The role of robustness in phenotypic adaptation and innovation

Wagner, Andreas

Abstract: Phenotypes that vary in response to DNA mutations are essential for evolutionary adaptation and innovation. Therefore, it seems that robustness, a lack of phenotypic variability, must hinder adaptation. The main purpose of this review is to show why this is not necessarily correct. There are two reasons. The first is that robustness causes the existence of genotype networks–large connected sets of genotypes with the same phenotype. I discuss why genotype networks facilitate phenotypic variability. The second reason emerges from the evolutionary dynamics of evolving populations on genotype networks. I discuss how these dynamics can render highly robust phenotypes more variable, using examples from protein and RNA macromolecules. In addition, robustness can help avoid an important evolutionary conflict between the interests of individuals and populations-a conflict that can impede evolutionary adaptation.

DOI: https://doi.org/10.1098/rspb.2011.2293

The role of robustness in phenotypic adaptation and innovation

Invited review for the Proceedings of the Royal Society

Andreas Wagner

University of Zurich
Institute of Evolutionary Biology
and Environmental Studies, Y27-J-54
Winterthurerstrasse 190
CH-8057 Zurich Switzerland

Office: +41-44-635-6141
FAX: +41-44-635-6144
Email: andreas.wagner@ieu.uzh.ch

Abstract. Phenotypes that vary in response to DNA mutations are essential for evolutionary adaptation and innovation. Therefore it seems that robustness, a lack of phenotypic variability, must hinder adaptation. The main purpose of this review is to show why this is not necessarily correct. There are two reasons. The first is that robustness causes the existence of genotype networks, large connected sets of genotypes with the same phenotype. I discuss why genotype networks facilitate phenotypic variability. The second reason emerges from the evolutionary dynamics of evolving populations on genotype networks. I discuss how these dynamics can render highly robust phenotypes more variable, using examples from protein and RNA macromolecules. What is more, robustness can help avoid an important evolutionary conflict between the interests of individuals and populations, a conflict that can impede evolutionary adaptation.

Introduction. A feature or *phenotype* of an organism is robust if it persists when perturbed. Phenotypes encompass a broad range of traits, from macroscopic, visible traits, to molecular traits, such as the expression level of a gene, or the three-dimensional conformation of a protein.

The perturbations that can affect a phenotype fall into two broad categories. The first comprises environmental perturbations. These include changes in an organism's exterior environment, such as changes in temperature, in available nutrients, or in the abundance of other organisms, such as potential prey. They also include changes in an organism's internal environment, such as temporal fluctuations in gene expression levels, which are caused by ubiquitous intracellular noise. The second category of changes are mutations, changes in an organism's DNA, its *genotype*. Mutations affect an organism more permanently than environmental change, because the changes they cause are readily inherited from generation to generation. For this reason, they are especially important study objects for students of evolution. I will here focus on robustness to genetic mutations – mutational robustness. A huge body of literature shows that living systems on all levels of organization – from macromolecules to whole organisms – are to some extent robust to mutations (Kitano 2004; Stelling et al. 2004; Wagner 2005a; Wagner 2005b). I note that mutationally robust systems are often also robust to environmental change (Ancel & Fontana 2000; Lehner 2010; Masel & Siegal 2009; Meiklejohn & Hartl 2002; Szollosi & Derenyi 2009; Wagner 2005b), even though exceptions may exist (Cooper et al. 2006; Milton et al. 2003). Thus, most observations I make here about mutational robustness apply to environmentally robust systems as well.

During a population's evolutionary adaptation to a new environment, which takes place over multiple generations, its members undergo mutations. Most of these mutations are detrimental, and only few change an existing phenotype into a new, better-adapted phenotype (Eyre-Walker & Keightley 2007; Sawyer et al. 2007; Soskine & Tawfik 2010; Wloch et al. 2001). Very occasionally, one or more mutations may also bring forth evolutionary innovations – new phenotypes that are qualitatively different and superior to existing phenotypes. The ability of mutations to bring forth new phenotypes is important to Darwinian evolution. I will refer to it as *phenotypic variability*.

Phenotypic variability and robustness may seem opposite properties, because in a robust system, mutations do not easily change a phenotype. The main purpose of this review is to show why this view is not necessarily correct, and why robustness can instead be a prerequisite for phenotypic variability.

To make this point beyond individual case studies – anecdotes of natural history – one needs to study the relationship between genotypic change and the resulting phenotypic change *systematically*, either experimentally, through comparative data, or through computational modeling. I will focus here on protein and RNA macromolecules, for which this has become possible in recent years, but similar principles also hold for very different classes of systems (ESM and (Wagner 2011c)).

Other recent reviews have focused on phenotypic variability and its relationship to recombination (Masel & Trotter 2010), enzyme promiscuity (Khersonsky & Tawfik 2010), commonalities among different system classes (Wagner 2011b), and phenotypic constraints (Wagner 2011a). In contrast, this review's focus is the role of robustness in phenotypic variability. In part one I introduce some concepts and discuss the most difficult problem organisms face in evolutionary adaptation and innovation. Parts two and three focus on the two mean respects in which robustness affects phenotypic variability. Specifically, part two shows that robustness influences how genotypes with the same phenotype are organized in a vast space of genotypes. This *static* or *structural* influence of robustness contrasts with its influence on the evolutionary *dynamics* of populations that is the focus of part three. A brief part four discusses the evolutionary conflict between the interests of an individual and that of a population in producing phenotypic variation. It points out that robustness as a variability mechanism can avoid this conflict. I will not discuss the important but controversial claim that mechanisms endowing living systems with robustness originated in evolution *because* they favor variability, for lack of sufficient evidence (Pigliucci 2008).

Exploring the new while conserving the old. Proteins and RNA catalyze all chemical reactions, provide structural support, help cells and organisms move, guide cell communication, and carry out many other functions. The *genotype* of each such molecule, is a sequence of amino acids or RNA nucleotides. *Genotype space* comprises all sequences of a given length L. It is astronomically large, and comprises 20^L protein

genotypes and 4^L RNA genotypes. A molecule's *phenotype* refers to the fold or conformation that it forms in space, and to the biochemical function that this fold makes possible. New phenotypes in molecules arise through genotypic changes that cause phenotypic change. Macromolecules are well-studied, with many known adaptations and innovation (Cheng 2006; Khersonsky & Tawfik 2010; Li 1997; Wagner 2011c).

