Plasticity, Evolvability and Modularity in RNA

Lauren W. Ancel
Walter Fontana

SFI WORKING PAPER: 1999-12-077

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu
Plasticity, Evolvability and Modularity in RNA

Lauren W. Ancel and Walter Fontana

1Department of Biological Sciences
Stanford University
Stanford, CA 94305 USA
ancel@charles.stanford.edu

2Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501 USA

and

3Institute for Advanced Study
Program in Theoretical Biology
310 Olden Lane
Princeton, NJ 08540 USA
walter@santafe.edu

Submitted to
Journal of Experimental Zoology (Molecular and Developmental Evolution)
SYNOPSIS

RNA folding from sequences into secondary structures is a simple yet powerful, biophysically grounded model of a genotype-phenotype map in which concepts like plasticity, evolvability, epistasis and modularity not only can be precisely defined and statistically measured, but reveal simultaneous and profoundly non-independent effects of natural selection. Molecular plasticity is viewed here primarily as the capacity of an RNA sequence to assume a variety of energetically favorable shapes by equilibrating among them at constant temperature (microenvironmental plasticity). Through simulations based on experimental designs, we study the dynamics of a population of RNA molecules that evolve towards a predefined target shape. Each shape in the plastic repertoire of a sequence contributes to the overall fitness of the sequence in proportion to the time the sequence spends in that shape. Plasticity is costly, since the more shapes a sequence can assume, the less time it spends in any one of them. Unsurprisingly, selection leads to a reduction of plasticity (environmental canalization). The most striking observation, however, is the simultaneous slow-down and eventual halting of the evolutionary process. The reduction of plasticity entails genetic canalization, that is, a dramatic loss of variability (and hence a loss of evolvability) to the point of lock-in. The causal bridge between environmental canalization and genetic canalization is provided by a correlation between the set of shapes in the plastic repertoire of a sequence and the set of dominant (minimum free energy) shapes in its genetic neighborhood. This statistical property of the RNA genotype-phenotype map, which we call plastogenetic congruence, steers and then traps populations in regions where most genetic variation is phenotypically neutral. We call this phenomenon neutral confinement. Analytical models of neutral confinement, made tractable by the assumption of perfect plastogenetic congruence, formally connect mutation rate, the topography of phenotype space and evolvability. These models identify three mutational regimes: that corresponding to neutral confinement, a classic error threshold corresponding to the loss of the dominant phenotype, and an exploration threshold corresponding to a break-down of neutral confinement with the simultaneous persistence of the dominant phenotype. In a final step, we analyze the structural properties of canalized phenotypes. Surprisingly, the reduction of plasticity leads to an extreme modularity, which we define and analyze from several perspectives: thermophysical (melting behavior, the RNA version of a norm of reaction), kinetic (folding pathways, the RNA version of development), and genetic (transposability, the insensitivity of modular traits to changing genetic context). The model thereby suggests a possible evolutionary origin of modularity.
# Contents

1 INTRODUCTION ........................................ 4

2 RNA FOLDING AS A GENOTYPE-PHENOTYPE MAP ................. 9
  2.1 Why RNA? ........................................ 9
  2.2 Secondary structure ................................ 9
  2.3 Robust properties of RNA folding .................... 12
  2.4 Plasticity in RNA ................................ 17
  2.5 Epistasis in RNA ................................ 21

3 SIMULATED EVOLUTION IN PLASTIC AND NON-PLASTIC RNA POPULATIONS .......... 24
  3.1 Model Setup .................................... 24
  3.2 Discussion of Sample Runs ........................ 26
  3.3 Summary ....................................... 37

4 PLASTOGENETIC CONGRUENCE AND NEUTRAL CONFINEMENT ............... 39
  4.1 Plastogenetic congruence .......................... 39
  4.2 Neutral confinement .............................. 53
  4.3 Structural lock-in ............................... 59

5 ANALYTIC MODELS OF NEUTRAL CONFINEMENT ........................... 64
  5.1 Model Assumptions ............................... 64
  5.2 Frequency distribution within a neutral network ... 64
  5.3 Three-Tiered Model: Exploration and Error Thresholds ... 69
  5.4 Neutrality in the sub-network $C_0$ ................. 71
  5.5 Exploration and Error Catastrophes .................. 74
  5.6 Finite population birth-death model .................. 76

6 MODULARITY ....................................... 81
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Norms of Reaction - Melting Behavior</td>
<td>81</td>
</tr>
<tr>
<td>6.2</td>
<td>Context Insensitivity</td>
<td>86</td>
</tr>
<tr>
<td>6.3</td>
<td>Kinetics - Modularity and Funnels</td>
<td>87</td>
</tr>
<tr>
<td>7</td>
<td>DISCUSSION AND OUTLOOK</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>REFERENCES</td>
<td>97</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Biological evolution is the phenotypic transformation of organisms through time. Evolution is fueled by the introduction of novel phenotypes, and steered by population-level interactions including natural selection and genetic drift. The predominant route to heritable phenotypic change originates with genetic mutation. The processes that translate genetic variation into phenotypic variation give rise to an association between genotype and phenotype which we represent as a map that is sensitive to environmental boundary conditions. Concepts such as canalization, epistasis and modularity underlie our understanding of phenotypic variability (for a sweeping perspective see Wagner and Altenberg (1996) and Schlichting and Pigliucci (1998)). Yet, a micro-foundation of these concepts and of their interconnections in terms of the relation between genotype and phenotype is largely missing. Our goal here is to initiate such a foundation in the specific context of a conceptually, computationally and empirically simple yet powerful genotype-phenotype model based on the folding of RNA sequences (genotypes) into shapes (phenotypes).

RNA folding seems, at first, unlikely of being able to address canalization, epistasis and modularity. Although we shall mostly use language appropriate for RNA, readers versed in genetics are invited to bluntly metaphorize an RNA sequence to stand for a genome, and a position along the sequence to play the role of a locus with four possible alleles (nucleotides). A phenotype (RNA shape) then is a simple pattern of gene-gene interactions (base pairs, see section 2.2). Since this paper weaves together several seemingly diverse concepts, we begin with an overview.

Our study provides a molecular illustration for the Simpson-Baldwin effect (Baldwin, 1896; Simpson, 1953; Ancel, 1999a) using RNA as an example. Central to the initial “discovery” stage of the Simpson-Baldwin effect is phenotypic plasticity, that is, the genetically influenced capacity of an individual to assume a range of phenotypes. In an evolutionary context, a fixed environment will convey a selective advantage to those individuals that can access an improved phenotype within their plastic repertoire over those who cannot. The assimilation stage in the Simpson-Baldwin effect arises from the fitness costs of plasticity. Among the individuals selected in the first stage, those that can still access the improved phenotype while reducing their range
of phenotypic plasticity will have a selective advantage. This process of reducing plasticity is also known as environmental canalization (Waddington, 1942). The Simpson-Baldwin effect describes the genetic determination of a phenotype that previously seemed to be acquired anew in each generation. A frequently considered mechanism of plasticity is learning (Hinton and Nowlan, 1987; Maynard-Smith, 1987).

Although molecules do not learn, biopolymers like RNA are plastic in the sense that a given sequence can realize a repertoire of alternative structures, rather than being frozen in its minimum free energy configuration (section 2). An RNA sequence samples a variety of energetically low lying structures by wiggling among them under thermal fluctuation. The overall time a sequence spends in a shape reflects the thermodynamic stability of that shape. Consider now a population of replicating and mutating RNA sequences that are subject to selection for structural proximity to a constant target shape (section 3). Each shape attainable by a sequence contributes to that sequence’s overall fitness in proportion to the time the sequence spends in it.

The Simpson-Baldwin stages are immediate. Sequences with an advantageous but energetically suboptimal shape will be selected over those that lack that shape. Plasticity entails a fitness cost because the more alternative shapes an RNA molecule can fold into, the less time it will spend in each one. Sequences with an advantageous but energetically suboptimal shape $S$ will therefore first evolve to lower the energy of the shape to the point of making $S$ the minimum free energy structure, and subsequently evolve to fine tune the thermodynamic stability of $S$ by removing as many alternative structures as possible from its energetic vicinity. We show that in our RNA model such genetic assimilation occurs extremely rapidly and covers several orders of magnitude in thermodynamic stability (section 3).

Plasticity has often been linked to a speed-up in evolution (Baldwin expediting effect). The contrary turns out to be the case in our RNA model, where plasticity slows down evolution to the point of a phenotypic dead-end. In sections 3 and 4 we explain this behavior in terms of features that are intrinsic to the RNA genotype-phenotype map. Several threads come together as we argue that genetic assimilation requires a genotype-phenotype map in which plastic accessibility corresponds to genetic accessibility. In other words, the shapes appearing in a sequence’s repertoire of energetically favorable structures correlate significantly to the minimum free energy structures of the
one-error mutants of that sequence. This turns out to be a general property of the RNA genotype-phenotype map (section 4). We call this phenomenon plastogenetic congruence.

Phenotypic variability describes the extent of phenotypic variation accessible to a genotype through mutation. The evolvability of an individual is the likelihood of reaching a phenotype with improved fitness through mutation (Altenberg, 1995). As such, it is linked to variability via some fitness function. Plasticity, on the other hand, captures the phenotypic variation at a fixed genotype, typically induced by environmental heterogeneity. In this sense, the impacts of the environment on plasticity are analogous to those of mutation on genetic variability. Plastogenetic congruence means that plasticity and variability mirror each other: low plasticity (that is, environmental canalization or strong genetic determination) implies low variability (genetic canalization or strong mutational buffering), and vice versa.

Plastogenetic congruence implies that an evolutionary reduction of plasticity has a flip side: a decline in variability. This duality results in a self-defeating process in which the loss of plasticity through natural selection leads to the loss of phenotypic variability, which favors the further loss of plasticity to the extent of evolutionary lock-in.

Plastogenetic congruence also sheds light on Waddington’s theories for the evolution of organismal development (Waddington, 1957). He introduced two modes of evolution: environmental canalization is the honing of developmental pathways to reduce environmental noise, and genetic canalization is the integration of genetic factors to reduce the deleterious effects of mutation. Under plastogenetic congruence then, genetic canalization will ensue as a by-product of environmental canalization. This yields a mechanistic explanation of a hypothesis put forward by Wagner et al. (1997).

The genetical underpinning of variability is epistasis. Epistasis is the influence that an alteration of gene $i$ has on the phenotypic consequences of a subsequent alteration of gene $j$ (Wagner et al., 1998). In other words, epistasis is the genetic control of variability. In RNA, low plasticity coincides with low variability, maintained by the fixation of epistatic interactions that control neutrality. Remarkably, epistatic interactions in RNA can shut down variability (or increase mutational robustness) almost completely. This phenomenon, which we call neutral confinement, underlies the evolutionary lock-in mentioned earlier. Under certain fitness assignments to phenotypes,
neutral confinement leads to structures that are no longer evolvable even if the underlying sequences are removed from neutral confinement. We refer to this as **structural lock-in**, a phenomenon reminiscent of Wimsatt’s generative entrenchment (Wimsatt, 1986) and Riedl’s notion of burden (Riedl, 1975).

Analytical models built on a stylized version of plastogenetic congruence predict that, for limited parameter regimes, a sufficiently high mutation rate should prevent or undo the exploration catastrophe resulting from neutral confinement (while preserving the dominant phenotype). We discuss this issue in sections 4 and 5.

The RNA folding genotype-phenotype map enables not only an exploration of evolutionary dynamics but also a characterization of the morphological endpoints of evolution. In section 6 we compare three classes of sequences that share the same dominant (that is, minimum free energy) structure. One set is derived from a random sample of sequences with the given dominant structure, another set has evolved on a neutral network, and the third set results from genetic assimilation under the plastic genotype-phenotype map. The characteristic which best distinguishes among the three classes is modularity. We study modularity of shape characters from a variety of perspectives, all contributing to a definition and quantification of modular traits as “transposable characters” that maintain their structural integrity in different sequence and environmental contexts. Although the three classes share the same minimum free energy structure, that structure is not even remotely modular on the random sequences, while it is extremely modular on sequences that have experienced genetic assimilation.

We can parse minimum free energy structures to identify coherent characters (sub-structures), but not to assess the modularity of such characters. At a qualitative level, modularity is evident in comparisons of the melting behavior (norms of reaction) of modular and non-modular, yet syntactically identical traits. Sequences whose dominant shape is modular lose their shape characters in a discrete fashion at distinct temperatures. Moreover, the presence or absence of a character does not influence the remaining ones.

Genetic assimilation leaves us with sequences that possess modular shapes and are at the same time evolutionarily locked in. This seems to contradict the hypothesized evolutionary advantage of modularity: modularity partitions quantitative traits into independently and easily evolvable units (Wag-
ner and Altenberg, 1996). While modularity may indeed facilitate the quantitative polishing of a trait, it leads to an evolutionary lock-in with respect to significant structural modifications of that trait. Resistance to structural change is the hallmark of a module. As many scholars have suggested, the generation of further evolutionary novelty (or plasticity) then must shift from locked-in modules to the combinatorial arrangement of modules into new units.

We conclude and provide an outlook to future work in section 7.
2 RNA FOLDING AS A GENOTYPE-PHENOTYPE MAP

2.1 Why RNA?

RNA combines genotype and phenotype into a single molecule. This makes RNA folding in many respects a limited, but also a simple model of a genotype-phenotype map (section 2). As a model system, RNA has the advantages of both computational tractability (Waterman, 1995) and suitability as a substrate for test tube evolution experiments (Joyce, 1989; Landweber, 1999). The RNA sequence-structure relation also occupies a rare intermediate level of abstraction bridging the empirical and the formal. RNA folding algorithms are sufficiently realistic for computational discoveries to suggest worthwhile empirical investigations. At the same time, the RNA structures are sufficiently abstract for these discoveries to provide insight and to suggest axioms for the construction of simplified models that are analytically tractable.

2.2 Secondary structure

RNA molecules are heteropolymers of (predominantly) four units called ribonucleotides. Ribonucleotides have a ribose phosphate in common, but differ in the base attached to the sugar. The essence of an RNA sequence is therefore captured by a string over a four letter alphabet, each letter representing a particular base - A for adenine, U for uracil, C for cytosine, and G for guanine. Hydrogen bonds give rise to stereoselective recognition between certain base pairs, specifically A·U and G·C. This base pairing enables an RNA sequence to be copied into a negative and back again into a positive. The pairing is not always exact. Error rates depend on the molecular machinery that assists in pairing and ligating the bases. For example, the per nucleotide error rate is $7 \cdot 10^{-5} - 2.7 \cdot 10^{-4}$ for Influenza, $3 \cdot 10^{-4}$ for Coliphage Qβ (Eigen et al., 1989) and $2 \cdot 10^{-4} - 5 \cdot 10^{-4}$ for HIV (Loeb et al., 1999). In this way, base pairing enables heredity. We therefore treat an RNA sequence as a genotype. The same base pairing mechanism, however, also enables segments of a sequence to pair with other segments within the same sequence, causing it to fold back on itself into a three-dimensional structure. (For the
formation of an intramolecular structure \( \text{G} \cdot \text{U} \) is also an admissible pair.)

This structure conveys chemical behavior to the sequence and constitutes its phenotype.

The extremely reduced replication time and the minimal complexity of the phenotype make RNA a tractable laboratory model. RNA molecules can be evolved in the test tube using a variety of techniques for amplification, variation and selection. In fact, such experiments have shown that evolutionary optimization of RNA properties in the test tube occurs readily and effectively. Examples include the optimization of replicative efficiency (Mills et al., 1967; Spiegelman, 1971); the production of RNA molecules binding optimally to prespecified target molecules by means of the SELEX technique (Ellington and Szostak, 1990; Tuerk and Gold, 1990; Ellington, 1994) (see also a discussion on SELEX generated arginine aptamers and the genetic code (Knight and Landweber, 1998)); evolutionary induced changes in the activity and specificity of catalytic RNA molecules (Beaudry and Joyce, 1992), so-called ribozymes; and the evolutionary design of ribozymes with new functions (Bartel and Szostak, 1993; Ekland et al., 1995) or even multiple functions (Landweber and Pokrovskaya, 1999). For a recent review see Landweber (1999).

Molecular structure in RNA can be characterized at many levels of resolution. One empirically well-established notion is the secondary structure, which is the topology of binary contacts as they result from base pairing (Figure 1). The secondary structure is a useful abstraction, since the pattern of base pairs provides both a geometric and thermodynamic scaffold for the tertiary structure of the molecule. This puts the secondary structure in correspondence with functional properties of the tertiary structure. Consequently, selection pressures, and hence functional elements, become observable at the secondary structure level in terms of conserved base pairs.

A secondary structure on a sequence of length \( n \) can be represented as a graph of base pair contacts (Figure 1). The nodes of the graph stand for bases at positions \( i = 1, \ldots, n \) along the sequence. The set of edges includes the unspecific covalent backbone connecting node \( i \) with node \( i + 1 \), for \( i = 1, \ldots, n - 1 \), and a specific set of pairs \( i \cdot j \) indicating a base pair between positions \( i \) and \( j \), \( P = \{ i \cdot j \mid i \neq j \text{ and } j \neq i + 1 \} \). The set \( P \) has to satisfy two conditions: (i) every edge in \( P \) connects a node to at most one other node, and (ii) if both \( i \cdot j \) and \( k \cdot l \) are in \( P \), then \( i < k < j \) implies that
Figure 1: An RNA secondary structure graph. A secondary structure is a graph made of structural elements listed in Figure 2. Unpaired positions not enclosed by any base pair are called “external”. Components are shape features delimited by external bases.

Failure to meet condition (ii) results in pseudoknots which are interactions that belong to the next - the tertiary - level of structure. Both conditions distinguish RNA structure from protein structure, in particular condition (i) which builds RNA secondary structure exclusively from binary contacts. We use a picture of the graph (Figure 1) as the visually most immediate representation of a secondary structure. We sometimes use a more convenient line oriented representation of nested parentheses, such as “(((((((.))).))).(((.))))” , in which a dot stands for an unpaired position, and a pair of matching parentheses indicates positions that pair with one another.

The elements of a secondary structure are classes of loops (Figure 2): a hairpin loop occurs when one base pair encloses a number of unpaired positions, a stack consists in two base pairs with no unpaired positions, while an interior loop has two base pairs enclosing unpaired positions. An internal loop is called a bulge, if either side has no unpaired positions. Finally, the class
of multiple loops consists of loops delimited by more than two base pairs. A position that does not belong to any loop class is called external, such as free ends or joints between loops (see Figure 1). These loop classes provide a basis for assembling secondary structures. We make the reasonable assumption that the overall energy of a secondary structure is the sum of its loop energies. These have been measured and tabulated (Freier et al., 1986; Turner et al., 1988; Jaeger et al., 1989; He et al., 1991) as a function of loop size and the nature of the delimiting base pairs.

