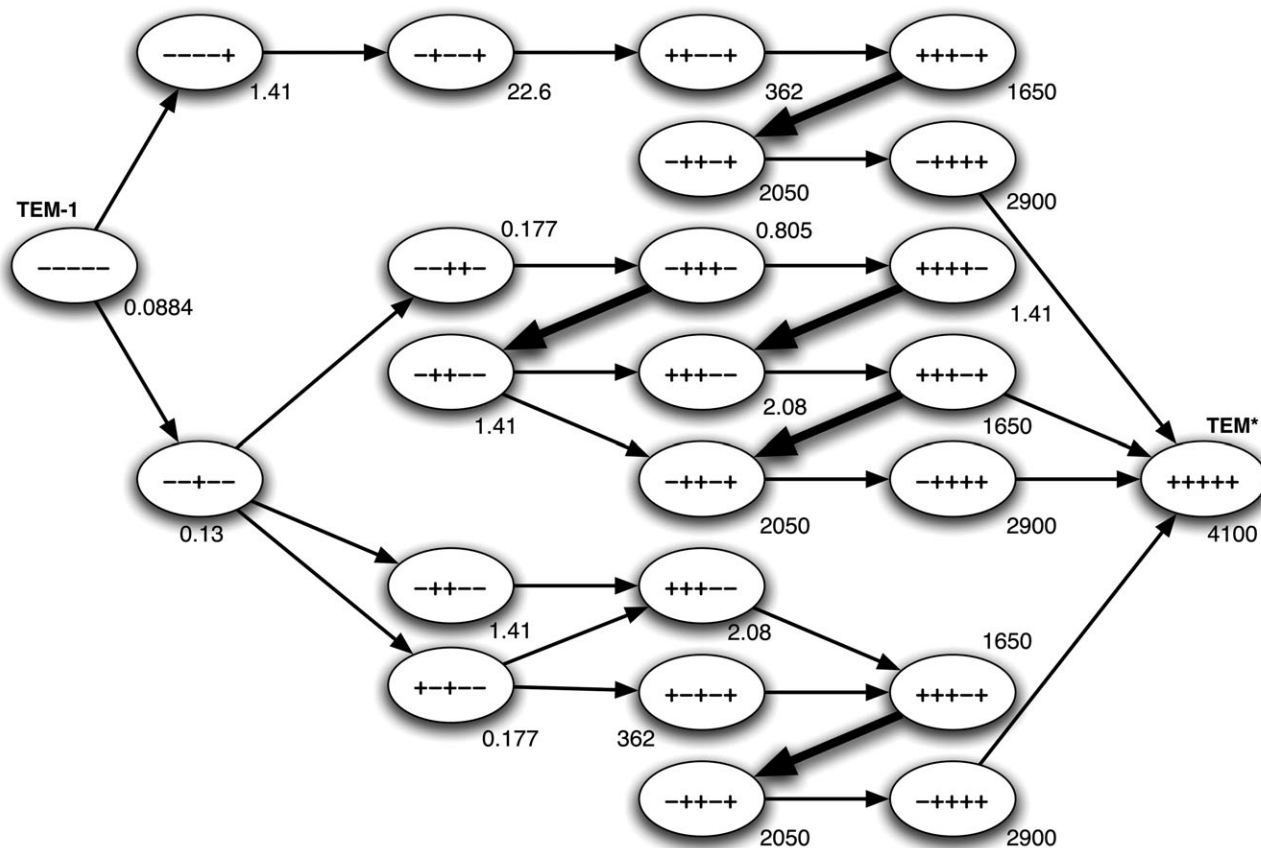


FIG. 1.—Mutational trajectories from TEM-1 allele (-----) to the highest-resistance quintuple mutant TEM* (+++++) that involve at least one reversion. Nodes represent alleles whose identities are given by a string of five plus or minus symbols corresponding (left to right) to the presence or absence of mutations g4205a, A42G, E104K, M182T, and G238S, respectively. Small arrows indicate a forward mutation, large backward arrows indicate reversions. Alleles in the same column are mutationally equidistant from the quintuple mutant allele. Adjacent to each allele is its minimum inhibitory concentration (in µg/ml), the lowest concentration of cefotaxime required to inhibit growth of *E. coli* cells carrying that β-lactamase allele. Note that some alleles have been duplicated for clarity of presentation.

selection can follow during adaptation, a potentially common but unappreciated feature of molecular evolution.