Evolutionary change takes place in populations of organisms. Each member of a population has some genotype. For the purpose of this review, it is useful to think of a population as a collection of genotypes in genotype space. The members of this population change their location in this space through mutations. An especially important class of mutations are point mutations, which transform a genotype into one of its neighbors – a genotype that differs in one amino acid or nucleotide. The individuals in an evolving population also have some phenotype. Natural selection preserves individuals with well-adapted phenotypes, and eliminates mutants with poorly adapted phenotypes.

Somewhere in genotype space a superior genotype may exist whose phenotype is better adapted to the current environment. The central problem in evolutionary adaptation is how a population can find such a superior genotype. The problem is difficult, because this genotype may exist far away from the population's current genotype, and because the vast majority of mutant phenotypes a population explores are inferior, not superior to existing phenotypes (Eyre-Walker & Keightley 2007; Sawyer et al. 2007; Soskine & Tawfik 2010). What is more, during a population's "search" for this genotype, an existing well-adapted phenotype must be preserved. Perhaps the most compact way to express this problem is with an analogy from politics: evolving populations need to be both "conservative" and "progressive" at the same time. A tall order indeed.

Two generic features of genotype space make it possible to reconcile these conflicting demands. The first is the existence of connected genotype networks, vast sets of genotypes whose members all have the same phenotype (Figure 1). These sets extend far through genotype space, and can be traversed in many small steps of individual mutations with little or no phenotypic change. Their existence has first been suggested by computational models of phenotype formation (Lipman & Wilbur 1991; Schuster et al. 1994), but they also exist in real macromolecules. A paradigmatic example is the family of oxygen-binding globins. They comprise hundreds of known members that are

connected through single amino acid changes to a common ancestor. They share a common fold and biochemical function – oxygen binding – but have diverged in more than 95 percent of their amino acid residues (Goodman et al. 1988; Hardison 1996). Macromolecules like this are the rule rather than the exception (Bastolla et al. 2003; Rost 2002; Thornton et al. 1999; Todd et al. 1999).

Genotype networks are sometimes called neutral networks (Schuster et al. 1994). However, evolutionary change on such networks is usually all but neutral. That is, such change may affect fitness. For example, weakly deleterious mutations are more abundant than neutral mutations in most macromolecules, but they are often followed by compensatory changes that allow a preservation of phenotype. Similarly, in large populations the simultaneous occurrence of multiple mutations can help a population "tunnel" through a region of low fitness in genotype space, and thus help preserve a phenotype (Eyre-Walker et al. 2002; Hartl & Clark 2007; Kern & Kondrashov 2004; Kulathinal et al. 2004; Sawyer et al. 2007; Weinreich & Chao 2005). Because phenotype preservation does not require neutrality of individual mutations, I refrain from using the word neutrality in this context.

The second central feature of genotype space regards the collection of those genotypes that can be reached from any one genotype through one or few mutations. This collection is also called a genotype's *neighborhood*. Neighborhoods are important, because the set of different phenotypes in a neighborhood are easily accessible by mutation. The size of this set is thus a simple measure of how phenotypically variable a genotype is in response to mutations (Wagner 2008). The second feature of genotype space is that neighborhoods of different genotypes typically contain different novel phenotypes (Figure 1, see also the ESM).

The first feature, genotype networks, allows individuals in a population to preserve their phenotype, while changing their genotype in many small steps that, cumulatively, can add up to substantial divergence. Because of the second feature, the different genotypes on a genotype network can explore different phenotypes, precisely because their neighborhoods contain different novel phenotypes.

Thus far, I implicitly assumed that one genotype has one phenotype, a simplification that helps illustrate key concepts in simple terms. However, it is important

to be aware that many molecules can form multiple folds and exert multiple functions (O'Brien & Herschlag 1999; Tokuriki & Tawfik 2009b). Such multifunctionality can play an important role in the origin of novel phenotypes (Khersonsky & Tawfik 2010), and can further enhance the variability caused by genotype networks (Wagner 2011c, Ch. 13).

In bringing forth genotype networks, robustness facilitates phenotypic variability. To appreciate the role of robustness in phenotypic variability, it is useful to define robustness in the genotype space framework. Specifically, a genotype is to some extent robust to mutations if it has some neighbors with the same phenotype *P* as itself. One can show that robustness thus defined is both necessary and sufficient for the existence of genotype networks (see ESM).

Now compare, as a thought experiment, two kinds of genotypes. The first is a minimally robust genotype, that is, a genotype that has no neighbors with phenotype *P*. Figure 2a shows such a hypothetical, minimally robust genotype *G* (black circle) with eight neighbors (dashed black lines). The second genotype is a genotype with some robustness, as exemplified by the left-most hypothetical genotype *G* (black circle) in Figure 2b. Half of the eight neighbors of this genotype have the same phenotype *P* as itself (solid lines), whereas the other half (dashed dark blue lines) have new phenotypes (not shown in the figure), all of which might be different from each other.

Which of these two genotypes is phenotypically more variable, in the sense that it can access a greater number of novel phenotypes through mutation? The answer is genotype *G* in Figure 2a. Because it is minimally robust, all of its eight neighbors have a phenotype different from *P*. In contrast, the left-most genotype in Figure 2b is more robust but less phenotypically variable, because only some of its neighbors have a phenotype different from *P*. This is the core argument why robustness reduces phenotypic variability, cast in the abstract but precise language of genotype space.

Figure 2b also illustrates why this argument is flawed. The robust left-most genotype *G* has neighbors with the same phenotype *P* as itself, one of which, *G'*, is shown as the middle circle in Figure 2b. This neighbor *G'* itself has a neighborhood, which contains five genotypes with new phenotypes (dashed medium blue lines) that may not occur in the neighborhood of *G* itself. *G'* has three further neighbors (solid lines) that have the same phenotype *P*, one of which is shown as *G''*. The neighborhood of *G''*

contains four genotypes with novel phenotypes (dashed light blue lines). In comparison with the non-robust genotype, from which up to 8 different novel phenotypes are accessible, the robust genotype can – merely through the neighbors shown in the figure – access up to (4+5+4=13) new phenotypes, more than if it was not robust. This argument takes only into account the neighborhoods of *G'* and *G''*, not the neighborhoods of several other neighbors of *G* with the same phenotype *P*. Thus, the actual number of different accessible phenotypes may be even higher for the robust genotype.