A stack of base pairs (a double-stranded or helical region) is the major stabilizing element. The formation of an energetically favorable stacking region, however, implies the formation of an energetically unfavorable loop that constrains unpaired bases. This “frustrated” energetics leads to a vast combinatorics of stack and loop arrangements constituting the structural repertoire of an individual RNA sequence. Structures possessing extremal properties with respect to a given sequence, such as minimizing the free energy, are of particular relevance. RNA is an excellent model system for sequence-structure relations in biopolymers precisely because of our ability to compute such minimum free energy secondary structures and statistical mechanics quantities, such as the partition function, by dynamic programming (Nussinov et al., 1978; Waterman and Smith, 1978; Nussinov and Jacobson, 1980; Zuker and Stiegler, 1981; Zuker and Sankoff, 1984; McCaskill, 1990). The work in this paper makes use of the Vienna RNA folding package (Hofacker et al., 1994-1998).

2.3 Robust properties of RNA folding

In the present work, RNA folding serves primarily as a model system for discovering and understanding qualitatively robust properties of genotype-phenotype relations and their influence on adaptive dynamics. RNA folding algorithms vary considerably in the accuracy of secondary structure prediction for individual instances (Huynen et al., 1997). We are seeking, however, robust properties that depend only on the logic of base-pair optimization intrinsic to RNA secondary structure, and this logic is adequately captured by present-day folding algorithms. The reader should keep in mind, therefore, that our main concern is with statistical properties of the sequence to structure map as a whole, rather than with specific cases. We neverthe-
Figure 2: Secondary structure elements.
less employ state-of-the-art computer programs (Zuker and Stiegler, 1981; Hofacker et al., 1994) that are routinely applied in the laboratory.

Our present work builds on robust statistical properties discovered computationally in previous work (Fontana et al., 1993a,b; Schuster et al., 1994; Gruner et al., 1996a,b). Some of these properties we describe next. They arise from two fundamental facts. Sequence space and shape space are both very high dimensional, and the sequence space is substantially larger than the shape space. Analytical tools developed in Stein and Waterman (1978) yield an upper bound of only $S_n = 1.48 \times n^{-3/2}(1.85)^n$ shapes vis à vis $4^n$ sequences, where $n$ is the sequence length (Hofacker et al., 1999). The mapping from sequences to minimum free energy shapes is therefore significantly many-to-one (even for a binary RNA alphabet).

The “typical shapes” property states that some shapes are realized much more frequently than others. These relatively few typical shapes are set apart from many rare shapes. More precisely, as sequence length goes to infinity, the fraction of “typical shapes” tends to zero (their number grows nevertheless exponentially), while the fraction of sequences folding into them tends to one. Consider a numerical example: The space of GC-only sequences of length $n = 30$ contains $1.07 \times 10^9$ sequences folding into 218,820 structures of which 22,718 ($= 10.4\%$) classify as typical (Gruener et al., 1996a,b), in the sense of shapes formed by more sequences than the average (Schuster, 1997). In this case, 93.4% of all sequences fold into these 10.4% shapes. This also implies that any statistical statement we make about the folding map, and a fortiori about its influence on adaptive dynamics, can only be expected to hold for the set of typical shapes.

**Neutrality and neutral networks**

There exist many sequences that have a given (typical) shape $\alpha$ as their minimum free energy structure. We call such sequences “neutral” (with respect to $\alpha$). In terms of the map from sequences to structures, a structure $\alpha$ identifies an equivalence class of sequences. A one-error mutant of a sequence that shares the same minimum free energy structure as that sequence is called a “neutral neighbor”. By “neutrality” of a sequence we mean the fraction of its $3n$ one-error mutants that are neutral neighbors. This notion of neutrality should not be confused with fitness neutrality which is a further layer of
degeneracy deriving from different phenotypes mapping to the same fitness value.

A typical sequence has a significant fraction of neutral neighbors, and the same holds for these neighbors. In this way, jumping from neighbor to neighbor, we can map an extensive mutationally connected network of sequences that fold into the same minimum free energy structure (Schuster et al., 1994; Reidys et al., 1997). We termed such networks “neutral networks” (Schuster et al., 1994).

Neutral networks are among the most remarkable properties of the RNA folding map. Although the structures remain the same throughout evolution on a neutral network, the sequences may change substantially from the point of entry on the network to the point of exit to another network associated with some superior phenotype. The ability to change the underlying sequence in the presence of strong selection pressure on its phenotype is critical for evolutionary dynamics. The role of neutrality has been typically viewed as a conservative one, buffering the phenotypic effects of mutations. Connected neutrality, however, also enables phenotypic change, because it permits the phenotypically-silent variation of a genotype in the context of which the next mutation can become consequential.

The boundary of a neutral network is the set of sequences that differ by one mutation from a sequence in the network, but are not themselves contained in the network. Transitions between structures are transitions between adjacent neutral networks. This suggests a measure of nearness between RNA structures based on the fraction of common boundary shared between their corresponding networks in sequence space (Fontana and Schuster, 1998a,b). RNA shape space is then organized by an accessibility relation based on the adjacency of neutral networks in genotype space. Such a topology enables a formal definition of continuous and discontinuous phenotypic change independently of any fitness criteria (Fontana and Schuster, 1998a). It thereby isolates properties of adaptive trajectories caused by the genotype-phenotype map, rather than by some arbitrary fitness landscape imposed on phenotypes. Figure 3 describes the RNA shape changes that are discontinuous in this accessibility topology.

In addressing structural plasticity, we depart from the above sense of neutrality, since two sequences with the same minimum free energy structure will differ in their suboptimal configurations. The notion of “neutral network”
Figure 3: **Continuous and discontinuous RNA shape transformations.** The figure illustrates transformations between RNA secondary structure motifs. Solid (dashed) arrows indicate continuous (discontinuous) transformations in the topology of Fontana and Schuster (1998b). Three groups of transformations are shown. Top: the loss and formation of a base pair adjacent to a stack are both continuous. Middle: the opening of a constrained stack (e.g., closing a multiloop) is continuous, while its creation is discontinuous. This reflects the fact that the formation of a long helix between two unpaired random segments upon mutation of a single position is a highly improbable event, whereas the unzipping of a random helix is likely to occur as soon as a mutation blocks one of its base pairs. Bottom: generalized shifts are discontinuous transformations in which one or both strands of a helix shift ending up with or without an overlap. Accordingly, generalized shifts are divided into the four classes shown.
remains, however, a very useful one. When we speak of neutrality in the context of plasticity, we always mean neutrality with respect to minimum free energy structures and not with respect to the entire plastic repertoire.

2.4 Plasticity in RNA

Plasticity is the genetic determination of a range of phenotypic possibility. A plastic individual can change phenotype during its lifetime through learning or by reacting to environmental stimuli. At the level of a single non-interacting molecule, plasticity appears in its simplest form. Unspecific contact with a heat bath triggers transitions between molecular configurations, provided the energy barriers between configurations are sufficiently low. For a particular RNA molecule, the range of accessible configurations depends on its energy landscape in fixed environmental conditions. The kinetic process of folding into a native configuration is the RNA analogue of development. Random transitions between kinetically accessible structures that are energetically close to the native configuration are an unavoidable consequence of the stochasticity of molecular motion, akin to “developmental noise”. A microscopic kinetic folding model for RNA has been developed recently (Flamm et al., 1999), but at present it is too involved for the goals of this paper. Instead, we exploit an extension of the standard thermodynamic minimum free energy folding algorithm which permits the computation of all secondary structures within some energy range above the minimum free energy (Wuchty et al., 1999). This suboptimal folding algorithm yields fast access to the low energy portion of the secondary structure space of a given sequence. We neglect energy barriers and assume that a sequence equilibrates among all structures whose free energy is within $5kT$ from the minimum free energy configuration. The $5kT$ choice amounts to approximately 3 kcal at 37° C, and corresponds to the loss of two $G\cdot C/C\cdot G$ stacking interactions. Under thermodynamic equilibration, the Boltzmann probability of a structure $\sigma$, $\exp(-\Delta G_\sigma/kT)/Z$, corresponds to the overall fraction of time that the molecule spends in $\sigma$, where $\Delta G_\sigma$ is the free energy of structure $\sigma$, $k$ is the Boltzmann constant, $T$ the absolute temperature, and $Z = \sum_\tau \exp(-\Delta G_\tau/kT)$ the partition function. The latter is computed by an algorithm described in McCaskill (1990).

This defines a genotype-phenotype map that assigns to a sequence a set of
Figure 4: Genotype-phenotype maps. The lower part illustrates the simple map that assigns to a sequence its minimum free energy structure as the only phenotype. The upper part schematizes the plastic map that assigns to a sequence the thermodynamic spectrum of shapes within $5kT$ (typically $T = 310.15K$). If shape $\sigma$ has free energy $\Delta G_\sigma$, the sequence is assumed to spend a fraction of time in $\sigma$ that is given by its Boltzmann factor, $\exp(-\Delta G_\sigma/kT)/Z$. The mfe structure is the dominant phenotype, in the sense that the sequence spends the largest fraction of time in it.
structures and their occupation times. We shall refer to this map as the “plastic map”, to distinguish it from the “simple map” where a sequence is associated with its minimum free energy structure only (Figure 4).

In an evolutionary context (section 3), the obvious advantage for a plastic sequence that covers a broad spectrum of structures is the increased likelihood that some of them contribute to an improved fitness. As pointed out in Scheiner (1993) the cost of plasticity may be in terms of maintaining the cellular machinery required for plasticity, rather than the direct impact on fitness due to the realization of a particular plastic trait. In the stripped down world of RNA there is no such machinery beyond the molecule itself. Yet, the cost of plasticity is evident: the broader the range of structures, the less time the sequence will spend in any one of them. Thus, even if some structures constitute an improvement, this can easily be undermined by a small occupancy time. Incidentally, this is analogous to Schmalhausen’s argument (Schmalhausen, 1949) that one cost of plasticity is given by “erroneous” phenotypic changes (Ancel, 1999a). The erroneous phenotypes in this case are inevitable residencies in detrimental structures.

Biologists have drawn a distinction between two kinds of plasticity: phenotypic plasticity proper refers to macroenvironmental sensitivity, while environmental variance refers to flexible responses to microenvironmental parameters (Waddington, 1957; Gavrilets and Hastings, 1994). The distinction is appropriate for RNA as well. The diversity of structures for a given sequence as described above is environmental variance, the microenvironment being a heat bath at constant temperature. Phenotypic plasticity proper would be revealed, for example, by the melting behavior of an RNA sequence, that is, the suite of minimum free energy structures (plus suboptimals) at distinct temperatures. Most of the work reported here pertains to environmental variance which we simply call plasticity. In section 6, however, we focus on phenotypic plasticity proper (macroenvironmental variance) when comparing the melting behavior of sequences obtained by various processes. The correlation we observe between the phenotypic consequences of microenvironmental and macroenvironmental variation suggests that the distinction may be a matter of degree and not of kind.

The number of suboptimal structures per RNA sequence varies over several orders of magnitude. A simple listing of all suboptimal structures is therefore not a practical characterization of plasticity. A more statistical indication
Figure 5: **Base pairing uncertainties.** Base pairing uncertainties are given by equation (1). The figure shows two dramatically different distributions of high and low uncertainty pertaining to sequences that share the same minimum free energy structure. The upper graph shows the base pairing entropy of a randomly chosen sequence (not shown) that folds into the structure shown at the top of the figure; the lower part refers to a low-plasticity sequence (not shown) with the same minimum free energy structure. This sequence was obtained through a canalizing process discussed in section 3.
is provided by a measure introduced originally by Huynen et al. (1997) to assess the well-definedness of an RNA minimum free energy structure of a given sequence $S$. We measure the Shannon entropy of the base pairing probability distribution of position $i$ in sequence $S$. Let $p_{ij}$ be the fraction of time that $i$ spends paired with $j$ in the ensemble of all suboptimal structures of $S$ (calculated with the partition function algorithm (McCaskill, 1990)). Additionally, let $p_{i0}$ be the fraction of time that $i$ is unpaired in the suboptimals of $S$, $p_{i0} = 1 - \sum_j p_{ij}$. The entropy $\Pi^S_i$ of the $p_{ij}$-distribution for a given $i$ then indicates the diversity and probability of alternative interactions available to position $i$:

$$\Pi^S_i = -\sum_{j=0}^{L} p_{ij} \log(p_{ij})$$

where $L$ is the length of $S$. The maximum value of $\Pi^S_i$ is $\log L$. If $i$ always occurs paired to some site $j$, or if $i$ is always unpaired in the suboptimals of $S$, then $\Pi^S_i = 0$. Although this scenario never occurs, different sequences with the same minimum free energy structure can differ substantially in their plasticity. An example is shown in Figure 5. The upper part shows the distribution of $\Pi^S_i$ for a sequence with high plasticity whose minimum free energy structure is marginally stable, while the lower part refers to a canalized sequence with extremely reduced plasticity that is the outcome of an evolutionary process described later in section 3.

### 2.5 Epistasis in RNA

While plasticity refers to an individual RNA sequence, epistasis is a phenomenon related to mutational effects. The phenotypic consequences of a mutation at position $i$ depend on the genetic background provided by the remaining sequence. In particular, we assay the effect that a modification of position $j$ has on the phenotypic consequences of a subsequent mutation at position $i$ (Wagner et al., 1998).

Consider, for example, the sequence and its minimum free energy structure at the center of Figure 6. We call a sequence position “neutral” if at least one out of three possible mutations at that locus leave the structure invariant. In Figure 6, neutral positions, like the one labelled $x$, are marked with grey bullets. Mutating position $x$ from $C$ to $G$ does not alter the structure.
Figure 6: Epistasis.
two structures at the bottom of the figure illustrate the effects of this neutral substitution at $x$ on the phenotypic consequences of mutating position $y$ from $G$ to $A$.

The top part of Figure 6 emphasizes another facet of epistasis. Consider again the neutral mutation from $C$ to $G$ at position $x$ yielding the (same) structure shown at the top left of Figure 6. The grey bullets mark the neutral positions in the new sequence. The '+' symbols indicate positions that have become neutral as a result of this mutation, while the '-' symbol marks a position that has lost its neutrality. This illustrates that even if a point mutation does not affect the structure, it can alter the extent of neutrality across the sequence. Epistasis as the genetic control of neutrality plays an important role in what follows.
3 SIMULATED EVOLUTION IN PLASTIC AND
NON-PLASTIC RNA POPULATIONS

3.1 Model Setup

In this section we describe evolving populations of RNA sequences. Under
the simple genotype-phenotype map (section 2.4) a sequence $S_j$ has a single
minimum free energy structure $\sigma_j^0$ as its phenotype. To determine the fitness
of $S_j$, we compare $\sigma_j^0$ to a prespecified target structure $\tau$.

This fitness function is motivated by an experimental protocol in which RNA
sequences are evolved, for example, to optimally bind a ligand. RNA se-
quences are artificially selected by running the RNA sample through a column
that has the ligand tethered to the filling material. The desirable portion
of the sample remains bound in the column and is eluted by a solvent with suit-
able ionic strength (Tuerk and Gold, 1990; Ellington and Szostak, 1990). The
evolutionary end product is typically unpredictable in the laboratory situa-
tion. Its possible shape(s) are, however, implicitly specified by the choice of
the ligand, like a simple lock specifies its key. We are not seeking RNA shapes
with particular chemical properties, since little is known about the link be-
tween an RNA structure and its binding properties or catalytic behaviors.
Instead, we specify the optimal shape directly at the outset, shortcircuit-
ing the indirection through ligands in the laboratory. We then study the
evolutionary dynamics, the evolutionary histories, and the thermodynamic
properties of evolutionary outcomes.

We define the selective value $f(\sigma)$ of shape $\sigma$ as a hyperbolic function of the
Hamming distance $d(\sigma, \tau)$ between the parenthesized representations (section
2.2) of $\sigma$ and the target $\tau$:

$$f(\sigma) = \frac{1}{0.01 + d(\sigma^0, \tau)/n},$$

where $n$ is the length of the structure. The results reported here are robust
to changes in the functional form of the selective value. In particular, they
also hold under linear and exponential forms.

The fitness $r_j$ of a sequence $S_j$ is its replication rate constant, and for the
non-plastic case we simply take it to be:

\[ r_j = f(\sigma^j_0). \]  

(3)

Our population evolves in a chemical flow reactor whose outflow is regulated to maintain a nearly constant total sequence population size (Eigen, 1971). The entire system is simulated in terms of continuous time stochastic chemical reaction kinetics. (For a description see Fontana and Schuster (1987); Huynen et al. (1996); for a general simulation technique see Gillespie (1976, 1977).) The number of replication events per time unit in the flow reactor depends on the replication rate constants of the individual sequences comprising a population, and changes over time as the population evolves to larger rate constants. To meaningfully compare different runs, we plot statistics along replication events, rather than the time shown by an external clock.

Point mutations are the sole source of genetic variation in our simulations. Unless otherwise stated, the replication accuracy per nucleotide position is 0.999, the sequence length is 76 (which corresponds to a short tRNA sequence), and the reactor capacity is 1000 sequences. Initial conditions typically consist of a homogenous population of a randomly generated sequence species.

The plastic case is a simple extension of the control case above. For a sequence \( S_j \), we consider all of its suboptimal structures \( \sigma^j_i \) whose free energy \( E(\sigma^j_i) \) is within an interval of size \( \Delta \) above the minimum free energy \( E(\sigma^j_0) \). The \( \sigma^j_i \) are indexed with increasing energy, where index 0 refers to the ground state, and the set of all suboptimal structures accessible within \( [E(\sigma^j_0), E(\sigma^j_0) + \Delta] \) is denoted with \( \sigma^j(\Delta) \). Unless otherwise specified, \( \Delta = 5kT \), where \( T \) is the absolute temperature, which is fixed at 310.15K (37° C), and \( k \) is the Boltzmann constant, \( k = 1.98717 \cdot 10^{-3} \) kcal/K. In other words, \( \Delta = 3.08 \) kcal. Given our model setup, the “environment” consists in the target shape \( \tau \) and the temperature \( T \). Throughout this paper we only consider (macroscopically) fixed environments.