Methods

Mutational Trajectory Calculation

A full treatment of the calculation of the probability of a trajectory of mutations from TEM-1 to TEM* can be found in the Supplementary Materials of Weinreich et al (Weinreich et al. 2006). Briefly, assuming the strong-selection, weak-mutation (SSWM) model (Gillespie 1991), the probability of following a trajectory of alleles:

$$P(TEM^1 \rightarrow A \rightarrow B \rightarrow C \rightarrow D \rightarrow TEM^*) \quad (1)$$

is the product of the individual transition probabilities:

$$P(TEM^1 \rightarrow A) \cdot P(A \rightarrow B) \cdot P(B \rightarrow C) \\ \cdot P(C \rightarrow D) \cdot P(D \rightarrow TEM^*) \quad (2)$$

We employ 2 different models to calculate probability $P(i \rightarrow j)$ of fixing a single mutant neighbor j of allele i . The first is the equal fixation model, in which the probability of fixing a beneficial mutation is independent of the magnitude of its resistance improvement. Such truncated selection is often observed for bactericidal antibiotic, as bacteria either live in a particular antibiotic concentration or die out completely (Pankey and Sabath 2004). Mathematically, we have that:

$$P(i \rightarrow j) = 1 / |M_i^+| \quad (3)$$

where M_i^+ is the set of beneficial single mutant neighbors of allele i . A single mutant neighbor j is beneficial with respect to allele i if and only if the minimum inhibitory concentration (MIC) of j is greater than i .

The second, the correlated fixation model, uses extreme value theory (EVT) to introduce a correlation between the magnitude of a mutation's phenotypic effect and its probability of fixation. Under this model, $P(i \rightarrow j)$ is:

$$P(i \rightarrow j) = \frac{\sum_{x=r_j}^{r_i-1} \frac{1}{x}}{\sum_{k \in M_i^+} \sum_{x=r_k}^{r_i-1} \frac{1}{x}} \quad (4)$$

where r_i is the fitness rank of allele i among all alleles regardless of mutational adjacency.

The mutational trajectories discussed in the main text were calculated using the MIC data presented in Weinreich et al. (2006), with a corrected estimate of 0.5 $\mu\text{g/ml}$ for allele A42G/M182T, determined in triplicate.

Acknowledgments

D.M.W. was supported by the NSF under Award DEB-0343598. M.A.D. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-1861-05).

Literature Cited

- Gillespie JH. 1991. The causes of molecular evolution. Oxford, UK: Oxford University Press.
- Hartl DL, Clark AG. 1997. Principles of population genetics. Sunderland, MA: Sinauer Associates, Inc.
- Horn J, Goldberg DE, Deb K. 1994. Long path problems. In: Davidor Y, Schwefel H-P, Manner R, editors. Parallel problem solving from nature. Berlin, Germany: Springer-Verlag.
- Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York, NY: Academic Press. p. 21-132.
- Negri MC, Lipsitch M, Blazquez J, Levin BR, Baquero F. 2000. Concentration-dependent selection of small phenotypic differences in TEM beta-lactamase-mediated antibiotic resistance. Antimicro Agents Chemo. 44:2485-2491.
- Orr HA. 2002. The population genetics of adaptation: the adaptation of DNA sequences. Evolution. 56:1317-1330.
- Orr HA. 2003. A minimum on the mean number of steps taken in adaptive walks. J Theo Biol. 220:241-247.
- Pankey GA, Sabath LD. 2004. Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of Gram-positive bacterial infections. Clin Infect Dis. 38: 864-870.
- Poelwijk FJ, Kiviet DJ, Tans SJ. 2006. Evolutionary potential of a duplicated repressor-operator pair: simulating pathways using mutation data. PLOS Compu Biol. 2:58.
- Stemmer WP. 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature. 370:389-391.
- Wang X, Minasov G, Shoichet BK. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol. 320:85-95.
- Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 312:111-114.
- Weinreich DM, Watson RA, Chao L. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 59:1165-1174.
- Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Jones DF, editor. Proceedings of the Sixth International Congress of Genetics. Menasha, WI: Brooklyn Botanic Garden. p. 356-366.

Marcy Uyenyama, Associate Editor

Accepted June 04, 2007