Three concrete examples show how much higher. These examples are based on three different natural RNA molecules, a guide RNA, a ribozyme, and a telomerase (Figure 2c, same color coding as Figures 2a and 2b). The phenotype in question is RNA secondary structure, a planar fold that occurs when an RNA molecule folds onto itself through internal base pairing. Secondary structure is essential for the function of many RNA molecules, and thus in itself a phenotype worthy of study. It can be predicted computationally using known biophysical RNA folding principles (Hofacker et al. 1994).

Each of the three panels of Figure 2c shows, in a black bar, the maximally possible number of novel secondary structure phenotypes accessible to an RNA genotype if its phenotype were minimally robust (as in Figure 2a). This number equals the total number of neighbors of an RNA molecule, which equals three times the molecule's total length *L* in nucleotides, because every one of the molecule's nucleotides can mutate into three other nucleotides. For example, for the guide RNA with length *L*=40, these would be 3×40=120 neighbors and novel phenotypes.

The dark blue bars indicate the *actual* number of accessible new phenotypes in the neighborhood of a genotype. This number was obtained by computing the minimum free energy secondary structure phenotype of each neighbor of a genotype with an RNA folding algorithm (Hofacker et al. 1994). For the guide RNA, this number is 40, many fewer than the maximally 120 new phenotypes without robustness, thus confirming the principle illustrated Figures 2a and 2b.

The medium and light blue bars indicate the total number of different phenotypes that are accessible up to two and three mutations away from the starting genotype. This number – obtained again by computing the structures for all genotypes in these neighborhoods – is much larger than the 120 maximally attainable phenotypes in the

absence of robustness. Specifically, for the guide RNA discussed here 746 (medium blue bars) and 1174 (light blue bars) distinct new phenotypes become accessible two and three mutations away. Thus, robustness allows access to many novel RNA phenotypes.

I note that the genotypes considered in this computation (up to two mutations away from *G* and with the same phenotype *P*) are a tiny fraction of the genotypes that form a typical genotype network. For example, a simple calculation shows that there are only 7.2×10^3 total genotypes that differ from the guide RNA with L=40 in no more than two mutations. However, the size of this guide RNA's genotype network – which can be computed – equals approximately 9.1×10^{17} ($\pm 3.3\times10^{16}$) genotypes, and is thus more than 14 orders of magnitude larger (Jörg et al. 2008). (Astronomically large genotype networks are typical for natural RNA molecules.) It is currently not feasible to compute the number of distinct phenotypes near a genotype network this large, but this number would surely also be astronomical.

Taken together, these observations mean that the mere existence of robustness makes a dramatic difference in phenotypic variability. The difference is that between the few novel phenotypes accessible in the immediate neighborhood of a non-robust genotype, and the extremely large number of new phenotype accessible from the neighbors of a genotype network. By bringing genotype networks into existence, robustness makes vastly more new phenotypes accessible.

I will next discuss two lines of experimental evidence that indicate how important this principle is for the discovery of new molecular phenotypes.

The first experiment revolves around one natural and one synthetic ribozyme (Schultes & Bartel 2000). The natural ribozyme which catalyzes its own cleavage is encoded by the human hepatitis delta virus. The synthetic RNA is the so-called class III self-ligating ribozyme, which joins an oligonucleotide substrate to its own 5' end. The two ribozymes are unrelated in sequence and fold. (Schultes & Bartel 2000). Schultes and Bartel (Schultes & Bartel 2000) were able to design a mutational path through RNA genotype space that starts from either one of the ribozymes and leaves its native activity largely intact, until it reaches a hybrid ribozyme that is more than 40 mutational steps from each starting point. This hybrid can act both as a self-cleaving ribozyme and as a ligase, albeit with lower catalytic activity than the starting enzymes. By constructing a

hybrid ribozyme and constructing a path through sequence space back to its ancestors, this work makes two key points. First, many consecutive changes in a genotype are possible that do not affect an RNA's (catalytic) phenotype. Without robustness, this would not be the case. Second, these changes can be very important intermediate steps to create a new catalytic function. Similar principles have been suggested for other ribozymes (Beckert et al. 2008; Huang & Szostak 2003).

This experiment demonstrated the role of robustness in the origin of new phenotypes using an engineered path through genotype space. Biological evolution does not use such pre-meditated paths, but random changes in evolving populations. The next experiment shows that robustness is also highly relevant in such populations (Hayden et al. 2011). The experiment revolves around the concept of cryptic variation. Cryptic variation is genotypic variation in a population that is not visible on the level of phenotype (Gibson & Reed 2008). An example is variation in genotypes on the same genotype network. Cryptic variation cannot exist without mutational robustness. The experiment asks whether cryptic variation can help a population find a new and superior genotype during an evolutionary search.

The study system was again an RNA ribozyme, the so-called *Azoarcus* ribozyme (Tanner & Cech 1996), which is a naturally occurring ribozyme that can ligate a short RNA fragment to its own end. One can subject populations of ribozymes like this to repeated cycles of mutagenesis, and to selection to maintain or to modify this catalytic activity (Beaudry & Joyce 1992).

The experiment in question compared two different kinds of populations, one that consisted mostly of identical or similar genotypes, all of them copies of a single ribozyme sequence, and another that consisted of many diverse genotypes (Figure 3). Ribozymes in the two kinds of populations had similar catalytic activities on a specific RNA substrate, such that the average activities of the populations were indistinguishable. In other words, the first kind of populations contained little or no cryptic variation, whereas the second kind contained lots of it. The experiment then changed the chemical environments in which these populations existed. That is, it exposed both kinds of populations to a new, chemically modified RNA substrate, on which the starting ribozyme has low catalytic

activity. Both kinds of populations then experienced repeated rounds of mutagenesis and selection that favored high activity on the new substrate (Hayden et al. 2011).