The selective value of a suboptimal structure \( \sigma^j_i \) is given by \( f(\sigma^j_i) \) (equation 2), exactly as in the non-plastic case. For a plastic sequence, however, \( \sigma^j_i \) contributes this value to the overall fitness of the sequence in proportion to the time the sequence spends in it. We think of a sequence flowing through the selection column as switching among its alternative structures, such that
each structure contributes to the overall binding probability of the sequence in proportion to its Boltzmann factor. Hence,

\[
r_j = \sum_{\sigma_i^j \in \sigma^j(\Delta)} f(\sigma_i^j) \cdot \frac{\exp(-E(\sigma_i^j)/kT)}{Z},
\]

(4)

where \(Z\) is the partition function. The assumption here is that a sequence actually achieves thermodynamic equilibration among its alternative structures. Due to possibly high energy barriers between the structures this may be kinetically unrealistic on time scales shorter than the experiment. Our understanding of the kinetic process of RNA folding is too immature at present to assess this assumption.

3.2 Discussion of Sample Runs

RNA shape space topology is invariant to plasticity

Figures 7 and 8 juxtapose the progress of a simple and a plastic population that evolve towards a tRNA target shape. In the following we will denote population averages with \(\langle \cdot \rangle\). The black curves trace \(\langle d(\sigma_0^j, \tau) \rangle\), the average distance between the target and the minimum free energy shapes realized by the sequences in the population. The grey curves monitor \(\langle d(\sigma^j, \tau) \rangle\), the population average of the weighted shape distance, which is, for any given sequence, the Boltzmann weighted average distance between its suboptimal shapes and the target:

\[
da(\sigma^j, \tau) = \frac{1}{Z} \sum_{\sigma_i^j \in \sigma^j(\Delta)} d(\sigma_i^j, \tau) \cdot \exp(-E(\sigma_i^j)/kT).
\]

(5)

A fitness independent relation of (genetic) accessibility between shapes (neutral networks), induced by the RNA folding map, organizes the space of RNA shapes and underlies the observed punctuated dynamics. This is discussed in detail elsewhere (Fontana and Schuster, 1998a,b; Schuster and Fontana, 1999), and is briefly reviewed in section 2.3. Steps correspond to, or are barely
Figure 7: Evolution under plastic and simple genotype-phenotype maps, Example 1. These graphs depict the evolution of the population average of the minimum free energy distance to target (black) and the average weighted shape distance to target (grey). (See text for definitions.) The inset magnifies the trajectories through the first few million replication events. In the plastic case, fitness improvements (not shown) correspond to transitions in the average weighted distance to the target (not to the distance of the minimum free energy structure). The target is a tRNA cloverleaf structure, the population size is logistically constrained to fluctuate around 1000 sequences, and the replication accuracy is 0.999 per position.
preceded by, hard-to-achieve transitions between neutral networks that share only a small common boundary (these are, technically speaking, discontinuous transformations in the topology of RNA phenotype space). Shape transformations associated with such transitions are characterized by the simultaneous shift of several base pairings in a single mutational event. Flat periods correspond to genetic diffusion along a shape-neutral network or both genetic and phenotypic diffusion along several shape-neutral networks that are fitness-neutral with respect to each other. While natural selection generally pushes the population toward the target, the long periods of apparent stasis occur when the population shifts among a set of phenotypically equivalent genotypes that are mutationally isolated from higher fitness phenotypes.

Figure 8: Evolution under plastic and simple genotype-phenotype maps, Example 2. See caption to Figure 7.

In comparing the plastic to the control runs, we immediately recognize the same pattern of punctuated dynamics after a rapid initial relaxation. Plasticity seems not to change the topology of shape space in a way that alters the punctuated character of the dynamics. Below we discuss this further.
In the plastic case, $\langle d(\sigma_j^0, \tau) \rangle$ dictates the trajectory of $\langle d(\sigma_j, \tau) \rangle$. The minimum free energy structure of a sequence $S$ is usually closer to the target than any other suboptimal structure of $S$. If we were to let $d_{\sigma_j}$ be the Boltzmann weighted average distance to target over all suboptimals, not just those within $\sigma^j(\Delta)$, then $\langle d_{\sigma_j} \rangle$ would necessarily be larger than $\langle d(\sigma_j^0, \tau) \rangle$. But $\langle d(\sigma_j, \tau) \rangle$, which considers only the shapes in $\sigma^j(\Delta)$, remains less than $\langle d(\sigma_j^0, \tau) \rangle$. This suggests that the suboptimal structures are not much worse than the minimum free energy structure and that their probabilities are relatively small. This is dramatically in contrast to the statistics for the control run with the simple map where the $\langle d(\sigma_j^0, \tau) \rangle$ fluctuates wildly at values much higher than $\langle d(\sigma_j^0, \tau) \rangle$.

No Baldwin expediting effect

Figures 7 and 8 represent the outcomes of a large set of similar simulations: evolution proceeds much more slowly under the plastic map, and plastic populations never approach the target as closely as do control (simple map) populations, which often attain the target (not shown). We take notice of this striking effect here, but defer its explanation to section 4 in which we develop the necessary concepts.

Here we take issue with a generalization often drawn from the Hinton and Nowlan class of models (Hinton and Nowlan, 1987; Maynard-Smith, 1987) that phenotypic plasticity expedites evolution by effectively smoothing the adaptive landscape. More specifically, they claim that plastic individuals scan a wider variety of nearby phenotypes and therefore have a better chance of locating advantageous ones than non-plastic individuals who are blind to nearby possibilities. This (implicitly) assumes a correlation between the plastically accessible and the genetically accessible phenotypic repertoires of an individual. In section 4 we find this indeed (and surprisingly) to be true in RNA (plastogenetic congruence). We will show, however, that plastogenetic congruence ultimately works against evolution. In this section, we argue against a significant speed-up, even in the early stages of the evolutionary process. On average, plasticity does not significantly shorten the duration of the early periods of stasis preceding phenotypic innovations. This is revealed in a comparison of the early evolutionary trajectories under the plastic map and the simple map (see, for example, the inset of Figure 7, which magni-
fies the first few million replications). Recent work specifies the restrictive population-genetic conditions for a Baldwin expediting effect (Ancel, 1999b). In the present case, however, the absence of an expediting effect results from features intrinsic to the RNA genotype-phenotype map itself.

Consider the discovery of the first plastic individual $S$ with $\alpha$ as the minimum free energy structure and some advantageous suboptimal configuration $\beta$. Assume further that $\beta$ is not a neighbor of $\alpha$ in the shape space topology sensu Fontana and Schuster (1998a), which is typically the case at the major phenotypic transitions in the simulations. We claim that, on average, the discovery of individual $S$ cannot happen significantly faster in the plastic population than the first appearance of an individual with $\beta$ as its minimum free energy shape in a non-plastic control population dominated by $\alpha$.

First we note that the individual $S$ sought above is always possible in principle. If we relax conditions by asking for a sequence $S'$ that can assume both shapes $\alpha$ and $\beta$ irrespective of their energies, that is, dropping the requirement of $\beta$ being in the $5kT$-band and of $\alpha$ being the minimum free energy structure, then it is easy to find such a sequence $S'$ for any choice of $\alpha$ and $\beta$. This was proved by Reidys et al. (1997). We define a sequence to be compatible with a structure, if it can form that structure, regardless of where it ranks energetically. A given structure $\alpha$, then, determines a compatibility set consisting in all sequences compatible with $\alpha$. Consequently, our individual $S'$ must be an element of the intersection between the compatibility set of $\alpha$ and the compatibility set of $\beta$. The theorem of Reidys et al. (1997) states that the intersection of any two compatibility sets is non-empty. The actual size of these intersection sets is computed in Weber (1997).

Figure 9 provides a non-technical illustration of a sequence that satisfies two structures. A sequence can only obtain a structure if positions that are supposed to pair are occupied by nucleotides that are actually able to pair. Satisfying two structures at once entails additional constraints. If the two structures are “independent”, meaning that positions that pair in one structure are disjoint from positions pairing in the other, the constraints simply add up. If, however, the two structures are non-independent, additional constraints arise because certain positions must now satisfy two pairings simultaneously, and, depending on how the two structures are related, such constraints propagate along chains of dependencies. The constraints on nucleotides for a sequence that must satisfy two particular structures increase
Figure 9: **Satisfying two structures at once.** (After Weber (1997).) Top: Two structures are shown that differ in a “shift” – a transformation in which one strand slides past the other (see Figure 3). The two structures are represented in the middle as circles with the base pairs drawn as chords. The additional constraints on a sequence that arise from having to satisfy both structures are easily seen by superimposing the circle representations (bottom). Position 1 must have a nucleotide that can pair with both position 13 and 12. But the nucleotide choices at 12 must take into account its complementarity to position 2 as well, and so on until the end of this chain is reached at position 9. It is fairly immediate that any such chain can be satisfied, and hence that there exists at least one sequence compatible with any two structures. (A secondary structure can be thought of as a permutation acting on sequence positions. For example, the permutation corresponding to the structure on the left specifies that positions \{5, 6, 7, 8, 9\} are fixed points, and that 4 is assigned to 10, 10 to 4, 3 to 11, 11 to 3, etc. The two (involutive) permutations corresponding to the two structures generate a dihedral group, and the chain of constraints in the bottom circle is one orbit obtained by the action of this group on the sequence positions (Reidys and Stadler, 1996; Reidys et al., 1997).)
with the number and length of such chains. The most severe constraints arise from structure pairs that differ in shifts and their generalization (see Figures 3 and 9). These highly constrained sequences bridge precisely the transformations that are discontinuous in the topology developed in Fontana and Schuster (1998a,b).

Under the simple map, the discontinuous transformations responsible for the evolutionary transitions of Figures 7 and 8 are all transitions between rarely adjacent neutral networks. Under the plastic map, such transformations would be expedited if it were easy for a population to locate the intersection region of the corresponding compatibility sets. Yet, this region is smallest precisely for these difficult discontinuous transformations. We conclude that if the transformation of $\alpha$ into $\beta$ is discontinuous, an evolutionary transition between two minimum free energy structures $\alpha$ and $\beta$ in the simple control case is approximately as difficult as finding a sequence that realizes both $\alpha$ and $\beta$ among its suboptimals. In the plastic case, the intermediary must not only be compatible with $\alpha$ and $\beta$, but it must have $\alpha$ as its minimum free energy structure and $\beta$ within $5kT$. This makes a possible expediting effect even more unlikely. These arguments suggest that the organization of the RNA genotype-phenotype map prevents a speed-up of the evolutionary discovery process in our plastic simulations. The difficulty of discontinuous phenotype transitions is not lessened by allowing for plasticity.

Environmental canalization (reduction of plasticity)

In Figure 10 we monitor the time course of five population plasticity statistics for the plastic case. We discuss three of them here. Curve $i$ in Figure 10 shows the average fraction of the partition function realized within $5kT$, which is

$$z^j = \frac{1}{Z} \sum_{\sigma^j_i \in \sigma^j(\Delta)} \exp(-E(\sigma^j_i)/kT).$$  \hspace{1cm} (6)

This indicates the total probability of the structures in $5kT$ relative to all structures attainable by $S_j$ at any energy, that is, how much of the partition function falls within $5kT$. Curve $iv$ shows the average number of structures realized in $5kT$, $\langle|\sigma^j(\Delta)|\rangle$, while curve $v$ tracks the average structure variance,
\[ \langle \text{var} \sigma^j(\Delta) \rangle, \]

with

\[ \text{var} \sigma^j(\Delta) = \frac{d(\sigma^j_i, \sigma^0_i)^2}{d(\sigma^j_i, \sigma^0_i)^2} - \frac{d(\sigma^j_i, \sigma^0_i)^2}{d(\sigma^j_i, \sigma^0_i)^2}, \tag{7} \]

where the bar indicates an average over the structures in \( \sigma^j(\Delta) \). In equation (7) we make use of the fact that the notion of variance can be generalized to sets (here of structures) for which a mean does not exist, but for which one can compute pairwise distances between the elements. Finally \( \text{var} \sigma^j(\Delta) \) conveys information about the structural diversity in the plastic repertoire of sequence \( S_j \), while \( |\sigma^j(\Delta)| \) simply counts the structures. These are three quantitative measures of RNA plasticity. We juxtapose these to adaptive events indicated by the evolutionary traces (curves \( vi \) and \( vii \)).

The corresponding graphs for the control case with the simple map, which are not displayed here, fluctuate wildly without correlation to evolutionary events. After an initial period, \( \langle |\sigma^j(\Delta)| \rangle \) levels off to between 126 and 254 structures for the control case. In dramatic contrast, the plastic curve (iv) reach much lower, fairly constant levels of only about 30, 13, and 20 structures, indicating a population-wide reduced plasticity compared to both the control case and the initial condition (204 structures in \( 5kT \)). This loss of plasticity occurs very rapidly. The reduction is also evidenced by the average fraction of the partition function (curve i) that rapidly climbs to 0.9 and then settles at 0.95, while the control case remains in the interval \([0.65, 0.75]\).

The structure variance of curve v (plastic case: \([7, 15]\), versus control case: \([46, 92]\)) indicates a reduction in plasticity as alternative structures become more similar to the ground state.

The transition from the first appearance of a better shape as a suboptimal to its appearance as the ground state of a sequence (phase (iia) above) is rapid and does typically not transpire at the population level. From \( \langle z^j \rangle \) (curve i) in Figures 10 and 11 we see that phase (iib), plasticity reduction, is more prominent. Note, however, that aside from the initial relaxation, a major reduction in plasticity occurs only after the first transition. Transitions are marked by dashes running vertically through all figures. Subsequent transitions (2 and 3 in Figure 10) are associated with a dip in \( \langle z^j \rangle \) (arising from the change to a new minimum free energy structure whose Boltzmann probability is at first smaller compared to the structure it replaces) and a subsequent rapid assimilation to pre-transition levels. This pattern appears more blurred at transition 6 in Figure 10, which is associated with a small adaptive improvement.
Figure 10: **Plastic evolution, Example 1.** The time evolution of five plasticity statistics (i-v) are shown in conjunction with the major shape transitions (curves vi and vii). See text for the definitions of the quantities monitored. The inset enlarges a population-level Simpson-Baldwin effect discussed in the text.
Example of a population-level Simpson-Baldwin effect

Transitions 4 and 5 are somewhat unique and have a more classical Simpson-Baldwin flavor. (Refer to the Introduction for a description of these stages.) At transition 4, a new advantageous sequence species $S$ with a suboptimal structure $\beta$ that is closer to the target than its minimum free energy shape $\alpha$ begins to invade the population. After $10^6$ replication events the first sequences $S'$ appear with $\beta$ as their minimum free energy structure. The population retains $\alpha$, however, as a frequent suboptimal shape with high Boltzmann weight. Mutation actively shifts around weights between $\alpha$ and $\beta$. Although $\alpha$ is selectively worse than $\beta$, this can be offset by a higher Boltzmann probability. Consequently $\alpha$ remains the most prevalent minimum free energy structure. This shows up as a discrepancy between the population averages of the weighted distance to target – $\langle d(\sigma^j, \tau) \rangle$ – and the distance based only on the minimum free energy structures – $\langle d(\sigma^0_j, \tau) \rangle$. Upon $2 \cdot 10^5$ further replications, sequences with $\beta$ as the minimum free energy structure take over the population, and $\langle d(\sigma^0_j, \tau) \rangle$ consequently drops towards $\langle d(\sigma^j, \tau) \rangle$. The situation reverses, however, as Boltzmann probabilities shift. After $6.5 \cdot 10^5$ replications $\beta$ disappears completely as a minimum free energy structure, and is observed only as a suboptimal. Finally, $3 \cdot 10^5$ replications later the transition back to sequences with $\beta$ as their minimum free energy structure occurs, accompanied by a realignment of $\langle d(\sigma^j, \tau) \rangle$ with $\langle d(\sigma^0_j, \tau) \rangle$. The plot of $\langle z^j \rangle$ indicates a significant subsequent reduction in plasticity that stabilizes $\beta$ (label “f” in graph (i) of Figure 10). In terms of minimum free energy structures, events 4 and 5 taken together constitute one single transition that occurs in two stages. Event 4 is the onset of the Simpson-Baldwin discovery phase which lasts until event 5 when the assimilation phase begins. This is the only example we encountered of both Simpson-Baldwin phases being revealed at the population level.

Fitness-neutral plasticity

We now examine the composition of structural repertoires through the joint behavior of the two remaining plasticity statistic. Curve $ii$ (black) monitors the average of the Boltzmann probabilities of the minimum free energy structures across the population. For the sake of brevity, we refer to it as
the “mfe-probability”. Curve iii tracks an important feature that we have not yet discussed. A sequence frequently has suboptimal configurations with the same fitness as its minimum free energy structure. We have discussed neutrality predominantly as defined by minimum free energy structures or sequence fitness. Here we measure the neutrality within the plastic repertoire of a sequence. For a given sequence $S$, we add up the Boltzmann probabilities of all suboptimal structures that are at the same structural distance from the target shape as the minimum free energy structure of $S$. In other words, this is the amount of time $S$ spends in configurations other than the minimum free energy structure that have the same selective value as the minimum free energy structure. Curve iii (grey) in Figure 10 displays the population average of this measure which we call “fitness-neutral plasticity”.

![Graph](image)

**Figure 11: Plastic evolution, Example 2** As in Figure 10, but only the population averages of three quantities are shown for the initial evolutionary trajectory. The top curve is the fraction of the partition function, the next to bottom curve is the number of structures in the plastic repertoire, and the bottom one is the structure variance in the plastic repertoire. The two middle curves graph the evolutionary trajectory of the plastic population.
Label “e” in Figure 10 marks the trajectories of the average mfe-probability and average fitness-neutral plasticity during the Simpson-Baldwin event of transitions 4 and 5. Both decrease in unison, reflecting the drop in stability of the minimum free energy structures, the presence of fitness-superior suboptimal structures (that is, structures with higher selective values than the minimum free energy structure), and the relative unimportance of plastic neutrality to this event. Contrast this with the events labelled “a”, “b”, “c” and “d”. During these periods of relative fitness stasis, mfe-probability spikes down, while fitness-neutral plasticity spikes up. This signals the appearance of high-probability fitness-neutral alternative configurations within the plastic repertoire of the most frequent sequences. These fitness-neutral alternatives detract probability from the minimum free energy structure (and from each other), displaying phenotypic diversification within the plastic repertoire. For example, at event “d”, sequences appear that have, as their lowest energy suboptimals, three fitness-neutral shapes with Boltzmann weights of 0.45, 0.28, and 0.17. Such sequences also tend to have an increased overall number of suboptimal configurations, because each fitness-neutral shape comes with its own small family of thermodynamically linked structures. This effect is small for “b”, “c” and “d”, but quite pronounced for “a”. Conversely, event “g”, shows the opposite behavior. Shortly after transition 6, the dominant minimum free energy structure is very well-defined. Then a considerable amount of fitness-neutral plasticity arises and persists for a long time. At event “g”, that neutrality disappears and different sequences with well-defined minimum free energy structures dominate the population for a relatively short time.