Populations with much cryptic variation adapted up to six times faster to the new chemical environment (Figure 3b) (Hayden et al. 2011). They did so through genetic changes that improved the ribozyme's catalytic activity on the new substrate. DNA sequencing of thousands of genotypes from evolving populations subsequently showed why: There exists a superior genotype, and populations with much cryptic variation discover this genotype faster, because they are genotypically more diverse, and contain genotypes that are already close in genotype space to the superior genotype. In sum, this experiment shows that cryptic variation – a consequence of robustness – can accelerate evolutionary adaptation to a new chemical environment. Similar phenomena are likely to exist in proteins (Amitai et al. 2007; Bloom et al. 2007), even though we do not yet have proof that cryptic variation can accelerate the rate of adaptation for them.

Robustness can facilitate phenotypic variability by affecting evolutionary dynamics on large genotype networks. Thus far, I argued that robustness can facilitate variability through its *static, structural* role in organizing genotypes with the same phenotype into genotype networks. I will next turn to a second role of robustness, which builds on the first role: Robustness can increase variability through its influence on the evolutionary *dynamics* of populations on genotype networks. To this end, I will first introduce some further terminology from the genotype space framework. After that, I will explain how robustness can affect the evolutionary dynamics of populations, and then discuss a mix of pertinent experimental, computational, and comparative data.

Some phenotypes have very large associated genotype networks and are formed by many different genotypes. Others have much smaller genotype networks and are formed by fewer genotypes (Ciliberti et al. 2007a; Ciliberti et al. 2007b; Jörg et al. 2008; Li et al. 1996; Samal et al. 2010; Schuster et al. 1994). This difference in genotype network size is accompanied by a difference in the average robustness of genotypes encoding these phenotypes. Specifically, the larger a phenotype *P*'s genotype network is, the greater is also the average fraction of each genotype's neighbors with this phenotype *P*. In other words, genotypes on a large genotype network are more robust to mutations than genotypes on a small genotype network (Reidys et al. 1997; Wagner 2008). This

observation allows one to extend the definition of robustness I used thus far – the number of a *genotype*'s neighbors with the same phenotype. Specifically, one can define the robustness of a *phenotype* as the average robustness of all genotypes encoding it. Phenotypes with large genotype networks are more robust.

Now consider a population of initially identical genotypes with the same phenotype *P*. Subject the population to repeated cycles ("generations") of mutations and selection that confines the population to the genotype network of the phenotype *P*. After a given number of generations, examine the neighborhood of each individual in the population, and enumerate the number of *different* or *unique* phenotypes that occur in these neighborhoods. That is, if the same phenotype is formed by two or more genotypes in these neighborhoods, count it only once. This number is a measure of the phenotypic variability of an entire population, not just of a single individual. It encompasses all phenotypes that a population can access through a single nucleotide change in some individual.

To understand how phenotypic variability is affected by phenotypic robustness, it is necessary to examine how populations evolve on genotype networks that vary in size. Recent work on populations of evolving RNA molecules has done that for computationally predicted secondary structure phenotypes (Wagner 2008). It found that populations whose phenotypes have greater robustness also show greater phenotypic variability. This observation is based on thousands of randomly sampled phenotypes, and is thus independent of any one particular phenotype. It is a generic feature of RNA genotype space.

To understand this observation, one needs to understand two different phenomena with opposite effects on phenotypic variability. The first of these is the number of different phenotypes in the neighborhood of any one genotype. This number will be lower for highly robust phenotypes, because their genotypes have, on average, more neighbors with unchanged phenotype.

The second phenomenon is the rate at which a population spreads through a genotype network. This rate is determined by the likelihood that a mutation is deleterious, that is, that it does not preserve the phenotype *P*. Individuals suffering deleterious mutations are eliminated from the population, which slows the population's

355 diversification. The greater the incidence of such mutations, the slower a population
356 spreads through genotype space. A lack of robustness thus acts like a brake on the
357 genotypic diversification process of a population. This diversification process is
358 important, because the fraction of unique phenotypes in the neighborhoods of two
359 genotypes increases with the distance between them (Figure 1). This means, as I
360 discussed earlier, that populations with greater (cryptic) genotypic diversity can access
361 more novel phenotypes through mutations. They have greater phenotypic variability.
362 In sum, considering only the first phenomenon, high phenotypic robustness entails
363 low variability. In contrast, considering the second phenomenon, high robustness entails
364 high variability. In evolving populations, these two phenomena have opposite effects on
365 variability. For RNA secondary structure phenotypes, the second phenomenon – greater
366 population diversity – exerts the dominant influence on phenotypic variability. This is
367 why more robust phenotypes have higher phenotypic variability overall (Wagner 2008)).
368 Observations from computational analyses like these can help us appreciate that
369 we must study the dynamics of evolving populations – not just individual genotypes – to
370 understand the quantitative link between robustness and phenotypic variability. A
371 combination of experimental evolution work and comparative analyses further indicate
372 that robustness also matters for real molecules.
373 One class of experiments worth highlighting regards chaperones, proteins that
374 assist other proteins in folding, and help maintain their fold and function. Chaperones can
375 reduce the effects of environmental stress, such as high temperature, and they can
376 eliminate the deleterious effects of some mutations that reduce protein stability and
377 abolish a protein's activity (Fares et al. 2002; Hartl & Hayer-Hartl 2002). In the language
378 of genotype space, one could say that a chaperone increases the size of the genotype
379 network of a particular phenotype, because it can render some mutations neutral that
380 would otherwise be deleterious or lethal. In other words, a chaperone can make a
381 phenotype more robust. Recent laboratory evolution experiments on four different
382 enzymes expressed in *E. coli* support this notion. Specifically, populations of these
383 enzymes tolerated twice as many amino acid changes and evolved greater genotypic
384 diversity when large amounts of a chaperone were present. One of these enzymes, a
385 phosphotriesterase that can hydrolyze the pesticide paraoxon, was also subjected to

laboratory evolution for activity on a new catalytic substrate, 2-naphtylhexanoate. Populations of this enzyme attained higher activity and specificity on the new substrate when the chaperone was overexpressed. In sum, high robustness – in this case induced by a chaperone – is associated with superior evolutionary adaptation (Tokuriki & Tawfik 2009a).

Laboratory evolution experiments of enzymes also provide relevant evidence independent from that of chaperones (Amitai et al. 2007; Bloom et al. 2006; Bloom et al. 2005). A case in point is cytochrome P450, which belongs to an enzyme superfamily whose members hydroxylate many different substrate molecules. The relevant experiments mutagenized different variants of this enzyme that differed in their thermodynamic stability, and in their robustness to mutations. The stable and more robust variants of cytochrome P450 more readily evolved the ability to hydrolyze new substrates, such as the anti-inflammatory compound naproxen (Bloom et al. 2006; Bloom et al. 2005).

Experiments like these can show how robustness can facilitate evolutionary adaptation on short, laboratory timescales. They are silent about how this relationship translates onto the enormous timescales on which proteins diversified in life's history. Only a comparative analysis of the phenotypic diversity of proteins – a record of past evolutionary innovation – can answer this question. That is, it can answer whether highly robust protein phenotypes have adopted many different functions in their evolutionary history.

Such an analysis has become possible with the ability to estimate the robustness of protein folds (not just genotypes) to point mutations (England & Shakhnovich 2003), and to estimate the functional diversity of proteins, for example through well-catalogued enzyme functions. A recent study of 112 ancient protein folds showed that highly robust folds have evolved greater functional diversity, using different and complementary measures of functional diversity (Ferrada & Wagner 2008).

In sum, evidence that ranges from computational to comparative and experimental suggests that more phenotypic robustness can increase the ability of RNA and protein molecules to adapt and diversify in evolution. The computational work I discussed earlier in this section helps explain why: Phenotypic robustness accelerates the

spreading of populations through a genotype network, makes a broader spectrum of phenotypes accessible through mutation, and thus increases the odds of encountering a beneficial phenotype.

Robustness can help avoid conflicts between individuals and populations in bringing forth phenotypic variation. It is sometimes stated that biological systems bring forth novel features because this ability has been "selected for". This assertion is naïve and problematic. To see why, consider mutator alleles, variants of genes that can increase an organism's mutation rate (and phenotypic variability) dramatically (Sniegowski et al. 2000; Taddei et al. 1997). Mutators can be quite abundant in bacterial populations (Taddei et al. 1997). A facile explanation for their abundance resorts to the advantage they confer to a *population*: they help create many new phenotypes. Even though most of these new phenotypes may be deleterious, the few beneficial phenotypes may help the population survive in a challenging environment. However, this advantage is overshadowed by a great disadvantage to the *individual* – typically just one in a large population – who first acquires a mutator: Because most mutations are deleterious, carrying the mutator genotype is detrimental to this individual, and will thus often lead to its extinction (Sniegowski et al. 2000). A conflict thus exists between the interests of a population and that of an individual. How this conflict is resolved may depend on details of a population's life history and environment. Sometimes the conflict may be resolved in favor of the population, at other times in favor of the individual. In the latter case, variability would be reduced. Thus, the emergence of phenotypic variability in evolution is not a foregone conclusion. Similar conflicts exist for other mechanisms that facilitate phenotypic variability (Kirschner & Gerhart 1998).

Robustness as a variability principle, however, has a remarkable property: it can avoid this conflict. In RNA and proteins, where more robustness promotes greater variability, the interests of the individual and the lineage can be perfectly aligned. This is a simple consequence of how robustness influences the evolutionary dynamics of populations. Consider a population where stabilizing selection maintains a well-adapted phenotype. If this phenotype is highly robust, it is not easily perturbed through mutation or environmental changes, because the two kinds of robustness are usually correlated (Ancel & Fontana 2000; Lehner 2010; Masel & Siegal 2009; Meiklejohn & Hartl 2002;

Szollosi & Derenyi 2009; Wagner 2005b). Such robustness is advantageous for an individual that has this phenotype, because this individual experiences fewer perturbations with deleterious effects. At the same time, it is also advantageous for populations of such individuals. The reason lies in the evolutionary dynamics I discussed in the preceding section: Robust phenotypes in both RNA and protein molecules show greater phenotypic variability, and can become phenotypically more diverse on evolutionary time scales (Ferrada & Wagner 2008; Wagner 2008). Robustness can thus benefit both an individual and its lineage. Evolutionary conflicts are among the most serious impediments to adaptation, which makes their avoidance here even more significant (Futuyma 2009). Their general role in the evolution of phenotypic variability needs further study.

Summary and open questions. In sum, I have distinguished between two roles of robustness in evolutionary adaptation and innovation, a structural and a dynamic role. First, robustness causes the existence of genotype networks, complex web-like structures formed by genotypes with the same phenotype, which facilitate phenotypic variability. Second, a robust phenotype can help the evolutionary exploration of new phenotypes in macromolecules by accelerating the dynamics of change in an evolving population.

Many open questions persist in this young research field. They fall into two broad classes. The first regards quantitative aspects of evolutionary dynamics. How do the sizes of evolving populations and their mutation rates interact with robustness to influence phenotypic variability? Do the principles I discuss here also apply in environments that change rapidly and continually, where populations always track a moving optimal phenotype? Do these principles apply to systems with extremely high or low robustness? Does robustness also accelerate the evolutionary exploration of new phenotypes in systems other than macromolecules, such as evolving regulatory circuits or metabolic networks? Population genetic models and computational analyses of genotype-phenotype relationships are beginning to tackle these questions (Ancel & Fontana 2000; Draghi et al. 2010; Draghi & Wagner 2008; Espinosa-Soto et al. 2010; Rodrigues & Wagner 2011). However, we still lack a sufficient body of concordant evidence from different approaches to draw general conclusions.

A second class of questions regards evolutionary changes in robustness itself. Experimental and comparative work suggests that the robustness of macromolecules can change on evolutionary time scales (Montville et al. 2005; Sanjuan et al. 2006; Wagner & Stadler 1999). If robustness benefits both individuals and populations, then natural selection may favor robust phenotypes. If so, the robustness of phenotypes might increase over time. Only tentative evidence exists that naturally occurring phenotypes may be unusually robust (Cowperthwaite et al. 2008; Jörg et al. 2008). We do not yet know the causes of this robustness, we do not yet have relevant evidence from other system classes, and we are ignorant about the population genetic conditions under which such an increase would occur. Only with such evidence will we be able to answer a last and most fundamental question: Does robustness evolve in a way that facilitates evolutionary adaptation and innovation?

Acknowledgments:

I acknowledge support through Swiss National Science Foundation grants 315230-129708, as well as through the YeastX project of SystemsX.ch, and the University Priority Research Program in Systems Biology at the University of Zurich.