3.3 Summary

A quantitative characterization of RNA evolution under the plastic and the simple genotype-phenotype maps, reveals the following:

- The punctuated character of the evolutionary dynamics holds for RNA populations evolving under both the simple and the plastic genotype-phenotype map.

- The populations evolving under the plastic map halt evolutionary progress much farther away from the target structure than those evolving under
the simple map.

- Plastic populations show no evolutionary speed-up (or Baldwin expediting effect); overall they exhibit a strong slow-down.

- A strong reduction in plasticity occurs in the plastic case. The Simpson-Baldwin effect is observed along individual founding lineages causing the transitions. The discovery phase, however, typically occurs at the micro-level and does not transpire at the population level. We observed one example for a classical Simpson-Baldwin effect with both phases observable at the population level.

- Fitness-neutral plasticity is pervasive and can prevent plasticity from being reduced to the biophysical limit.
4 PLASTOGENETIC CONGRUENCE AND NEUTRAL CONFINEMENT

4.1 Plastogenetic congruence

Section 3 demonstrates that a plastic population evolving towards a constant target undergoes a dramatic loss of variation of plastic repertoires, and ultimately hits an evolutionary dead-end. In the current chapter, we construct a causal bridge from the loss of plasticity to the loss of evolvability by arguing a significant correlation between the structures available to a sequence in its plastic repertoire and the structures present in the mutational vicinity of that sequence. First we use a concrete example to provide intuition for the mechanistic underpinnings of such a correlation. Then through statistical evidence, we claim that it is fundamental property of the genotype-phenotype map.

The top left corner of Figure 12 shows a short sequence $A$ and its minimum free energy structure $\alpha$ together with the list of all suboptimal configurations within $5kT$ of the minimum free energy. While the sequence spends 51% of the time in $\alpha$, its relative stability is only marginal, since it competes with a number of energetically close suboptimal structures. These easily accessible alternative structures exist because base pair interactions holding together the stem in $\alpha$ are weak. The entire stem can switch into a different position, as in the 4th suboptimal configuration $\beta$. We say therefore that $\alpha$ is not "well-defined" as the minimum free energy structure.

Significantly, $\alpha$’s marginal stability also translates into an increased sensitivity to mutations. For example, mutating a position inside $\alpha$’s loop (arrow) suffices to tip the thermodynamics in favor of $\beta$. The result is a change in the minimum free energy structure from $\alpha$ to $\beta$ (upper right corner of Figure 4).

Now consider a neutral mutation - one that does not displace $\alpha$ as the minimum free energy structure - that transforms $A$ into $A'$ (lower left corner of Figure 4). Although the structure of $\alpha$ is not altered, the mutation strengthens the thermodynamic stability of the stacking region of $\alpha$. The fraction of the partition function taken up by $\alpha$ increases from 51% to 88%. Notice that the set of alternative structures available to $A'$ within $5kT$ has become
considerably smaller compared to \( A \). In addition, the few alternative structures that occur with appreciable probability have become more similar to the ground state. The relative stability of \( \beta \), which has disappeared from the set of \( 5kT \)-suboptimal structures, has dropped by one order of magnitude. The consequence is an increased robustness of \( \alpha \) towards mutations. In fact, when we make the same nucleotide change in \( A' \) that we did in \( A \) to acquire \( B \), \( \alpha \) remains the minimum free energy structure.

The upper arrow from \( A \) to \( B \) illustrates the correlation between structures in the plastic repertoire and structures in the mutational vicinity that we introduced at the beginning of this chapter. In particular, this point mutation changes a \((5kT-)\)suboptimal structure into the minimum free energy structure. It is not difficult to imagine that a sequence \( A \) with \( \beta \) in its plastic repertoire is much more likely to have a genetic neighbor with \( \beta \) as the minimum free energy structure, than a sequence \( A' \) that lacks \( \beta \) in its plastic repertoire. The arrows from \( A \) to \( A' \) and further to \( B' \) illustrate again the epistatic control of neutrality (section 2.5). This is a special case of the correlation just described: the more time a sequence spends in its minimum free energy structure, the higher the fraction of neutral neighbors in its genetic vicinity.

To summarize, we call a shape \( \sigma \) “plastically accessible” to a sequence \( S \) if \( \sigma \) is within the plastic repertoire of \( S \). Likewise, \( \sigma \) is “genetically accessible” from \( S \), if there exists a one-error mutant \( S' \) of \( S \), such that \( \sigma \) is the minimum free energy structure of \( S' \). The alignment of plastic accessibility with genetic accessibility means that the set of shapes a particular sequence can fold into (plastic accessibility) strongly correlates to the minimum free energy shapes realized by its one-error mutants (genetic accessibility). We call this property of the genotype-phenotype map “plastogenetic congruence”. Figure 13 illustrates the concept.

The genetic accessibility of phenotypes constrains the evolutionary trajectory of the control simulations with the simple map (Fontana and Schuster, 1998a,b). In the plastic simulations, however, plastic accessibility, through its alignment with genetic accessibility, has an equally profound impact on the evolutionary dynamics. We return to this after a quantitative demonstration of plastogenetic congruence for the RNA folding map.

Ideally we would perform statistics on a broad sampling of genotypes to show the extent of overlap between plastic repertoires and mutant-neighboring
Figure 12: Plasticity and minimum free energy shapes of genetic neighbors. The figure illustrates how the plasticity of a sequence correlates with the minimum free energy structures of one-error mutants. Arrows show point mutations. The mutation from B to B' involves the same nucleotide substitution at the same position as the mutation from A to A'. The only difference is the slight change in context due to the neutral mutation from A to B. See text for details.
Figure 13: **Plastogenetic congruence.** The top part shows a sequence with its alternative configurations in $5kT$ (energy increases meandering left to right, then right to left, bottom to top, as in Figure 4). The minimum free energy structure is framed by the light grey box. The white structure $\alpha$ in the dark box is a shape that is plastically accessible to the sequence. The arrow indicates a single point mutation (the affected nucleotide is shown enlarged) to a neighboring sequence for which $\alpha$ is the minimum free energy structure. This shows that $\alpha$ is genetically accessible from the top sequence. Plastogenetic congruence means that (generally) a structure $\alpha$ is plastically accessible from from a sequence $S$ if and only if $\alpha$ is genetically accessible from $S$. In RNA, this is realized in a statistical sense.
minimum free energy structures. In lieu of this computationally prohibitive approach, we present three pieces of partial evidence for plastogenetic congruence:

1. The frequency of a structure \( \beta \) as a minimum free energy structure in the one-mutant neighbors of a sequence \( S \) is significantly larger for sequences that have \( \beta \) in their plastic repertoires than for sequences that do not.

2. The minimum free energy structure \( \alpha \) of a sequence \( S \) is present at high frequency in the plastic repertoires of one-mutant neighbors.

3. For any advantageous shape \( \beta \) in the plastic repertoire of any sequence \( S \), \( S \) can typically evolve to another sequence \( S' \) with \( \beta \) as its minimum free energy structure, in only (on average) five point mutations.

In the first approach we exploit our ability to easily generate random sequences \( S_\alpha \) with a particular structure \( \alpha \) as the minimum free energy structure ("inverse folding", see Hofacker et al. (1994)). Since two sequences that share the same minimum free energy structure often share other suboptimal structures, this simple procedure yields sets of sequences \( S_\beta \alpha \) that share the same ground state \( \alpha \) and that also have a particular shape \( \beta \) in their \( 5kT \)-plastic repertoire. We simultaneously obtain control samples of sequences \( S_\alpha \) that specifically lack the shape \( \beta \) from their plastic repertoire. Next, we compute the minimum free energy shapes of all one-mutant neighbors of sequences in the sample \( S_\beta \alpha \) and in the control sample \( S_\alpha \), and compare the frequencies of sequences with \( \beta \) as the minimum free energy shape. A few examples are collected in table 4.1. A systematic exploration of the possible shape combinations \( \alpha \) and \( \beta \) is unfeasable. Table 4.1 provides anecdotal evidence for our proposition: A sequence (with minimum free energy structure \( \alpha \)) that has a structure \( \beta \) among its \( 5kT \) suboptimals has a significantly increased likelihood of having a one-error mutant with \( \beta \) as its minimum free energy structure, compared to a sequence (with minimum free energy structure \( \alpha \)) that lacks \( \beta \) among its \( 5kT \) suboptimals.

Table 4.1 shows ratios of likelihoods. The probability of finding a structure \( \beta \) as the suboptimal of a sequence with minimum free energy structure \( \alpha \) increases if \( \beta \) is in the neighborhood of \( \alpha \) in the sense of the shape topology developed in Fontana and Schuster (1998a,b). We were therefore able to
Plastogenetic congruence I. Samples of sequences with a tRNA cloverleaf as the minimum free energy structure $\alpha$ (bottom line) and various structures $\beta$ as suboptimal configurations were generated as described in the text. The frequency of $\beta$ becoming the minimum free energy structure upon one point mutation was computed for the sample of sequences having $\beta$ as a suboptimal and for a control sample of sequences lacking $\beta$ within a $5kT$ range from $\alpha$. The ratio of the former to the latter is tabulated. Two kinds of frequencies were computed: the frequency with which $\beta$ occurs as a ground state among all 1-error neighbors (“occurrence frequency”, and corresponding o-ratio) and the frequency with which it occurs at least once in a 1-error neighborhood (“neighborhood frequency”, and corresponding n-ratio). For example, row 1 states that the shown shape occurs 8 times more frequently as a minimum free energy structure in the 1-error neighborhood (ratio $^b$) of a sequence that has that shape as a suboptimal compared to one that lacks it. The first part of the table is based on samples with more than 200 sequences with $\beta$ as a suboptimal, while the second part is based on sample sizes of more than 100 but less than 200 sequences. A given sequence can have several suboptimal configurations listed in this table.
attain larger samples (top of table) for suboptimal structures that are topologically near $\alpha$ (4.1). The likelihood ratios indicate that these structures are easily converted through mutation from a suboptimal to a minimum free energy structure. Even sequences with rarer suboptimals (bottom of table 4.1) have a significantly increased likelihood compared to the control sequences for making $\beta$ the ground state in a single mutation. The last row of table 4.1 shows the only counterexample we found to this claim for the tRNA cloverleaf case. Similar trends hold for samples based on other ground states $\alpha$, as well as for sequence lengths other than 76.

Figure 14: Plastogenetic congruence II. For each point in the graph, we generate a new sample of 100 sequences with the tRNA cloverleaf as their minimum free energy structure. All one-error mutants were scanned for the presence of the tRNA cloverleaf as a suboptimal configuration. The graph shows the frequency of such sequences as a function of the energy range $\Delta$ that defines the plastic repertoire. Curve $i$ refers to the frequency with which a sequence in the sample has at least one one-error mutant with the desired suboptimal structure (“neighborhood frequency”), while curve $ii$ shows the “occurrence frequency”. See text for details.

In a second approach, we generate a sample of sequences with minimum free
energy structure $\alpha$ (a “neutral set” of $\alpha$), and check for the presence of $\alpha$ in the plastic reperatoires of sequences in the set of one-mutant neighbors of the neutral set. We search and calculate statistics over only the one-mutant neighbors that are compatible (see section 3.2) with $\alpha$, since only these can have $\alpha$ as a suboptimal in the first place. (Given a sequence $S$, the fraction of one-error mutants compatible with $S$ is $1 + (n_{GU} - 5n_{bp})/3n$, where $n$ is the sequence length, $n_{bp}$ is the number of base pairs in $\alpha$ and $n_{GU}$ is the number of $GU$ pairings that would occur when $S$ folds into $\alpha$.) While sequences that are compatible with $\alpha$ have, by definition, $\alpha$ among their suboptimals at some energy, we are only interested in the limited plasticity range of $0 \leq \Delta \leq 10kT$ from the minimum free energy, since structures outside this range have Boltzmann probabilities too low to influence fitness. Figure 14 shows the fraction of sequences that have $\alpha$, the tRNA cloverleaf, among their suboptimal configurations and that are located in the compatible one-error boundary of a tRNA neutral set. This fraction is shown as a function of the energy interval $\Delta$. The fraction of sequences in the sample that have at least one one-mutant neighbor with the tRNA as a suboptimal (curve $i$) hits 1 at only $0.6kT$, while the fraction of all one-error mutants (curve $ii$) is above 0.8 at $5kT$, and reaches 1 at $9.6kT$ (for $T = 37^\circ C$).

The first numerical observations demonstrate that the plastically accessible configurations of a sequence indicate the immediate phenotypic vicinity of $\alpha$’s neutral network. The second method shows that $\alpha$’s neutral network casts a “shadow” into the energetically low lying suboptimal configurations realized by its one-error compatible boundary. Together these facts suggest that plastic accessibility and genetic accessibility mirror each other in our RNA model. From a biophysical point of view, this matches intuition. It means that the thermodynamic stability of a suboptimal structure $\alpha$ that is realized with non-negligible probability can frequently be improved (even to the point of making $\alpha$ the ground state) by one suitably placed point mutation. This occurs “positively” through mutations in the base pairing positions of $\alpha$ to nucleotides that yield better stacking energies; and through similar modifications to mismatches at the termini of helical regions (for example, the mutation from $A$ to $A'$ in Figure 4). It also occurs “negatively” through mutations that cause incompatibility between the sequence and the ground state or other low lying configurations. Intuitively, the mutational stabilization of structures toward which a sequence is already predisposed occurs more readily than the construction of a minimum free energy structure.
Plastogenetic congruence in conjunction with the cost of plasticity are the reasons plasticity can be decreased through genetic evolution. Our first two pieces of evidence argue generally for the correlation between plastic and genetic accessibility. Figures 15 and 16 provide a third perspective which emphasizes this evolutionarily enabling function of plastogenetic congruence. We generate a sample of sequences with a given $\alpha$ as a $5kT$-suboptimal configuration with no constraints on the minimum free energy structure. The likelihood of obtaining sequences with a predefined $\alpha$ among their $5kT$ repertoires can be tuned by enriching those sequence segments that should fold into the stacking regions of $\alpha$ with GC pairs. These contribute the largest stacking energy and, therefore, induce the segments to fold in the desired way (see inset of Figure 15).

Every sample sequence is now made the starting point of a gradient walk. At each step of the walk, we choose the one-error mutant that most increases the Boltzmann probability $P(\alpha)$ of $\alpha$. Figure 15 shows the distribution of the walk lengths until $\alpha$ becomes the ground state and until $P(\alpha)$ cannot further be improved. The graphs on the left of Figure 15 confirm that there is a large probability of making $\alpha$ the ground state in a single point mutation. Furthermore, this one-step probability is lower for samples that permit $\alpha$ to lie higher in the energy interval $\Delta$. The difference is one order of magnitude as $\Delta$ increases from $5kT$ to $8kT$. A higher GC content in stacking regions also increases the one-step probability, since any bias toward $\alpha$ helps its evolution to lower free energy.

Two to three mutations are a remarkably few to convert a suboptimal $\alpha$ into the minimum free energy structure. Equally remarkable is the high average number of steps it takes to reach a local maximum of $P(\alpha)$.

Figure 16 shows the distributions of $P(\alpha)$ when $\alpha$ has become the minimum free energy structure for the first time and when $P(\alpha)$ could not be further improved. Two aspects are worth noting. First, the degree to which the thermodynamic sharpness of tRNA structures can be improved is large. Second, the final $P(\alpha)$-distribution is much narrower and the average $P(\alpha)$ attained is substantially closer to one for a tRNA structure (I) than for another structure (II) that we chose randomly. For this random structure only 68% of the gradient walks found a sequence that had that shape as the minimum free energy structure within 100 mutations. This suggests that certain structures
Figure 15: **Plastogenetic congruence III.** The inset shows the frequency with which a random sequence compatible with a tRNA cloverleaf (structure I) has that structure among its $5kT$ suboptimal configurations (curve i). This frequency is shown as a function of the fraction of GC or CG pairs in the sequence segments that fold into the stacking regions of the cloverleaf shape. Curve ii is the frequency with which such sequences have structure I as a minimum free energy structure. The main graph shows the distribution of gradient walk lengths as described in the text. There are two sets of curves, black (left hump) and grey (right hump). The left set pertains to walk lengths up to the first appearance of the given structure (I or II) as a minimum free energy structure. The right set is the distribution of walk lengths until the Boltzmann probability of that structure could not be further improved. Only walks that terminated within 100 steps are considered. Circles pertain to structure II, $\Delta = 5kT$, GC fraction in helical regions is 0.5, 1000 distinct walks were performed of which 686 terminated. Squares pertain to structure I, $\Delta = 5kT$, GC fraction is 0.5, 1000 walks of which 999 terminated. Triangles up pertain to structure I, $\Delta = 5kT$, GC fraction is 0.33 (i.e., no bias), 916 walks of which 909 terminated. Triangles down pertain to structure I, $\Delta = 8kT$, GC fraction is 0.33 (i.e., no bias), 322 walks of which 322 terminated.
are more canalizable, i.e. susceptible to reduction of plasticity than others, and, therefore, that there are intrinsic limits to canalization.

Figure 16: **Plastogenetic congruence III.** The frequency distributions of Boltzmann probabilities for structures I and II are shown. See Figure 15 for details. The left set of curves show the probabilities when the structures first appear as the ground state, and the right set shows the probabilities at local fitness optima of the gradient walk.

In sum, this third perspective on plastogenetic congruence looks beyond the one-error neighborhood. If a conversion of a suboptimal shape into a minimum free energy structure cannot be achieved in a single step, it can occur over several steps in a gradual fashion. Moreover, once a structure has become the ground state, there is still room for a significant reduction of plasticity.

Plastogenetic congruence occurs because plasticity and mutational robustness have the same genetic basis. The top part of Figure 17 shows four examples of gradient walks, chosen for their diversity of approaches to a local optimum. The walk criterion optimizes the Boltzmann probability of the structure depicted at the bottom of Figure 17. It is the structure that domi-
nated the population when the evolutionary process of Figure 8 (section 3.2) became trapped. The bottom part of Figure 17 shows the concurrent changes in neutrality along the same gradient walks. First a structure descends to the minimum free energy configuration, and the neutrality drops to a minimum. Then the neutrality increases dramatically to levels above 0.4, which are in the tail of the distribution of neutralities for all sequences with that structure as the minimum free energy configuration (not shown). The average of this neutrality distribution is 0.3 (Matt Bell, personal communication). In one case, the sequences attain, but are unable to sustain, an extreme neutrality of approximately 0.5 (up-triangles). Note that the fraction of one-error neighbors compatible with the shown structure is approximately 0.55. (The exact compatibility fraction depends on the U content of a sequence, see above.) The neutralities associated with the maximum Boltzmann probability are therefore close to the theoretically achievable maximum.