Figure Captions

Figure 1: Connected genotype networks facilitate accessibility of diverse phenotypes. The figure schematically represents a hypothetical set of genotypes (small open black circles) in genotype space (rectangle) that share the same phenotype and form a genotype network; neighboring genotypes are connected by black lines. Colored circles indicate genotypes with different phenotypes. The two large dashed circles denote the neighborhood of two different genotypes on the genotype network. The upper left neighborhood contains two novel phenotypes (blue and orange), whereas the lower right neighborhood contains two different novel phenotypes (beige and green). The figure illustrates that many different novel phenotypes can be accessed from a connected genotype network that spreads far through genotype space. Note that a two-dimensional figure like this cannot capture many features of high dimensional genotype spaces. These include that individual genotypes can have thousands of neighbors, not just the few shown here, and that the phenotypes shown in color also have vast genotype networks that are not shown. Adapted from (Wagner 2011c). **Permission from OUP requested.**

Figure 2: Robustness makes many phenotypic variants accessible to mutations. Circles in a) and b) represent genotypes with some hypothetical phenotype *P*, straight lines connect a genotype to its neighbors (not all neighbors of a genotype are shown as circles), solid lines connect a genotype to a neighbor with the same phenotype *P*, dashed lines connect a genotype to a neighbor with a new phenotype. **a)** The hypothetical genotype *G* shown here has no robustness, that is, no neighbors with the same phenotype. All eight of its neighbors have new phenotypes. It thus shows maximal phenotypic variability. **b)** All three genotypes shown are to some extent robust, that is, they have neighbors with the same phenotype *P*. Dark, medium, and light blue dashed lines point to genotypes with new phenotypes that are one, two, and three mutations away from the left-most genotype in b). Robustness makes more new genotypes accessible. See text for details. **c)** illustrates this principle through the actual number of new accessible phenotypes for three different natural RNA molecules (horizontal axis), and for computationally predicted (Hofacker et al. 1994) minimum free secondary structure

phenotypes in their neighborhood. Each of these molecules has some phenotype *P* (not shown). The black bar in each of the three panels indicates the maximally possible number of different phenotypes one mutation away from an RNA genotype *G*. This number is equal to 3*L*, where *L* is the number of nucleotides in a molecule. It would be attained only in the absence of robustness, as in panel a). Dark, medium, and light blue bars indicate, just as in panel b), the number of distinct new phenotypes that are accessible in the neighborhood of the molecule *G* ("1 mutation away"), in the neighborhoods of all its neighbors *G'* with phenotype *P* ("2 mutations away"), and in the neighborhood of the neighbors *G''* of *G'* with phenotype *P* ("3 mutations away). Data in c) are averages (error bars: one standard error of the mean) from ten inversely folded (Hofacker et al. 1994) RNA genotypes per RNA secondary structure phenotype. The individual RNA molecules have been obtained from the functional RNA database (http://www.ncrna.org/frnadb) (Kin T et al. 2007). They include a guide RNA (*Trypanosoma brucei*, fRNAdb accession number L25590, *L*=40nt), a hammerhead ribozyme (*Schistosoma mansoni*, acc. no: AF036740, *L*=43), and a telomerase (*Moneuplotes crassa*, acc. no: AF061109; *L*=33nt). See (Jörg et al. 2008, Table S1) for predicted secondary structure phenotypes *P* of these RNA molecules.

Figure 3: Cryptic variation can facilitate evolutionary adaptation. a) The large rectangles in both panels represent a genotype space into which a hypothetical genotype network is inscribed (gray open circles connected by gray lines). The colored circles symbolize individuals in a population on this genotype network. The left population (blue circles) is less genotypically diverse, and thus contains less cryptic variation than the population on the right (yellow circles). **b)** A laboratory evolution experiment showing how fast two populations of ribozymes with indistinguishable phenotype (catalytic activity on an RNA substrate) but different amounts of cryptic genotypic variation adapt evolutionarily to a new RNA substrate. As in panel a), blue and yellow correspond to populations with little and much cryptic variation. The horizontal axis shows time in generations, the vertical axis shows a measure of the biochemical activity of each

560 population on the new RN substrate. The population with more cryptic variation adapts
561 faster (Hayden et al. 2011).
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583

584 **References**

585

586 Amitai, G., Gupta, R. & Tawfik, D. 2007 Latent evolutionary potentials under the neutral
587 mutational drift of an enzyme *HFSP Journal; doi:10.2976/1.2739115*.
588 Ancel, L. W. & Fontana, W. 2000 Plasticity, evolvability, and modularity in RNA.
589 *Journal of Experimental Zoology/Molecular Development and Evolution* **288**,
590 242-283.
591 Bastolla, U., Porto, M., Roman, H. E. & Vendruscolo, M. 2003 Connectivity of neutral
592 networks, overdispersion, and structural conservation in protein evolution.
593 *Journal of Molecular Evolution* **56**, 243-254.
594 Beaudry, A. A. & Joyce, G. F. 1992 Directed evolution of an RNA enzyme. *Science* **257**,
595 635-541.
596 Beckert, B., Nielsen, H., Einvik, C., Johansen, S. D., Westhof, E. & Masquida, B. 2008
597 Molecular modelling of the GIR1 branching ribozyme gives new insight into
598 evolution of structurally related ribozymes. *EMBO Journal* **27**, 667-678.
599 Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. 2006 Protein stability
600 promotes evolvability. *Proceedings of the National Academy of Sciences of the*
601 *United States of America* **103**, 5869-5874.
602 Bloom, J. D., Romero, P. A., Lu, Z. & Arnold, F. H. 2007 Neutral genetic drift can alter
603 promiscuous protein functions, potentially aiding functional evolution. *Biology*
604 *Direct* **2**.
605 Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C. & Arnold, F. H.
606 2005 Thermodynamic prediction of protein neutrality. *Proceedings of the*
607 *National Academy of Sciences of the United States of America* **102**, 606-611.
608 Cheng, C. 2006 Old proteins and DNA, new tricks: Molecular evolution of antifreeze
609 proteins in polar fishes. *Comparative Biochemistry and Physiology a-Molecular*
610 *& Integrative Physiology* **143**, S61-S61.
611 Ciliberti, S., Martin, O. C. & Wagner, A. 2007a Circuit topology and the evolution of
612 robustness in complex regulatory gene networks. *PLoS Computational Biology*
613 **3(2): e15**.
614 Ciliberti, S., Martin, O. C. & Wagner, A. 2007b Innovation and robustness in complex
615 regulatory gene networks *Proceedings of the National Academy of Sciences of*
616 *the U.S.A.* **104**, 13591-13596
617 Cooper, T. F., Morby, A. P., Gunn, A. & Schneider, D. 2006 Effect of random and hub
618 gene disruptions on environmental and mutational robustness in Escherichia coli.
619 *BMC Genomics* **7**.
620 Cowperthwaite, M. C., Economo, E. P., Harcombe, W. R., Miller, E. L. & Meyers, L. A.
621 2008 The ascent of the abundant: How mutational networks constrain evolution.
622 *PLoS Computational Biology* **4**.
623 Draghi, J., Parsons, T., Wagner, G. P. & Plotkin, J. 2010 Mutational robustness can
624 facilitate adaptation. *Nature* **463**, 353-355.
625 Draghi, J. & Wagner, G. R. 2008 Evolution of evolvability in a developmental model.
626 *Evolution* **62**, 301-315.
627 England, J. L. & Shakhnovich, E. I. 2003 Structural determinant of protein designability.
628 *Physical Review Letters* **90**, 218101.