We conclude that the level of neutrality (genetic canalization) increases with increasing Boltzmann probability of the minimum free energy structure (environmental canalization). After the structure becomes the minimum free energy structure, all point mutations that increase its Boltzmann probability are neutral. Figure 17 illustrates that these neutral mutations steer the sequences towards even higher neutrality (up to an average ceiling). This is a profound epistatic effect, of the kind we define in section 2.5 and Figure 6 (genetic control of neutrality). It also suggests that the predominant mechanism for environmental canalization in RNA is neutral mutations that knock out alternative structures in the plastic repertoire.

Genetic canalization means high phenotypic robustness towards mutations. In our model, environmental canalization - the reduction of plasticity - is clearly an adaptation, since it is a direct response to selection in a population. Genetic canalization, however, is not an adaptation. It is instead a byproduct of environmental canalization, to which it is linked by common genetic underpinnings. The possibility that environmental and genetic canalization share a genetic basis which might account for the evolution of canalization was recently hypothesized by Wagner et al. (1997). Our study of the evolution of plasticity in RNA provides a mechanistic realization for this hypothesis.

Figures 18 and 19 provide anecdotal support for the corresponding drops in plasticity and evolvability. The first of these figures illustrates the surprising
Figure 17: **Gradient walks in plastogenetic space.** Gradient walks are generated by moving from the current sequence to its one-error mutant that most improves the Boltzmann probability of a prespecified structure. Walks begin with sequences that have that structure within their $5kT$ plastic repertoire. Top: Progress profiles along sample walks. Bottom: The concurrent development of neutrality (fraction of neutral one-error mutants) along the same walks.
extent of plasticity reduction in RNA. It compares the density of structural states of three sequences that have been obtained through different processes but share the same minimum free energy structure. The common minimum free energy shape is the characteristic shape $\alpha$ found at the dead-end of the evolutionary process depicted in Figure 8 (section 3).

The random sequence obtained by inverse folding $\alpha$ (Hofacker et al., 1994) has 574 different structures within 3 kcal of $\alpha$. It spends only 3% of the time in $\alpha$, and the combined probability of the 574 alternative configurations accounts for only 58% of the total partition function. The different configurations cover a wide range of structural diversity.

The second sequence is the product of evolution under the simple map. We inoculated a simulated flow reactor with the aforementioned inverse folded sequence, and specified its minimum free energy structure $\alpha$ as the target structure. This leaves no room for phenotypic improvement. Despite the absence of direct selection pressure on the well-definition of the ground state, we observe a reduction of plasticity by one order of magnitude. In van Nimwegen et al. (1999), we learn that a population evolving on a neutral network under the simple map will concentrate on sequences with higher than average neutrality. (See Figure 19 below and Figure 24 in Section 5.2.) Because mutations off the neutral network yield, on average, much lower fitness phenotypes, there is indirect selection against sequences in the neutral network that have a high fraction of one-error mutants off the network, i.e. low neutrality. This is a second order effect that depends on the probability of deleterious mutation. Looking at the suite of low free energy configurations of these highly neutral sequences, we find corresponding low plasticity. This is similar to the phenomenon of “epigenetic stability” found by A. Wagner (1996). In our case it is caused by plastogenetic congruence in the absence of direct selection on plasticity. We see this indirect effect in the neutrally evolved sequence of Figure 18, where the time spent in the ground state $\alpha$ increases to 34%, although there remains considerable structural diversity in the plastic repertoire.

The last sequence in Figure 18 is evolved under the plastic map where reduced plasticity is the direct consequence of natural selection. The size of the plastic repertoire decreases by another order of magnitude to only 4 structures in addition to the ground state which is obtained 67% of the time. If the molecule equilibrates over all its states, it spends 94% of its time in these 5
configurations that are highly similar to each other.

Figure 19 shows the neutral positions of these sequences. We call a position neutral if at least one mutation at that position leaves the minimum free energy structure unchanged. Figure 19 additionally includes three sequences obtained similarly for the structural end-point in the evolutionary process of Figure 7, section 3. For the structures on the left, the neutrality (fraction of mutations that leave the minimum free energy structure intact) of the sequences increases from 0.184 for the random sequence to 0.412 for the neutrally evolved sequence to 0.456 for the canalized sequences. Similarly on the left, the neutrality increases from 0.158 to 0.311 to 0.430 from top to bottom. The intermediate neutrality of the neutrally-evolved sequences reflects the second-order selection for increased neutrality under the simple map as discussed above. The further increase in neutrality that occurs under the plastic map is the flip-side of a dramatic reduction in plasticity due to plastogenetic congruence.

Notice that the neutral coverage in the canalized case is almost perfect, in the sense that almost every position permitting a mutation that does not destroy the compatibility of the sequence with its minimum free energy structure is a neutral position. For example, in a GC pair only a mutation in the position occupied by the C can be compatible with the original structure and hence neutral. Replacing it by a U preserves the ability to pair with G, while there is no comparable mutation for the position occupied by the G.

4.2 Neutral confinement

Recall that plasticity is costly because more structures in the plastic repertoire implies less time spent in any one. Further, this cost leads to an evolutionary reduction of plasticity (section 3.2). Genetic assimilation - the movement of an advantageous suboptimal structure to the minimum free energy structure - and the further reduction of plasticity are facilitated by plastogenetic congruence. Yet this connection between plastic accessibility and genetic accessibility also produces a detrimental byproduct of the reduction in plasticity: a loss of variability.

Figure 20 compares the frequency distribution of neutralities at the end of the plastic simulations depicted in Figures 7 and 8 (filled circles) and the corresponding control runs (down-triangles). (The data represented by squares
Figure 18: **Density of states in $5kT$.** The structural density of states is shown for three sequences that have been obtained by inverse folding, neutral evolution under the simple map and canalization under the plastic map. All three sequences have the same ground state. For the first two sequences, we present a few sample structures with their energy on the right hand scale. We display the complete repertoire of the third sequence. The grey boxes indicate many non-displayed structural states.
Figure 19: **Neutral positions.** The emphasized positions are those with at least one neutral mutation. We highlight these neutral positions for inverse folded, neutrally evolved, and canalized sequences on two minimum free energy structures.
will be discussed below.) Recall that the neutrality of a sequence, the fraction
of its one-error mutants that share its minimum free energy structure, mea-
sures the lack of variability of the sequence. The plastic population evolves to
a distribution of neutralities with a much higher average and lower kurtosis
than that of the non-plastic control population.

Highly neutral sequences have only a small proportion of distinct phenotypes
in their mutational vicinities, and are therefore unlikely to mutate towards
a higher fitness phenotype. Furthermore, we argue that high neutrality not
only suggests a small number of non-neutral neighbors, but that the minimum
free energy structures of the few non-neutral neighbors offer little phenotypic
diversity, making the discovery of novel advantageous mutants even more un-
probable. In Wuchty et al. (1999), high neutrality is shown to correlate with
high structural similarity between the ground state structure and the other
configurations in the plastic repertoire. Again, by plastogenetic congruence,
this implies that a one-error mutant of a low-plasticity sequence either folds
into the same minimum free energy structure as the low-plasticity sequence
(neutrality) or into a structure that is very similar to it. Very similar here
means that the structures differ slightly with respect to stack lengths or loop
sizes, but have otherwise the same relative arrangement of stacks and loops.
In other words, they are topologically equivalent. As a consequence, the dis-
covery of advantageous shapes is considerably slowed down and eventually
halted. We say the population is “neutrally confined”.

Highly neutral regions of a neutral network appear to be highly extended
and connected (Matt Bell, personal communication). Genetic diversification,
therefore, still occurs within these regions. Yet the extent of genetic variation
is insufficient to produce discontinuous shape transformations such as a shift.
A shift from structure $\alpha$ to structure $\beta$ requires a sequence that can fold into
both structures (recall the discussion in section 3.2 and around Figure 9 on
discontinuous transitions), such that the next point mutation can alter the
ground state from $\alpha$ to $\beta$. Such a sequence is by definition more plastic
than a sequence highly defined on either structure alone. It is precisely this
necessary increase in plasticity that is prevented by neutral confinement.

Neutral confinement in RNA seems independent of the mutation rate. We sim-
ulate evolution on a neutral network by starting a population with sequences
that have the designated target structure as their minimum free energy struc-
ture. Figure 21 depicts such simulations that use the structure shown at the
Figure 20: **Neutral confinement.** The neutrality of population supports is the fraction of one-error mutants of sequence species present in a population. In other words, every sequence type is weighted the same, irrespective of its frequency in the population. Frequency-weighted plots look similar, with more dramatic high-neutrality peaks, and subdued the low-neutrality spectrum. These plots intentionally convey a more conservative picture. Top: Populations pertaining to the simulation of Figure 8. Filled circles: population support neutrality after $4.2 \times 10^6$ replications under the plastic map (the structure shown in the inset dominates). Down-triangles: neutrality at the end of the control run (Figure 8) with the simple map. Squares: neutrality of a population that has evolved for $5.6^6$ replications from the one underlying the filled-circle data when plasticity was switched off. Bottom: Similar analysis for populations pertaining to the simulation shown in Figure 7. Filled circles: with plasticity after $30 \times 10^6$ replications. Down-triangles: control run without plasticity after $32 \times 10^6$ replications. Squares: evolved from the filled-circle data for $4.4 \times 10^6$ replications with plasticity switched off.
Figure 21: **Neutrality as a function of replication accuracy.** Populations were obtained by neutral evolution with different replication accuracies on the structure $\alpha$ shown at the top of Figure 20. The graphs that decay exponentially with decreasing replication accuracy represent the stationary frequency of the target phenotype $\alpha$ in the population. The nearly constant graphs monitor the average neutrality (in the population support sense of Figure 20) of the sequences on the “master network” (i.e., whose minimum free energy configuration is the target $\alpha$). The replication accuracy at which the master network is lost is known as the phenotypic error threshold (Huynen et al., 1996; Reidys et al., 1998). For infinite populations the threshold accuracy $q_{\min}$ depends on the average neutrality of the $\alpha$-network, $\lambda_\alpha$, and the superiority of $\alpha$, $\sigma_\alpha$, indicating how much better the replication rate of $\alpha$, $r_\alpha$, is with respect to the remaining phenotypes in the population (Reidys et al., 1998; Schuster and Fontana, 1999):

$$q_{\min} = \frac{(1 - \lambda_\alpha \sigma_\alpha)/(1 - \lambda_\alpha \sigma_\alpha)^{1/n}}{1 - \lambda_\alpha}$$

with $\sigma_\alpha = r_\alpha / \sum_{\beta \neq \alpha} r_\beta x_\beta/(1 - x_\alpha)$, where $x_\beta$ is the fraction of sequences with ground state $\beta$. In the plastic case, $r_\alpha$ is not constant, since different sequences with $\alpha$ as a ground state will have different suboptimal configurations. We replace $r_\alpha$ by the average over the pertinent sequences in the population.

The graphs for the plastic case always dominate those of the simple case (higher neutrality, and hence lower threshold accuracy). Filled circles: neutral evolution at superiority 10. Down-triangles: neutral evolution at superiority 1.5. Solid curves are linear regressions and exponential fits. Note the near constancy of the population neutrality in the plastic case regardless of the superiority and error rate.
top of Figure 20 as both the starting minimum free energy shape and the target. Figure 21 monitors the fraction of sequences with that ground state structure as a function of the replication accuracy per position. This is done for two values of the superiority (see caption) and for both the plastic and the simple map. The phenotypic error threshold (Huynen et al., 1996; Reidys et al., 1998) is the replication accuracy at which that ground state structure is lost from the population. At the same time we monitor the average neutrality of sequences with that ground state structure. The independence of the average neutrality from error rate and superiority was predicted for the simple case by van Nimwegen et al. (1999). In the plastic case, however, the independence is rather unexpected. In section 5 we present formal models of evolution under the plastic map. They predict a shift in the equilibrium distribution from high neutrality (of neutral confinement) to lower neutrality at a mutation rate higher than the error threshold. This suggests that as the mutation rate increases, the population goes directly from neutral confinement into falling completely off the neutral network (error catastrophe).

We have demonstrated that plastogenetic congruence plays three critical roles. First it enables the genetic assimilation of suboptimal shapes into ground states. Second it facilitates the reduction of plasticity. This occurs under stabilizing selection following a transition, in the early periods of apparent stasis when the genetically accessible phenotypic novelty is either neutral or of lower fitness. Third, during this reduction, plastogenetic congruence yields the self-defeating advance towards sequences with low phenotypic variability.

4.3 Structural lock-in

In our model, neutral confinement is maintained by selection pressure to reduce plasticity. If we remove this pressure by reverting to the simple map, we expect to relax neutral confinement and proceed to target. Yet this does not occur. While neutral confinement is indeed lost, the evolutionary process remains stuck in the same phenotype $\alpha$ (not shown). To investigate this, we randomly generate sequences that have the minimum free energy structure $\alpha$. Recall that $\alpha$ is the structure found commonly in the population at the end of evolution under the plastic map. Because these random sequences have not been subjected to selection, they have significantly more diverse
plastic repertoires than the corresponding evolved sequences. We then evolve 
these random sequences under the simple map, and find that, even with the 
lower plasticity and hence relaxed neutrality, the population fails to make 
phenotypic progress towards the target. We conclude that the lock-in is due 
not just to high neutrality, but to the structure itself.

Figure 22: Structural lock-in. The target cloverleaf structure is shown in the 
middle. Top and bottom depict the succession of dominant structures in the runs 
of Figures 8 and 7, and show the sequential lock-in described in the text.

Figure 22 illustrates such a structural lock-in. The structure represented 
in the middle is the target cloverleaf. The top 5 (bottom 6) structures are 
the dominant phenotypes taken from the simulation of Figure 8 (Figure 7). 
During the evolutionary process, segments of a sequence evolve base pairings 
that match the pairing orientation (up- or downstream) of the corresponding 
segments in the target. In the evolved structure, these segments pair with 
one another. In the target, however, they pair with bases elsewhere.

This results from our fitness function based on the Hamming distance be-
tween a structure and the target. In this metric, the unit of comparison 
is a position. Consequently, a paired position is evaluated independently
from the position with which it is paired. What matters is only whether it is paired and how it is oriented - "(" or ")". For example, a stacking region "((((. . .))))" that differs by a shift from the corresponding stacking region in the target, ".((((. . .))))", would contribute positively to the fitness of the overall structure, because the pairing states of some positions in that region match the pairing states of their corresponding target positions.

Consider now a metric that is based on "base pair distance", where the unit of structural comparison is a base pair. A shifted stack would not contribute anything to this measure of fitness, since no base pair is aligned with the target structure. We consider base pair distance below.

![Figure 23: Neutral confinement without structural lock-in. Flow reactor evolution as in Figures 8 and 7, but with base pair distance rather than Hamming distance as the similarity measure between a structure and the target entering equation (2) and (4). Top curve: plastic case, bottom curve: control run with the simple map. The control population attains the target in the usual punctuated fashion, while the plastic population gets stuck in neutral confinement. Switching off plasticity during the confinement regime, reduces neutrality and the population proceeds to target.](image-url)
The motif shaded in shape 1 at the top of Figure 22 positively contributes to fitness despite the discrepancy with the target pairing combinations. Such misfolded motifs require discontinuous transitions (in this case a flip of the leftmost strand, see Figure transform) to come into perfect agreement with the target motif. These transitions are achieved by extensive neutral genetic drift in the simple case. In the plastic case, however, misfolded structural modules that convey a fitness advantage tend to get frozen by neutral confinement. Consequently, natural selection on the remaining parts of the sequence proceeds conditional on the pre-existing misfold. This produces another highly advantageous motif in structure 4 that is similarly frozen by neutral confinement. By this time, the structure has a misfold built on top of a misfold. The segments marked a, b, c and d (see also the structures shown in Figure 20) each convey high fitness since they match the pairing state of the corresponding target segments. Yet in the target, a pairs with d, not with b, and both b and c pair with other segments. This can only be undone by a suite of neutral mutations that poise segment a to pair with d in conjunction with another suite of neutral mutations to produce segments capable of pairing with b and c, which will be liberated when a becomes paired with d. A loss of the pairing states of b and c would entail too high a fitness cost. An escape from this phenotype requires three discontinuous transitions at once - a feat that is virtually impossible, even in the absence of neutral confinement. The ground state structure itself is evolutionarily locked-in, independent of the neutrality of the underlying sequence.

Structural lock-in is reminiscent of “generative entrenchment” (Wimsatt, 1986, 1999) in a developmental context. This is the constraint imposed on later events by consequences of earlier decisions (see also Buss (1987)). Neutral confinement (a consequence of the RNA genotype-phenotype map) is a necessary condition for the ensuing structural lock-in. The evolutionary trajectories of the control (simple map) simulations show that lock-in does not occur in the absence of neutral confinement, unless one starts with a locked-in structure. Neutral confinement, however, is not a sufficient condition, since lock-in also depends on the map from phenotypes to fitness. This is clearly illustrated by Figure 23 where the fitness of a sequence (equation 2) is based on the base pair distance mentioned previously. In this case, the class of discontinuous shape transformations is given by the nucleation of whole stacking regions (see Figure 3 and Fontana and Schuster (1998b) for details). Evolution under the plastic map becomes trapped, as with the orig-
inal Hamming distance. This time, however, switching off plasticity undoes neutral confinement and enables the process to proceed towards the target. Under this structural metric, then, neutral confinement alone explains the evolutionary dead-end. We conclude that neutral confinement and structural lock-in are distinct but non-independent phenomena that lead to a profound loss of evolvability under the original plastic map.
5 ANALYTIC MODELS OF NEUTRAL CONFINEMENT

5.1 Model Assumptions

In this section, we offer three simplified models of plastic RNA evolution. In each, we characterize the dynamics of a plastic population in the final stage of evolution, and derive equilibrium conditions in terms of neutrality and fitness. The two assumptions that make these approaches analytically tractable are: (1) perfect plastogenetic congruence, and (2) containment within a single neutral network. By perfect plastogenetic congruence, we mean that structures accessible to a sequence through plasticity are exactly those found as minimum free energy structures of sequences accessible through single mutations. Although simulation populations typically consist of sequences from several equally fit neutral networks, our containment assumption simplifies the population to sequences that share a single minimum free energy structure, although their suboptimal structures may vary.

5.2 Frequency distribution within a neutral network

Consider a population of RNA sequences with phenotypic plasticity. Each sequence can assume a range of structures, all within an energetic neighborhood of its minimum free energy structure. The fitness of a sequence is determined by a weighted average of the distances between its low energy configurations and a target structure.

The assumption that plasticity gives a perfect picture of evolutionarily adjacent structures, allows us to construe the fitness of a sequence by

\[
    w(S_i) = \sum_{S_j \in \mathcal{N}(S_i)} f(d(\alpha_{0j}, \tau)) \cdot \frac{\delta_\mu(S_j, S_i)}{\sum_{S_j \in \mathcal{N}(S_i)} \delta_\mu(S_j, S_i)}
\]

where \( S_i \) is a sequence, \( \alpha_{0j} \) is the minimum free energy structure of \( S_i \), \( \mathcal{N}(S_i) \) is the mutational neighborhood of \( S_i \), \( d(\cdot, \cdot) \) determines structural distance between two shapes, \( f(\cdot) \) is a monotonically decreasing function of structural distance \( d(\cdot) \), and \( \delta_\mu(\cdot) \) is a monotonically decreasing function of mutational
distance (indexed by the mutation rate). This definition is to be seen in analogy to equation (4). Because of perfect plastogenetic congruence, the structures that determine the fitness of \( S_i \) are the minimum free energy structures in its mutational neighborhood. The second factor in 8 plays a role analogous to the Boltzmann probability by weighing a neighboring structure with the likelihood of reaching its sequence by mutation.

Suppose a population is concentrated on a neutral network \( G \) of relatively high fitness. That is, all sequences in \( G \) have the same minimum free energy structure, and most one-error mutants of these sequences that lie outside of \( G \) have relatively much lower fitness. Let \( |G| \) be the number of sequences in \( G \). We think of \( G \) in graph theoretic terms. Each sequence is analogous to a node in a graph, and two nodes are connected by an edge when the sequences they represent differ by exactly one mutation. Assuming perfect plasticity, we can approximate the fitness of an individual node \( S_i \) by \( f d_i \) where \( f \) is the selective value of the minimum free energy structure shared by all sequences in \( G \), and \( d_i \) is the degree of the node within the network. In other words, for a sequence \( S_i \), \( d_i \) is the number of one-error mutants that have the same minimum free energy structure as \( S_i \), and therefore lie within the same neutral network.

We can express the per generation change in the frequency distribution of sequences in \( G \) with a system of \( |G| \) equations. For any \( S \in G \),

\[
P'_s = \frac{f}{\bar{w}} (P_s d_s (1 - \mu) + \frac{\mu}{3n} \sum_{t \in N(s)} P_t d_t)
\]  

where \( P_s \) gives the frequency of sequence \( S \), \( \mu \) gives the per sequence per generation mutation rate, \( n \) is the length of the sequences and \( \bar{w} \) is the average fitness of the population. Note that this formulation ignores the possibility of mutations onto \( G \) from sequences outside of \( G \).

We now translate this system of equations into a transition matrix \( M \) such that \( \vec{P}' = M \vec{P} \) where \( \vec{P} \) is the frequency distribution vector. Let \( I \) denote the identity matrix; \( A \) denote the adjacency matrix where \( A_{ij} = 1 \) if \( S_i \) and \( S_j \) are one-error mutants of each other and \( A_{ij} = 0 \) otherwise; and \( D \) denote the diagonal matrix of degrees with \( D_{ii} = d_i \) for all \( i \) and \( D_{ij} = 0 \) for all \( i \neq j \). Then we derive,

\[
M = ( (1 - \mu) I + \frac{\mu}{3n} A ) D. 
\]
Assume there exists a unique node $c$ such that $d_c > d_i$ for all $i \neq c$. Without loss of generality, we let $c = 0$. Then $S_0$ is the most neutral sequence in $G$ with degree $d_0$. When $\mu = 0$, that is, there is no mutation, then $M = D$ which has a leading eigenvalue of $d_0$ and an associated eigenvector $[1, 0, 0, \ldots, 0]$. In the absence of mutation then, the equilibrium population is made up entirely of $S_0$.

In general, the equilibrium distribution is the solution $\vec{P}$ to the eigenvalue equation $M\vec{P} = \lambda \vec{P}$ where $\lambda$ is the leading eigenvalue of $M$. The $n$th eigenvalue and associated eigenvector of $M$ are denoted $\lambda^{(n)}$ and $\vec{P}^{(n)}$, respectively. For much of the following analysis we will treat only the leading eigenvalue and eigenvector $\lambda^{(0)}$ and $\vec{P}^{(0)}$, and in that case, we omit the superscript $(0)$.

Perturbation theory enables a first-order approximation of the effects of mutation on the leading eigenvector. First we rewrite the matrix in (10), and approximate $\vec{P}$ and $\lambda$ for small $\mu$ by standard first order perturbation theory:

$$M = D + \mu \left( \frac{1}{3n} A - I \right) D = [M_0] + \mu [M_1],$$

with

$$\lambda = \lambda_0 + \mu \lambda_1, \text{ and } \vec{P} = \vec{P}_0 + \mu \vec{P}_1.$$  \hspace{1cm} (12)

The eigenvalue equation becomes $(M_0 + \mu M_1)(\vec{P}_0 + \mu \vec{P}_1) = (\lambda_0 + \mu \lambda_1)(\vec{P}_0 + \mu \vec{P}_1)$. We now seek solutions for $\lambda_1$ and $\vec{P}_1$. This expression breaks down into two equations.

$$M_0 \vec{P}_0 = \lambda_0 \vec{P}_0$$ \hspace{1cm} (13)

and

$$M_0 \vec{P}_1 + M_1 \vec{P}_0 = \lambda_1 \vec{P}_0 + \lambda_1 \vec{P}_0.$$ \hspace{1cm} (14)

Equation (13) is just $D \vec{P}_0 = \lambda_0 \vec{P}_0$, which, as discussed above, has the leading eigenvalue solution $\lambda_0 = d_0$ and $\vec{P}_0 = [1, 0, 0, \ldots, 0]$. Rearranging equation (14), and multiplying both sides by the transpose $\vec{P}_0^T$, we find

$$\vec{P}_0^T (M_0 - I \lambda_0) \vec{P}_1 = -\vec{P}_0^T (M_1 - I \lambda_1) \vec{P}_0 \Rightarrow 0 = -\vec{P}_0^T (M_1 - I \lambda_1) \vec{P}_0,$$ \hspace{1cm} (15)
since \( \vec{P}_0^T M_0 \vec{P}_0 = M_0 \vec{P}_0 = \lambda_0 \vec{P}_0 = \vec{P}_0^T \lambda_0 \). This yields the eigenvalue solution 
\[ \lambda_1 = \frac{\vec{P}_0^T M_1 \vec{P}_0}{\vec{P}_0^T \vec{P}_0} = (M_1)_{0,0} = -d_0 \text{ since } \vec{P}_0^T \vec{P}_0 = 1. \]

Now let \( \vec{P}_1 = \sum_{i=1}^{[G]} a_i P_0^{(i)} \), where the \( a_i \) are scalar coefficients and \( P_0^{(i)} \) denotes the \( i \)th eigenvalue of the matrix \( M_0 \), which is the \( i \)th unit vector. This means that the \( i \)th entry of \( \vec{P}_1 \) is just \( a_i \). Substitution of this summation in equation (14) yields
\[
M_0 \sum_{i=1}^{[G]} a_i P_0^{(i)} + M_1 \vec{P}_0^{(0)} = \lambda_0 \sum_{i=1}^{[G]} a_i P_0^{(i)} + \lambda_1 \vec{P}_0^{(0)} \tag{16}
\]

\[ \Rightarrow (M_1 - \lambda_1 \vec{P}_0^{(0)}) \vec{P}_0^{(0)} = \sum_{i=1}^{[G]} a_i (\lambda_0 - \lambda_1^{(i)}) \vec{P}_0^{(i)}. \tag{17} \]

Multiplying both sides by \( \vec{P}_0^{(k)T} \), we find
\[
\vec{P}_0^{(k)T} (M_1 - \lambda_1 \vec{P}_0^{(0)}) \vec{P}_0^{(0)} = a_k (\lambda_0 - \lambda_1^{(k)}) \tag{18}
\]

Equation (18) simplifies to expressions for all \( a_k \), when \( k > 0 \).

\[
a_k = \frac{\vec{P}_0^{(k)T} (M_1 - \lambda_1 \vec{P}_0^{(0)}) \vec{P}_0^{(0)}}{(\lambda_0 - \lambda_1^{(k)})} = (M_1 - \lambda_1 \vec{P}_0^{(0)})_{k,0} = A_{0,k} \frac{d_0}{3n(d_0 - d_k)} \tag{19}
\]

Recall that \( A_{0,k} = 1 \) if \( S_k \) is a one-error mutant of \( S_0 \) and \( A_{0,k} = 0 \) otherwise. We normalize to solve for \( a_0 \). Since \( \sum_{i=1}^{[G]} a_i = 1 \), then \( a_0 = 1 - \frac{d_0}{3n} \sum_{k=1}^{[G]} \frac{A_{0,k}}{d_0 - d_k} \). From this we derive the complete first-order approximation of the equilibrium distribution.

\[
\hat{P}_k = \begin{cases} \frac{1}{1+\mu} \left( 1 + \mu \left( 1 - \frac{d_0}{3n} \sum_{k=1}^{[G]} \frac{A_{0,k}}{d_0 - d_k} \right) \right) & \text{if } k = 0, \\ \frac{\mu}{1+\mu} \left( \frac{d_0}{3n(d_0 - d_k)} \right) & \text{if } A_{0,k} = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{20}
\]

According to this approximation, the equilibrium distribution is made up of the most connected sequence \( S_0 \) and its one-error mutants. As \( \mu \) decreases, the proportion of non-\( S_0 \) sequences shrinks. Because the denominator \( 3n(d_0 - d_k) \) depends on \( d_k \), the frequency of a non-\( S_0 \) sequence is an increasing function of its degree. Even sequences with degree close to \( d_0 \) will
appear at low frequency. For example, consider an $S_k$ where $d_k = d_0 - 1$ and the maximal $\mu = 0.1$. Then $P_k = 0.03\frac{d_0}{n}$ which is very small for large $n$.

This model extends a graph theoretic conception of a non-plastic RNA neutral network (van Nimwegen et al., 1999). For two simple networks, we compare the equilibrium distribution of neutrality predicted by our plastic model to that of the original non-plastic model. Both networks contain 20 nodes (sequences). In the first, node 0 is connected to all 19 other nodes, and all other nodes are only connected to 0. In the second, nodes 0-10 are connected to all other nodes while nodes 11-20 are only connected to nodes 0-10. Figure 24 graphs the equilibrium distributions for non-plastic (dotted) and plastic (solid) populations. The plastic populations are much more concentrated on the nodes with highest degree.

![Network distributions](image)

Figure 24: **Network distributions.** The equilibrium distributions for a non-plastic population (dotted) and a plastic population (solid). In the network case (a), left ordinate, node 1 has much higher neutrality than all other nodes: $d_1 = 19$ and $d_n = 1$ for $n = 2 \ldots 20$. In case (b), right ordinate, nodes 1-10 form a high neutrality subnetwork: $d_n = 10$ for $n = 1 \ldots 10$ and $d_n = 1$ for $n = 11 \ldots 20$.

In these simple examples, and in general, plastogenetic congruence forces the
population into an extremely neutral region of the network, preventing access to phenotypes off of the neutral network. This model thereby demonstrates formally what we previously argued verbally: the reduction of plasticity leads to a loss of evolvability by way of neutral confinement.

5.3 Three-Tiered Model: Exploration and Error Thresholds

Again consider a population that has reached a relatively high fitness neutral network $G$. We break down the population into three distinct classes, two within the neutral network and one representing the rest of sequence space. For any sequence $S_i \in G$, the one-error mutants of $S_i$ that are also in $G$ are called its neutral neighbors, and the number of such neighbors is called $d_i$, the degree of $S_i$. Assume that the neutral network contains a single sequence $S_0$ with maximal connectivity $d_0$ within $G$, and the degree of all other sequences in $G$ is $d_1$. That is, for any $S_i \in G$ where $S_i \neq S_0$, $d_i = d_1 < d_0$.

We define three classes of sequences: $C_0$ is just the sequence $S_0$, $C_1$ consists of the sequences in $G - \{S_0\}$, and $C_2$ are all sequences not contained in $G$. Movement of the population among these classes results from a combination of mutation and natural selection.

Let $\mu$ be the probability of mutation per sequence per generation. Sequences in $C_0$ mutate to sequences in $C_1$ at a rate $\mu \frac{d_0}{3n}$, where $3n$ is the number of possible one-error mutants of a sequence. Mutations take sequences in $C_0$ off of the network, into $C_2$ at a rate $\mu \left(1 - \frac{d_0}{3n}\right)$. Likewise the mutation rates from $C_1$ to itself and to $C_2$ are $\mu \frac{d_1}{3n}$ and $\mu \left(1 - \frac{d_1}{3n}\right)$, respectively. We ignore back mutation of sequences in $C_1$ to $C_0$ and of sequences in $C_2$ to either $C_0$ or $C_1$.

The sequences in this model are endowed with perfect plastogenetic congruence as described above. We model this with a fitness function that considers the minimum free energy structures of all one-error mutants of a given sequence. Every sequence has a selective value which is a measure of the similarity between the sequence’s minimum free energy structure and a pre-determined target structure. Sequences within $G$ have a selective value of 1 while sequences off $G$ have a relative selective value of $f$. The relative fitness of a sequence in class $C_i$ is $w_i$, an average of the selective values of its
one-error mutants:

\[ w_0 = \frac{d_0}{3n} + \left( 1 - \frac{d_0}{3n} \right) f \]  
(21)

\[ w_1 = \frac{d_1}{3n} + \left( 1 - \frac{d_1}{3n} \right) f \]  
(22)

\[ w_2 = f \]  
(23)

Recall that \( \frac{d_i}{3n} (i = 1, 2) \) is, for any sequence in \( C_i \), the fraction of its one-error mutants that also lie within \( G \). We construct a transition matrix \( T \) that incorporates the effects of mutation and selection. If \( \vec{P} = (P_1, P_2, P_3) \) describes the occupancy of the three classes at some time, then \( \vec{P}' = T\vec{P} \) gives the frequency distribution in the next generation where \( T \) is as follows:

\[
T = \begin{pmatrix}
  w_0(1 - \mu) & 0 & 0 \\
  w_0\mu\frac{d_0}{3n} & w_1(1 - \mu) + w_1\mu\frac{d_1}{3n} & 0 \\
  w_0\mu(1 - \frac{d_0}{3n}) & w_1\mu(1 - \frac{d_1}{3n}) & w_2
\end{pmatrix}
\]  
(24)

The leading eigenvector of \( T \) provides the population equilibrium distribution. The eigensystem of \( T \) is given by

\[
\begin{pmatrix}
  \lambda_0 \\
  \lambda_1 \\
  \lambda_2
\end{pmatrix} = \begin{pmatrix}
  w_0(1 - \mu) \\
  w_1(1 - \mu \left( 1 - \frac{d_1}{3n} \right)) \\
  w_2
\end{pmatrix}
\]  
(25)

and

\[
\vec{P}_0 = \left( \frac{w_0(1 - \mu) - w_1(1 - \mu \left( 1 - \frac{d_1}{3n} \right))}{w_0\mu\frac{d_0}{3n}}, 1, \frac{w_1(1 - \mu \left( 1 - \frac{d_1}{3n} \right)) - w_0(1 - \frac{d_0}{3n} - \mu(1 - \frac{d_1}{3n}))}{-w_0\mu(1 - \mu) + w_2} \right)
\]  
(26)

\[
\vec{P}_1 = \left( 0, \frac{w_1(1 - \mu \left( 1 - \frac{d_1}{3n} \right)) - w_2}{w_1\mu(1 - \frac{d_2}{3n})}, 1 \right)
\]  
(27)

\[
\vec{P}_2 = (0, 0, 1)
\]  
(28)

The subscripts of the eigenvalues and eigenvectors refer to the class which dominates the distribution, and not to their magnitudes. As parameter values vary, so does the leading eigenvalue \( \lambda = \max(\lambda_0, \lambda_1, \lambda_2) \).

We seek parameter ranges that allow the population to explore phenotype space. A population concentrated in \( C_0 \) will have mostly neutral mutants,
and therefore will be unlikely to find a higher fitness phenotype through mutation. We call this an exploration catastrophe. A population lost in $C_2$ has retrogressed from the higher fitness network. This is the error catastrophe. $C_1$ on the other hand is a high fitness platform from which a population can explore phenotype space for better options. For these reasons, we say that a population with a high concentration of $C_1$ is below the error threshold which moves the population off of $G$, and above the exploration threshold which contains the population in a highly inward-looking subset of $G$.

In Figure 25, we display equilibrium distributions for various parameter ranges. In every case, we set the length of the sequences to $n = 100$. Each right-hand graph shows the concentration of $C_0$ labelled (0), $C_1$ labelled (1), and $C_2$ labelled (2). Populations dominated by $C_2$ are above the error threshold while populations dominated by $C_0$ are below the exploration threshold. On the left we graph the eigenvalues. The topmost plane at each point is the leading eigenvalue. Its associated eigenvector is that which determined the frequencies in the opposing graph. Light grey, medium grey and dark grey represent $\lambda_0, \lambda_1$ and $\lambda_2$ respectively.

The center distribution assumes neutralities of $\frac{d_0}{3n} = 0.45$ for $C_0$ and $\frac{d_1}{3n} = 0.3$ for $C_1$. These are the values we obtain from actual simulation. The equilibrium distribution suggests that for the low values of $\mu$ we use in simulations, the population should be trapped on $C_0$ in an exploration catastrophe. This is consistent with the neutral confinement we find for low-plasticity sequences.

### 5.4 Neutrality in the sub-network $C_0$

In simulation, RNA populations evolve through neutral networks that are much more complex than the idealized $C_0$ and $C_1$ of the three-tiered model. Equilibrium populations move among highly inter-connected sub-networks of multiple structurally similar neutral networks. In this extension of the three-tiered model, we still assume the population converges on a single neutral network $G$, but attempt a more realistic conception of the structure of $G$.