629 Espinosa-Soto, C., Martin, O. C. M. & Wagner, A. 2010 Phenotypic robustness can
630 increase phenotypic variability after non-genetic perturbations in gene regulatory
631 circuits. *Journal of Evolutionary Biology* **24**, 1284-1297.
632 Eyre-Walker, A. & Keightley, P. D. 2007 The distribution of fitness effects of new
633 mutations. *Nature Reviews Genetics* **8**, 610-618.
634 Eyre-Walker, A., Keightley, P. D., Smith, N. G. C. & Gaffney, D. 2002 Quantifying the
635 slightly deleterious mutation model of molecular evolution. *Molecular Biology*
636 *and Evolution* **19**, 2142-2149.
637 Fares, M. A., Ruiz-Gonzalez, M. X., Moya, A., Elena, S. F. & Barrio, E. 2002
638 Endosymbiotic bacteria - GroEL buffers against deleterious mutations. *Nature*
639 **417**, 398-398.
640 Ferrada, E. & Wagner, A. 2008 Protein robustness promotes evolutionary innovations on
641 large evolutionary time scales. *Proceedings of the Royal Society of London Series*
642 *B-Biological Sciences.* **275**, 1595-1602.
643 Futuyma, D. J. 2009 *Evolution*. Sunderland, MA: Sinauer.
644 Gibson, G. & Reed, L. K. 2008 Cryptic genetic variation. *Current Biology* **18**, R989-
645 R990.
646 Goodman, M., Pedwaydon, J., Czelusniak, J., Suzuki, T., Gotoh, T., Moens, L., et al.
647 1988 An evolutionary tree for invertebrate globin sequences. *Journal of*
648 *Molecular Evolution* **27**, 236-249.
649 Hardison, R. C. 1996 A brief history of hemoglobins: Plant, animal, protist, and bacteria.
650 *Proceedings of the National Academy of Sciences of the United States of America*
651 **93**, 5675-5679.
652 Hartl, D. & Clark, A. 2007 *Principles of population genetics*. Sunderland, MA: Sinauer
653 Associates
654
655 Hartl, F. U. & Hayer-Hartl, M. 2002 Protein folding - Molecular chaperones in the
656 cytosol: from nascent chain to folded protein. *Science* **295**, 1852-1858.
657 Hayden, E., Ferrada, E. & Wagner, A. 2011 Cryptic genetic variation promotes rapid
658 evolutionary adaptation in an RNA enzyme. *Nature* **474**, 92-95.
659 Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, L., Tacker, M. & Schuster, P. 1994
660 Fast folding and comparison of RNA secondary structures. *Monatshefte fuer*
661 *Chemie* **125**, 167-188.
662 Huang, Z. & Szostak, J. W. 2003 Evolution of aptamers with secondary structures from a
663 new specificity and new an ATP aptamer. *RNA-a Publication of the RNA Society*
664 **9**, 1456-1463.
665 Jörg, T., Martin, O. & Wagner, A. 2008 Neutral network sizes of biological RNA
666 molecules can be computed and are not atypically small. *BMC Bioinformatics* **9**,
667 464.
668 Kern, A. D. & Kondrashov, F. A. 2004 Mechanisms and convergence of compensatory
669 evolution in mammalian mitochondrial tRNAs. *Nature Genetics* **36**, 1207-1212.
670 Khersonsky, O. & Tawfik, D. S. 2010 Enzyme promiscuity: A mechanistic and
671 evolutionary perspective. In *Annual Review of Biochemistry, Vol 79*, vol. 79 (ed.
672 R. D. Kornberg, C. R. H. Raetz, J. E. Rothman & J. W. Thorner), pp. 471-505.