Suppose now that $C_0$ is enlarged to a sub-network of the neutral network $G$ in which all $|C_0|$ sequences have degree $d_0$. As above, the remaining sequences in $G$ occupy $C_1$ and share degree $d_1$. We define two new parameters $\chi_0$ and $\chi_1$ which are the fraction of neutral mutations that remain within $C_0$ and $C_1$ respectively. Then $1 - \chi_0$ is the fraction of neutral mutations from $C_0$ to $C_1$, 

...
Figure 25: Eigenvalues (left graphs) and equilibrium distributions (right graphs). Given values for $\mu$ and $f$, we graph $\lambda_0$, $\lambda_1$ and $\lambda_2$ in light, medium and dark grey, respectively, and equilibrium concentrations for $C_0$, $C_1$ and $C_2$ labelled as (0), (1), and (2), respectively.
and $1 - \chi_1$ is the fraction from $C_1$ to $C_0$.

Mutation operates according to the following table of mutation rates:

<table>
<thead>
<tr>
<th>Origin</th>
<th>$C_0$</th>
<th>$C_1$</th>
<th>$C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_0$</td>
<td>$\chi_0 \mu \frac{d_0}{3n}$</td>
<td>$(1 - \chi_0) \mu \frac{d_0}{3n}$</td>
<td>$\mu \left(1 - \frac{d_0}{3n}\right)$</td>
</tr>
<tr>
<td>$C_1$</td>
<td>$(1 - \chi_1) \mu \frac{d_1}{3n}$</td>
<td>$\chi_1 \mu \frac{d_1}{3n}$</td>
<td>$\mu \left(1 - \frac{d_1}{3n}\right)$</td>
</tr>
<tr>
<td>$C_2$</td>
<td>0</td>
<td>0</td>
<td>$\mu$</td>
</tr>
</tbody>
</table>

Again we ignore mutations from $C_2$ into $G$.

As in the original formulation, fitness is the weighted average of the selective values of the minimum free energy structures of all one-error mutants of a sequence. For any sequence in $C_0$ or $C_1$ the total number of one-error mutants that are also in $G$ is still $d_0$ or $d_1$ respectively. The extension of $C_0$ to a sub-network of $G$ does not therefore alter the fitnesses of the three classes.

Again we construct a transition matrix that describes flow between classes. In the following matrix $\mu_{ij}$, $0 \leq i, j \leq 2$, is the mutation rate from $C_i$ to $C_j$ as given in the mutation table above.

$$T' = \begin{pmatrix}
  w_0(1 - \mu) + w_0\mu_{00} & w_1\mu_{10} & 0 \\
  w_0\mu_{01} & w_1(1 - \mu) + w_1\mu_{11} & 0 \\
  w_0\mu_{02} & w_1\mu_{12} & w_2
\end{pmatrix} \quad (29)$$

The eigensystem of $T'$ is more complicated than that of $T$, because $T'$ is not triangular, and we have added the new parameters $\chi_0$ and $\chi_1$. In the following analysis, we assume $\chi_0 = \chi_1 = \chi$, and explore the affects of $\chi$ and $\mu$ on the predicted equilibrium distributions.

Figure 26 depicts equilibrium distributions for $f = 0.9, 0.5, 0.1$ from left to right. As $\chi$ increases, the equilibrium concentration of $C_0$ increases slightly and the frequency of $C_1$ decreases, making exploration even more difficult. The lower eigenvalue graphs reveal that under the extended model, populations are always in either the $\lambda_0$ phase or the $\lambda_2$ phase dominated by $C_0$ and $C_2$ respectively. If we define the error and exploration thresholds by these phase transitions, then this suggests that the error and exploration thresholds are identical. In other words, the population goes directly from confinement in $C_0$ to diffusion in $C_2$. 
5.5 Exploration and Error Catastrophes

In both versions of the model we find parameter ranges for which the population is confined to $C_0$ and therefore is unlikely to find phenotypic novelty through mutation (exploration catastrophe). For given values of $f$, $n$ and $\chi$ we can identify a threshold $\mu$ below which the population is in an exploration catastrophe. Generally, the exploration threshold decreases as $f$ increases, and is a slightly monotonically increasing function of $\chi$. This implies that as the size and interconnectivity of $C_0$ grows, so does the likelihood of reaching an evolutionary dead-end.

![Graphs showing equilibrium distributions and eigenvalues.](image)

**Figure 26:** Extended model equilibrium distributions and eigenvalues. Given values for $\mu$ and $\chi$, these display equilibrium concentrations for $C_0$, $C_1$ and $C_2$ labelled as (0), (1) and (2), respectively. The bottom row gives eigenvalues $\lambda_0$ in light grey, $\lambda_1$ in medium grey and $\lambda_2$ in dark grey. All graphs assume $L = 100$, $d_0 = 135$, and $d_1 = 95$. From left to right, $f = 0.9, 0.5, 0.1$. Note that the range for $\mu$ varies across graphs.

There also exist parameter ranges for which the population is lost in $C_2$. For
specified $f$, $n$ and $\chi$ we can find a threshold $\mu$ above which the population reaches such an error catastrophe. Intuitively, as $f$ – the relative fitness of $C_2$ – increases, the error threshold decreases.

We now contextualize these transitions within two established theoretical frameworks: the standard genotypic error threshold (SGET), denoting the mutation rate at which a genotype with optimal phenotype (“master genotype”) is lost (Eigen, 1971), and the phenotypic error threshold (PET), denoting the mutation rate at which the optimal phenotype is no longer maintained in the population (Huynen et al., 1996; Reidys et al., 1998).

In the simple version of our model, the exploration threshold corresponds to the SGET in which there is a single master sequence with superiority approximately $w_0 = \frac{d_0(1-f)+3nf}{3n}$. The SGET divides the low mutation rates at which the master class is preserved at equilibrium from the high mutation rates at which the master class is catastrophically lost. The exploration threshold similarly divides the low mutation rates at which the population is confined to $C_0$ from the high mutation rates at which the population reaches $C_1$. The theories differ only semantically in that the genotypic error catastrophe occurs below the SGET whereas the exploration catastrophe occurs at mutation rates above its threshold. These thresholds increase as superiority increases. In our case then, decreasing $d_1$ postpones the error threshold, and thereby increases the range of mutation rates for which a population will be locked into $C_0$.

The error catastrophe in our simple model is a PEC where the mean fraction of neutral neighbors is approximately $\nu \sim \frac{d_1}{3n}$. Reidys et al. (1998) showed that the PET threshold increases with $\nu$. In other words, the more sequences there are in the master class, the harder it is to lose the master class through mutation. This bears out in our model: as we increase $d_1$, we postpone the transition to an error catastrophe.

The exploration catastrophe takes place where the leading eigenvalue shifts from dark grey ($\lambda_0$) to medium grey ($\lambda_1$) in Figure 25, and the error catastrophe takes place where a medium grey ($\lambda_1$) leading eigenvalue shifts to light grey ($\lambda_2$). For some values of $f$, $n$, $d_0$ and $d_1$, we see phase transitions directly from $\lambda_0$ to $\lambda_2$. This corresponds to parameter values at which the error catastrophe precedes the exploration catastrophe. In other words, the minimum mutation rate at which the population can escape $C_0$ is already greater than the threshold above which the population is lost from $C_1$ into
In our extended model both the exploration catastrophe and the error catastrophe correspond to a PET because $C_0$ becomes a neutral network rather than a single sequence. Increasing the size of $C_0$, i.e. increasing the neutrality of $C_0$, postpones the exploration threshold. A larger $C_0$ entails extended ranges of mutation rates at which the population will be in an exploration catastrophe. Consequently the regions of parameter space in which the population goes directly from an exploration catastrophe into an error catastrophe also increase.

5.6 Finite population birth-death model

Following the example of Nowak and Schuster (1989) we use a stochastic birth-death model to examine the implications of finite population size on the exploration threshold. Consider a population already evolved to a relatively high fitness network in which all sequences share a minimum free energy structure. As in the graph theoretic approximation, we assume the neutral network is made up of a single highly connected node and a space of other equally but less connected sequences.

![Mutation Trajectories](image_url)
In this approximation, we let the relative fitness off the neutral network be $f = 0$. According to the fitness function described in the three-tiered model then, the sequence comprising $C_0$ has relative fitness $w_0 = d_0$ and the sequences in $C_1$ have relative fitness $w_1 = d_1$.

The diagram in Figure 27 illustrates mutation between classes. We have added back mutation from $C_1$ to $C_0$. Sequences in $C_0$ have connectivity $d_0$ and therefore mutate to $C_1$ at a rate $\mu \frac{d_0}{3n}$ and to $C_2$ at a rate $\mu \left(1 - \frac{d_0}{3n}\right)$.

There are $d_0$ sequences in $C_1$ that are one-error mutants of the sequence that makes up $C_0$. Each one of these sequences mutates into $C_1$ with a probability $\mu \frac{1}{3n}$. Assuming a uniform distribution of sequences across $C_1$, the total back mutation rate from $C_1$ to $C_0$ is $\mu \frac{d_0}{|C_1|3n}$ where $|C_1|$ is the number of sequences in $C_1$. The remaining neutral mutations $\mu \left(\frac{d_1}{3n} - \frac{d_0}{|C_1|3n}\right)$ take sequences in $C_1$ to other sequences in $C_1$.

We construct a model that maintains constant total population size $N$. Each time an individual sequence is lost through mutation to $C_2$, another sequence replicates with or without mutation to replace it. The matrix

$$
\mathbf{F} = \begin{pmatrix} w_0(1 - \mu) & w_1 \mu \left(\frac{d_0}{|C_1|3n}\right) & 0 \\
\frac{w_0 \mu d_0}{3n} & w_1 \left((-\mu) + \mu \left(\frac{d_1}{3n} - \frac{d_0}{|C_1|3n}\right)\right) & 0 \\
\frac{w_0 \mu (1 - d_0)}{3n} & w_1 \mu \left(1 - \frac{d_1}{3n}\right) & 0
\end{pmatrix}
$$

(30)

describes the movement between classes as a function of mutation and fitnesses. The rate of transition from class $j$ to class $i$ is given by $F_{ij}$.

The following equations describe changes in occupancy of the highly connected peak:

$$
P_{i,i+1} = \frac{\frac{N-i}{N} F_{2,1}}{\frac{i}{N} F_{2,0} + \frac{N-i}{N} F_{2,1}} \left(\frac{i}{N} F_{0,0} + \frac{N-i}{N} F_{0,1}\right)
$$

(31)

$$
P_{i,i-1} = \frac{\frac{i}{N} F_{2,0}}{\frac{i}{N} F_{2,0} + \frac{N-i}{N} F_{2,1}} \left(\frac{i}{N} F_{1,0} + \frac{N-i}{N} F_{1,1}\right).
$$

(32)

Both equations are made up of a death term and a birth term. For example, in the case of $P_{i,i+1}$, the death term gives us the likelihood that a individual chosen randomly to die will be from $C_1$. This is just the probability that a sequence in $C_1$ is lost to $C_2$ over the probability that any sequence from $C_0$ or $C_1$ is lost from the neutral network. The birth term is the rate at which
new members of $C_0$ are formed, and is a combination of accurate replication in $C_0$ and mutation from $C_1$. $P_{i,i-1}$ conversely combines the probability of a death in $C_0$ with the probability that a sequence in $C_1$ will arise to replace it.

Tools from stochastic processes allow us to calculate the probability distribution of the $C_0$ concentration at equilibrium (Nowak and Schuster, 1989; Karlin and Taylor, 1975). Let $\vec{Q} = (q_0, q_1, \ldots, q_N)$ be the vector in which $q_i$ is the probability that there are $i$ individuals in $C_0$ at equilibrium. Using the equations above we find that for all $k \in \{1, 2, \ldots, N\}$,

$$q_k = \frac{\pi_k}{\sum_{i=0}^{N} \pi_i}$$

where $\pi_k = \frac{P_{k-1,k}}{P_{k,k-1}} \pi_{k-1}$ and $\pi_0 = 1$. \hspace{1cm} (33)

Figure 28 graphs the equilibrium probability distributions for a range of mutation rates and $d_1$. For sufficiently large values of $d_1$, we see that there exists a threshold mutation rate above which the probability distribution becomes a monotonically decreasing function of the size of $C_0$. In Figure 28(C) for example, when $d_0 = 200$ and $d_1 = 170$, a qualitative shift in the distribution occurs around $\mu = 0.2$.

We can characterize this transition using the function

$$\Phi : i \rightarrow P_{i-1,i} - P_{i,i-1}$$ \hspace{1cm} (34)

This describes the net flow into $C_0$. When $\Phi$ is positive $C_0$ maintains a strong hold on the population distribution, whereas negative $\Phi$ implies skew towards $C_1$. As Figure 29 suggests, this function has two real roots in the interval $[0, 1]$ until the mutational threshold, at which point $\Phi(i)$ dives below 0 for all $i \in [0, 1]$. We call this transition the “conservative exploration threshold” since above it, the population is securely poised in $C_1$ to search the rest of phenotype space. Figure 30, which depicts the threshold for a range of $d_1$, was derived from equation 34. As the differences in $C_0$ and $C_1$ connectivities increase, the likelihood of being stuck in $C_0$ also increases. This effect is strongest for large populations.

These observations for a finite population corroborate the analysis of the three-tiered model. The lower the neutrality of $C_1$ relative to that of $C_0$, the harder it is to escape $C_0$. The barriers to exploration rise with population size.
Figure 28: Stationary state probability distribution. For all plots, $N = 100$, $d_0 = 200$, $n = 100$ and $C = 500$. Each figure depicts $q_x$ for $\mu \in \{0.05, 0.1, 0.25, 0.2, 0.25\}$, where the value of $\mu$ increases for each curve from left to right. (A) $d_1 = 150$ (B) $d_1 = 160$ (C) $d_1 = 170$
Figure 29: **Comparison of** $\Phi(x)$, labelled (i), and $q_x$, labelled (ii). **For all plots,** $N = 100$, $d_0 = 200$, $d_1 = 175$, $n = 100$ and $C = 500$. (A) $\mu = 0.05$ (B) $\mu = 0.15$ (C) $\mu = 0.25$.

Figure 30: **Conservative Exploration Threshold.** For all plots, $d_0 = 200$, $n = 100$ and $C = 500$. Population size $N$ varies as indicated.
6 MODULARITY

In section 3, we demonstrated that plasticity is rapidly reduced by natural selection under a biophysically motivated fitness function that weights the performance of each shape in the repertoire of a sequence by the probability that the sequence assumes that shape. The reduction of plasticity has, in addition to genetic canalization, a further side-effect which we call modularity. Since it is not \textit{a priori} obvious how to precisely define modularity, we suggest several quantitative perspectives on RNA sequences as evolvable manifestations of modularity.

6.1 Norms of Reaction - Melting Behavior

Throughout this paper, plasticity is a stochastic choice among alternative structural states of a biopolymer in contact with a heat bath at constant temperature ("microenvironmental fluctuations", see section 2.4). Biological plasticity is commonly characterized by a norm of reaction, a set of phenotypes triggered by the variation of some (macro)environmental parameter. The obvious parameter in our case is temperature, and the corresponding norm of reaction for RNA is known in biophysics as "melting". In Figures 31 and 32 we compare the (computed) melting behavior of the RNA sequences from Figure 19 at three levels.

First we compute the melting series, that is, the suite of minimum free energy structures in the temperature range from 0°C to 100°C. Melting behavior can also be measured through a more coarse grained statistic, the heat capacity (at constant pressure), labeled $H$ in Figures 31 and 32. We derive $H$ from the Gibbs free energy of the ensemble of structural states, $G$, via the partition function $Z$, using McCaskill's algorithm (McCaskill, 1990):

\[
H = -T \frac{\partial^2 G}{\partial T^2} \quad \text{with} \quad G = -RT \log Z, \tag{35}
\]

where $R$ the universal gas constant. This function can be measured empirically by differential scanning calorimetry (DSC). DSC is widely-used to determine the thermophysical properties of materials. As a sample is heated over a range of temperature, the material starts to undergo a phase change that releases or absorbs heat. The calorimeter measures the heat flow (enthalpy
change) into or out of the sample undergoing the phase change, thereby providing data from which the heat capacity can be quantitatively recovered. Our third perspective on structural transitions is the set of Boltzmann probabilities for each structure in the melting series as a function of temperature, as calculated by the suboptimal folding algorithm (Wuchty et al., 1999). Notably, the peaks in $H$ correspond to major structural phase transitions. The peak is typically at half-completion of the transition where the Boltzmann probabilities of the “outgoing” and “incoming” phases intersect.

In Figure 31, we compare the melting behavior of the three sequences on the left of Figure 19. The melting behavior of the inverse folded sequence is markedly more disorderly than that of the neutrally evolved and canalized sequences. Its melting series consists of as many as 10 shapes, and phase transitions often involve all parts of the structure at once. This is particularly the case at low temperatures, when there is a lot of structure to rearrange. The temperature intervals of each minimum free energy structure are fairly small. As a consequence, the graph of heat capacity consists of several small and closely-occurring humps. Notice that the $37^\circ C$ structure (#3 on the upper left of Figure 31) becomes unstable when the temperature is lowered, indicating that the molecule is already in the process of melting and just happens to have the desired structure at $37^\circ C$.

The canalized sequence, in contrast, has an almost digital melting behavior. Individual structural features undergo independently sharp melting transitions at distinct temperatures. This behavior identifies structural “traits”. In this case, such traits coincide with components that are not enclosed by any base pair (see Figure 1). The ground state structure is stable up to $61^\circ C$, when the large T-like feature at the 5′ end disappears almost entirely in a single step. Remarkably, the absence of that trait does not affect the structural integrity of the other trait at the 3′ end. This trait melts in a single step at $89^\circ C$. The major transitions from #1 to #2 and from #3 to #4 are well separated and marked by two sharp peaks in the heat capacity. (The spike in the Boltzmann probability of the ground state around $37^\circ C$ reflects the canalization of the sequence to this structure during evolution at this temperature.)

The neutrally evolved sequence occupies a middle ground. It has fewer transitions and more highly preserved structural similarity across transitions than the inverse folded sequence, but to a lesser extent than the canalized se-
Figure 31: **Melting behavior I.** Below each graph are the minimum free energy structures in the temperature range 0°C – 100°C for three sequences obtained by inverse folding, neutral evolution and canalization on the same structure (Figure 19 left). The graphs trace the temperature dependence of the specific heat (H) and the probabilities of the individual structures in the melting series.
Figure 32: Melting Behavior II. See caption to Figure 31. The 37°C ground state structure for these sequences is the one pictured on the right of Figure 19.
quence. The major departure from the canalized sequence is the lack of discrete trait transitions. For example, the transition from #1 to #2, involving the hairpin structure at the 3’ end, neither melts the feature completely nor preserves its components. The same holds for the trait at the 5’ end.