673 Kin T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, et al. 2007 fRNAdb: a
674 platform for mining/annotating functional RNA candidates from non-coding RNA
675 sequences. *Nucleic Acids Research (Database Issue)* **35** D145-D148.
676 Kirschner, M. & Gerhart, J. 1998 Evolvability. *Proceedings of the National Academy of*
677 *Sciences of the United States of America* **95**, 8420-8427.
678 Kitano, H. 2004 Biological robustness. *Nature Reviews Genetics* **5**, 826-837.
679 Kulathinal, R. J., Bettencourt, B. R. & Hartl, D. L. 2004 Compensated deleterious
680 mutations in insect genomes. *Science* **306**, 1553-1554.
681 Lehner, B. 2010 Genes confer similar robustness to environmental, stochastic, and
682 genetic perturbations in yeast. *PLoS ONE* **5**, e9035.
683 Li, H., Helling, R., Tang, C. & Wingreen, N. 1996 Emergence of preferred structures in a
684 simple model of protein folding. *Science* **273**, 666-669.
685 Li, W.-H. 1997 *Molecular evolution*. Massachusetts: Sinauer.
686 Lipman, D. & Wilbur, W. 1991 Modeling neutral and selective evolution of protein
687 folding. *Proceedings of the Royal Society of London Series B* **245**, 7-11.
688 Masel, J. & Siegal, M. L. 2009 Robustness: mechanisms and consequences. *Trends in*
689 *Genetics* **25**, 395-403.
690 Masel, J. & Trotter, M. V. 2010 Robustness and evolvability. *Trends in Genetics* **26**, 406-
691 414.
692 Meiklejohn, C. & Hartl, D. 2002 A single mode of canalization. *Trends in Ecology &*
693 *Evolution* **17**, 468-473.
694 Milton, C. C., Huynh, B., Batterham, P., Rutherford, S. L. & Hoffmann, A. A. 2003
695 Quantitative trait symmetry independent of Hsp90 buffering: Distinct modes of
696 genetic canalization and developmental stability. *Proceedings of the National*
697 *Academy of Sciences of the United States of America* **100**, 13396-13401.
698 Montville, R., Froissart, R., Remold, S. K., Tenaillon, O. & Turner, P. E. 2005 Evolution
699 of mutational robustness in an RNA virus. *Plos Biology* **3**, 1939-1945.
700 O'Brien, P. J. & Herschlag, D. 1999 Catalytic promiscuity and the evolution of new
701 enzymatic activities. *Chemistry & Biology* **6**, R91-R105.
702 Pigliucci, M. 2008 Is evolvability evolvable? *Nature Reviews Genetics* **9**, 75-82.
703 Reidys, C., Stadler, P. & Schuster, P. 1997 Generic properties of combinatory maps:
704 Neutral networks of RNA secondary structures. *Bulletin of Mathematical Biology*
705 **59**, 339-397.
706 Rodrigues, J. F. & Wagner, A. 2011 Genotype networks in sulfur metabolism. *BMC*
707 *Systems Biology* **5: 39**.
708 Rost, B. 2002 Enzyme function less conserved than anticipated. *Journal of Molecular*
709 *Biology* **318**, 595-608.
710 Samal, A., Rodrigues, J. F. M., Jost, J., Martin, O. C. & Wagner, A. 2010 Genotype
711 networks in metabolic reaction spaces. *BMC Systems Biology* **4:30**.
712 Sanjuan, R., Forment, J. & Elena, S. F. 2006 In silico predicted robustness of viroids
713 RNA secondary structures. I. The effect of single mutations. *Molecular Biology*
714 *and Evolution* **23**, 1427-1436.
715 Sawyer, S. A., Parsch, J., Zhang, Z. & Hartl, D. L. 2007 Prevalence of positive selection
716 among nearly neutral amino acid replacements in Drosophila. *Proceedings of the*
717 *National Academy of Sciences of the United States of America* **104**, 6504-6510.

718 Schultes, E. & Bartel, D. 2000 One sequence, two ribozymes: Implications for the
719 emergence of new ribozyme folds. *Science* **289**, 448-452.
720 Schuster, P., Fontana, W., Stadler, P. & Hofacker, I. 1994 From sequences to shapes and
721 back - a case-study in RNA secondary structures. *Proceedings of the Royal*
722 *Society of London Series B* **255**, 279-284.
723 Sniegowski, P., Gerrish, P., Johnson, T. & Shaver, A. 2000 The evolution of mutation
724 rates: separating causes from consequences. *Bioessays* **22**, 1057-1066.
725 Soskine, M. & Tawfik, D. S. 2010 Mutational effects and the evolution of new protein
726 functions. *Nature Reviews Genetics* **11**, 572-582.
727 Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J. & Doyle, J. 2004 Robustness of cellular
728 functions. *Cell* **118**, 675-685.
729 Szollosi, G. J. & Derenyi, I. 2009 Congruent evolution of genetic and environmental
730 robustness in micro-RNA. *Molecular Biology and Evolution* **26**, 867-874.
731 Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. & Godelle, B.
732 1997 Role of mutator alleles in adaptive evolution. *Nature* **387**, 700-702.
733 Tanner, M. A. & Cech, T. R. 1996 Activity and thermostability of the small self-splicing
734 group I intron in the pre-tRNA(IIe) of the purple bacterium Azoarcus. *RNA-a*
735 *Publication of the RNA Society* **2**, 74-83.
736 Thornton, J., Orengo, C., Todd, A. & Pearl, F. 1999 Protein folds, functions and
737 evolution. *Journal of Molecular Biology* **293**, 333-342.
738 Todd, A., Orengo, C. & Thornton, J. 1999 Evolution of protein function, from a structural
739 perspective. *Current Opinion in Chemical Biology* **3**, 548-556.
740 Tokuriki, N. & Tawfik, D. S. 2009a Chaperonin overexpression promotes genetic
741 variation and enzyme evolution. *Nature* **459**, 668-673.
742 Tokuriki, N. & Tawfik, D. S. 2009b Protein dynamism and evolvability. *Science* **324**,
743 203-207.
744 Wagner, A. 2005a Distributed robustness versus redundancy as causes of mutational
745 robustness. *Bioessays* **27**, 176-188.
746 Wagner, A. 2005b *Robustness and evolvability in living systems*. Princeton, NJ: Princeton
747 University Press.
748 Wagner, A. 2008 Robustness and evolvability: a paradox resolved. *Proceedings of the*
749 *Royal Society B-Biological Sciences* **275**, 91-100.
750 Wagner, A. 2011a Genotype networks shed light on evolutionary constraints. *Trends in*
751 *Ecology & Evolution* **26**, 577-584.
752 Wagner, A. 2011b The molecular origins of evolutionary innovations. *Trends in Genetics*
753 **27**, 397-410.
754 Wagner, A. 2011c *The origins of evolutionary innovations. A theory of transformative*
755 *change in living systems.* Oxford, UK: Oxford University Press.
756 Wagner, A. & Stadler, P. 1999 Viral RNA and evolved mutational robustness. *Journal of*
757 *Experimental Zoology/Molecular Development and Evolution* **285**, 119-127.
758 Weinreich, D. M. & Chao, L. 2005 Rapid evolutionary escape by large populations from
759 local fitness peaks is likely in nature. *Evolution* **59**, 1175-1182.
760 Wloch, D. M., Szafraniec, K., Borts, R. H. & Korona, R. 2001 Direct estimate of the
761 mutation rate and the distribution of fitness effects in the yeast Saccharomyces
762 cerevisiae. *Genetics* **159**, 441-452.
763

764