This comparative analysis of the melting behavior exploits one aspect of “modularity”: the thermodynamic independence of a structural trait from other traits over a wide temperature range.

Figure 33: Context insensitivity of Modules. The sequence segments underlying the 5’ structural component (A) and the 3’ component (B) of the shapes on the left and right of Figure 19, respectively, are embedded in random sequence contexts. The chart shows the frequency with which the segment retained its original structure if it originated in an inverse folded, neutrally evolved and canalized sequence.

We refine this characterization of modularity in light of Figure 32. It depicts the melting behavior of the sequences at the right of Figure 19. The claims we make for Figure 31 apply to this case as well. The melting behavior of the canalized sequence, however, is not as extreme as in the first example.
Neither structural trait disappears at once in a single event. Rather, both traits melt in an interleaving fashion. We note, however, that their paired substructures never rearrange in the process. They melt at distinct temperatures leaving other substructures unaffected. Stated differently, the melting process remains completely discrete at the level of components of traits, not necessarily at the level of entire traits. This simply means that there is modularity within modules, prompted by the syntactically hierarchical nature of RNA secondary structure, where larger structures are iteratively built from smaller paired segments.

6.2 Context Insensitivity

The temperature dependence of RNA structure reveals a thermodynamic integrity of components in canalized sequences that is absent from random sequences and less sharply established in neutrally evolved sequences. Thermodynamic integrity means that shape components are not disrupted as the dissolution of other components makes segments of the sequence available for base-pair interaction. This insensitivity to changing context within a fixed sequence suggests that such components should also exhibit increased insensitivity to contextual changes at the genetic level.

To test this hypothesis, we slice from each sequence the segment $s_\tau$ that folds into a particular structural trait $\tau$. We then flank $s_\tau$ on the left and right with 1000 random segments half its size, and count the instances in which $s_\tau$ still folds into $\tau$. The chart in Figure 33 confirms the higher robustness of structural components of canalized sequences relative to inverse folded sequences. In a few cases, the neutrally evolved sequences withstand contextual modification as well as the canalized ones.

Here, like in the case of plastogenetic congruence, phenotypic variation under environmental perturbation seems to mirror phenotypic variation under genetic change. Specifically, thermodynamic autonomy (or robustness) of traits, as manifest in the norms of reaction to temperature, implies autonomy of those same traits with respect to changing genetic contexts ("thermo-genetic congruence"). Computations of such context-independence suggest that it is significant for naturally occurring RNA sequences with functionally important structures (Wagner and Stadler, 1999).

Along a similar vein, canalization also entails localization of mutational ef-
fects. For example, Figure 5 of section 2.4 compared the base pairing entropies of the inverse folded to canalized sequences on the left half of Figure 19. The base pairs in the canalized sequence are extremely well-defined.

### 6.3 Kinetics - Modularity and Funnels

A comparison of folding kinetics of our three classes of sequences provides a final perspective on modularity. In forthcoming work (Flamm et al., 1999) we present a stochastic model of kinetic RNA folding at elementary step resolution that enables us to study folding pathways (or mechanisms). That is, we statistically predict the temporal successions of structure forming events. In RNA, folding pathways play a role analogous to developmental pathways in organisms. The lower right graph of Figure 34 shows the distribution of folding times for the three sequences on the left of Figure 19. The tree graphs represent the energetic barriers that exist between the 50 lowest free energy structures accessible to each of the three sequences (courtesy Christoph Flamm, University of Vienna). The differences are striking. Not only does the canalized sequence fold much more rapidly than the others, but it folds predominantly along one well-defined pathway, and has, therefore, one dominating time scale. (There is a folding trap [not shown], visited with low probability, that accounts for the tail of the distribution. It is due to an early configurational diffusion among many high energy shapes until the molecule drops into the funnel and folds.) In contrast, the distribution of folding times for the inverse folded and the neutrally evolved sequences indicate multiple, equally dominant mechanisms (and time constants), due to the presence of folding traps or the absence of any guiding or funnel structure (Bryngelson and Wolynes, 1987; Dill and Chan, 1997) of the energy surface.

The trees in Figure 34 provide a representation of which shapes, corresponding to local energy minima, are connected into basins and how these basins are linked with one another. The leaves are the 50 lowest states (local minima) and the internal nodes are the barrier heights (minimum of all maxima along paths) connecting two states or their basins (Flamm et al., 1999). In other words, the height of the lowest internal node connecting two leaves represents the energetic requirement of folding from one of the two structures into the other. (Stated in terms of a flooding metaphor, if the energy abscissa were to measure altitude and the landscape was flooded up to a given height,
Figure 34: Folding Kinetics and Energy Landscapes. The graph on the lower right shows the folding time distributions for the three classes of sequences shown on the left of Figure 19, based on a stochastic model of kinetic folding whose elementary moves consist in the making, breaking and shifting of a single base pair (Flamm et al., 1999). (1), (2) and (3) are the inverse folded, neutrally evolved and canalized sequences respectively. The trees depict the energy landscape associated with each sequence in terms of the hierarchical organization of barriers separating individual states and their basins. The distance from the root (top) of the tree represents the free energy.
a horizontal cut through the graph at that height reveals which states are joined under water (on the right), and which basins were to merge next as the water level is raised.)

The inverse folded sequence has a low energy landscape without much structure, as well as high barriers separating the individual states. The sequence can misfold in many ways, delaying the formation of the native minimum free energy structure. The consequence is a multitude of time constants that show up in the folding time distribution. The neutrally evolved sequence has a far more organized energy surface consisting of two major basins and, consequently, two major (that is, high probability) folding pathways: a direct one, if the early events end up in the right basin containing the minimum free energy structure, and an indirect one, when the sequence has to unfold after first misfolding into the wrong basin. The energy surface of the canalized sequence reflects funnel organization, first conceptualized in the context of protein folding (Bryngelson and Wolynes, 1987; Dill and Chan, 1997). Similar observations hold for other sequences, which are not displayed here.

These examples provide anecdotal evidence for the evolution of modularity. A complete statistics of modularity is outside the scope of this paper. This section shows that modularity of traits, understood intuitively as independence or autonomy of traits, manifests itself in a variety of modes - thermodynamic, genetic and kinetic. Recall, however, that sequences were never selected for modularity, only for the reduction of plasticity. It is easy to imagine direct selection pressures for modularity, but this analysis demonstrates that the emergence of modularity does not require them. Modularity arises, like genetic canalization, as a necessary by-product of environmental canalization.
7 DISCUSSION AND OUTLOOK

Biological evolution changes not only the frequencies of extant phenotypes, but the phenotypes themselves. A population genetic analysis of the fate of innovations under natural selection provides only a partial story that must be integrated with a theory of the attainability of phenotypic innovations (Buss, 1987; Wagner and Altenberg, 1996). What matters is not just the set of possible phenotypes in principle, but rather the topology of the possible. That topology conveys information about how (and whether) a particular phenotype can be accessed from another. For a given class of phenotypes, this organization of phenotype space is induced by the genotype-phenotype map, and critically shapes evolutionary histories independently of fitness criteria. To unravel that topology requires a formal perspective on biological objects. The choice of “phenotype” is clearly of strategic importance.

We have sidestepped the most severe complexity by turning our attention to the simplest (non-trivial) evolvable object: an RNA molecule. RNA provides both a theoretically and empirically well characterized high-dimensional relation between genotype (sequence) and phenotype (structure). The principle algorithms for RNA folding were developed twenty years ago (Waterman and Smith, 1978; Nussinov et al., 1978; Zuker and Stiegler, 1981), not with the present questions in mind, but rather as a tool to assist experimentalists. This divergence of applications suggests that inscription errors, (that is, the construction of models which predetermine the desired output) are less likely here than for genotype-phenotype models constructed with particular definitions of epistasis, plasticity or modularity in mind. Although RNA folding is not a completely formalized relation between sequences and structures, it is sufficiently abstract that insights into its statistical architecture (Schuster et al., 1994; Fontana and Schuster, 1998a) can inform the design of fully abstract axiomatic models of genotype-phenotype maps (Reidys et al., 1997; Cupal et al., 1999). We develop such an example in section 5.

Recent advances in RNA folding (McCaskill, 1990; Wuchty et al., 1999; Flamm et al., 1999) enable us to introduce and analyze a form of environment-gene interaction which we call plasticity. The result is a powerful model system in which concepts like plasticity, evolvability, epistasis and modularity not only can be precisely defined and statistically measured, but reveal simultaneous and profoundly non-independent effects of natural selection.
Although these concepts were introduced in the study of organismal evolution, we demonstrate that they apply to the molecular domain as well. RNA provides a simple microfoundation.

The secondary structure into which an RNA sequence folds is determined not by the primary sequence alone, but also by environmental inputs such as temperature and the presence of other potentially interacting molecules. As a surrogate for such environmental factors, we map a sequence to a repertoire of its thermodynamically most stable structures. We assume that the Boltzmann coefficient - a variable reflecting the thermodynamic stability of a structure relative to all other structures within the configuration space of a sequence - is proportional to the time that sequence would spend in the given structure under a heterogeneous environment.

The most striking outcome of our simulations is the dramatic loss of evolvability that accompanies the evolutionary reduction of diversity in the plastic repertoires. Recall that evolvability is the potential of a population of sequences to innovate phenotypically. We gain a deeper understanding of the loss of evolvability through two lines of inquiry. First we construct the causal bridge from the assumptions of our model - point mutation, a plastic genotype-phenotype map, and fitness based on the average structural distance to target over all structures in a plastic repertoire - to the evolutionary dead-end. Second, we characterize in as many dimensions as possible the typical genotype and phenotype distribution for a steady state population evolved under the plastic map. These two objectives are highly interrelated. The link between plasticity and evolvability is enabled by a close look at the evolving distribution of sequences and their shape repertoires.

The loss of evolvability stems from two simple observations:

1. The more variation in the plastic repertoire, the less time a sequence spends in its best structure. In this way, plasticity is costly, and is thereby reduced by natural selection.

2. There is a significant overlap between the shapes in the plastic repertoire of a sequence and the set of minimum free energy structures of genetically proximate sequences, i.e. of sequences that differ from it by one mutation. We call this property plastogenetic congruence.

The first observation is a straightforward consequence of our (biophysically
motivated) plastic fitness function. We verify the second, which rests on the intuition that a point mutation can tip the the folding landscape of a sequence in favor of any low-energy structure, through several statistical assays. Phenocopies - environmentally triggered traits that correspond to mutant phenotypes - provide evidence for the generality of plastogenetic congruence in nature. The effects of high temperature on moth antenna morphology (Goldschmidt, 1940), of ether on *Drosophila melanogaster* thoracic development (Gibson and Hogness, 1996; Waddington, 1942) and of gold foil on the fibular crest in birds (Müller, 1990) among many other environmental perturbations have been shown to mimick known mutants or ancestral morphologies (Stearns, 1993; Schlichting and Pigliucci, 1998). Because the alignment between environmental variability (plasticity) and genetic variability has met the skepticism of evolutionary geneticists, it is not well-integrated into mainstream evolutionary thinking. Our example of the plastic RNA map demonstrates its relevance to evolutionary theory, and instantiates the claim that plastogenetic congruence is a fairly ubiquitous property of genotype-phenotype relationships.

We then construct the causal bridge. Natural selection reduces plasticity. Plastogenetic congruence implies that a drop in diversity in the plastic repertoire entails a drop in the diversity of minimum free energy structures in the one-error neighborhood (the set of one-error mutants of a given sequence). Natural selection on plasticity indirectly curtails phenotypic novelty accessible by mutation, and hence the potential to evolve.

Through idealized models of a plastic population, we formalize this explanation for the decline in evolvability. These models become analytically tractable under the assumption that the plastogenetic congruence is perfect. In other words, the plastic repertoire for any given sequence is exactly the set of minimum free energy structures in the one-error neighborhood. The neutrality of a sequence is the fraction of its one-error neighborhood with identical minimum free energy structures. As neutrality increases, the ability to evolve decreases. A neutral network is a set of sequences that share a minimum free energy structure and are connected via mutation. In agreement with our simulations, the models predict that a plastic population evolving towards a fixed target will have much higher average neutrality at equilibrium than a comparable non-plastic control population.

Further, we formally connect mutation rate, the topography of phenotype
space and evolvability. Assuming that the population has reached a neutral network relatively close to the target shape, we identify three phases of equilibrium distributions: the exploration catastrophe, when the population is concentrated in a highly neutral region and so cannot access phenotypic novelty; the error catastrophe, when the population falls off the neutral network into the rest of genotype space with on average much lower fitness; and the ideal phase, when the population remains in regions of the neutral network that have mutational access to the rest of phenotype space. The fate of the population as mutation increases depends on the structure of the neutral network. For some, increasing mutation rate takes the population from the exploration catastrophe through the ideal phase to the error catastrophe. For many neutral networks, however, the exploration threshold exceeds the error threshold, so upon increasing mutation rate, a population goes immediately from an exploration catastrophe to an error catastrophe. Simulation suggests that this is the predicament of our steady state RNA populations under the plastic map.

While the loss of plasticity and plastogenetic congruence are sufficient to explain the evolutionary lock-in, they are not the whole story. The particular fitness function we use allows for the evolution of an additional layer of phenotypic rigidity which results in structures that are locked-in independently of how plastic the sequence is that folds into them. We describe this effect of structural lock-in in section 4, and demonstrate that populations evolving under fitness measures lacking this property still suffer a loss of evolvability through plastogenetic congruence.

Plastogenetic congruence is a robust statistical feature of the RNA folding map from sequences to secondary structures with broad population genetics implications. In particular, it instantiates the hypothesis put forward by Wagner et al. (1997) that genetic canalization – buffering against phenotypic effects of mutation – occurs as a byproduct of environmental canalization – the evolution of resilience to environmental perturbation.

In meeting the second objective, a multi-dimensional characterization of the phenotypic consequences of natural selection, we examined the thermophysical and kinetic properties of the phenotypes generated under plastic evolution. By thermophysical properties we mean the suite of minimum free energy structures as a function of temperature, or other quantitative measure such as the heat capacity. This is the RNA version of a norm of reaction. By kinetic
properties we mean features pertinent to the process of folding (as opposed to the mere map on which our model is based), such as time scales (rate constants) and folding pathways through the configuration/energy landscape of an individual sequence. This is the RNA version of “developmental pathway”. Strikingly, all indicators converge on modularity as the distinguishing characteristic of evolutionarily locked-in phenotypes (that is, of phenotypes of sequences that have lost plasticity).

By modules we do not mean a syntactical property of structures (that is trivial in RNA), but rather autonomous components that maintain their structural integrity across a broad range of environmental and genetic contexts, and that lose integrity through sudden and discrete steps. Modularity manifests itself as a resistance to sustained environmental or genetic perturbation, and the dissolution of modules translates into sharp and well separated phase transitions. In section 6, we compare thermophysical, genetic and kinetic aspects of modularity across sequences that were generated by different processes but share the same minimum free energy structure. The correlation between thermophysical and kinetic aspects of modularity is particularly striking. Modularity appears thermophysically as distinct melting temperatures of structural components that vanish upon melting, leaving an open chain segment, rather than a different structural arrangement. Modularity appears kinetically as a single primary folding funnel over the configurational landscape of a sequence. In other words, there is virtually no chance of misfolding because autonomous components fold independently of other components. This contrasts dramatically to the other two landscape prototypes of Figure 34 consisting of two or more major funnels (hence misfolds) in the neutrally evolved case, and no funnel structure whatsoever (hence a continuum of folding dead-locks) in the random case. The thermophysical and kinetic picture matches the genetic one, where modularity translates into a “cassette” principle, by which modular structural components have a markedly increased probability of maintaining their integrity if transplanted into different sequence contexts. This agrees with the principles of RNA architecture discovered through recent crystallizations of catalytic RNAs (e.g., Cech et al. (1994); for an overview see Westhof et al. (1996)).

Others have argued (Wagner et al., 1997) that modularity is unlikely to result from stabilizing selection. Our results suggest the contrary for RNA. Plastogenetic congruence has a significant effect during phases of evolution on a neutral network during which the population undergoes stabilizing selection
on the dominant phenotype. Our simulations show that modularity arises in conjunction with the loss of plasticity in such phases. Why? Because modularity is one mechanism (evidently favored by plastogenetic congruence) to achieve a selectively advantageous reduction of plasticity. For point mutations to lower plasticity, that is, well-define the minimum free energy structure, they must increase the thermodynamic stability of its structural components, perhaps simultaneously knocking out alternative configurations. This, in turn, localizes the consequences of further mutations by increasing the number of neutral positions (e.g., epistatic control of neutrality). This is reinforced by positive feed-back through plastogenetic congruence (any sequence that has the above properties makes it easier for mutations to enhance them further). These are precisely the effects needed to construct modularity at the component level. Increasing modularity is a pathway to reduced plasticity. The latter appears here as an adaptation to direct selection pressure, while the former arises as a mechanism to achieve that result.

Modularity both is a manifestation of evolutionary lock-in, and provides the basic tool for escaping it. The evolutionary stability of modules makes them, in conjunction with their context-insensitivity (transposability), natural building blocks for constructing novelty at a higher combinatorial level (Wagner and Altenberg, 1996). The shift towards a combinatorics of modular elements appears to be less of a convenient ad hoc innovation in evolutionary process than the only route left, once it arises as a byproduct of environmental canalization in a constant environment.

Finally we emphasize that our genotype-phenotype map was never subject to evolutionary modification. In nature, phenotype-phenotype relations are themselves evolving entities. The RNA folding map, however, does not evolve sensu strictu. This simply means that RNA, as discussed here, is not a useful framework to study the evolutionary construction of new object classes. Yet, the phenomena we describe evidently do not require such flexibility. The combinatorial architecture of the RNA folding map entails domains of phenotype-genotype space with sufficiently different properties to appear as if the map itself evolves.

The work described here has at least two broad continuations. In this paper, we hold the environment constant, and detail the consequences of an intuitive evolutionary loss of plasticity. The next question we ask is: what favors the maintenance of plasticity? Heterogeneous environments is a com-
mon response, and we can easily address this hypothesis within our model framework. Our characterization of modularity suggests, however, another hypothesis related to the maintenance of plasticity, one we hope to approach both theoretically and empirically. The autonomous structural components of modular sequences might limit interactions between molecules. Consider the hybridization into a joint secondary structure of two RNA sequences, $s_i$ and $s_j$. To form a joint phenotype, both sequences will have to give up parts of their intra-molecular structure to form inter-molecular base pairs. Suppose now that $s_i$ and $s_j$ are modular. In the light of our previous discussion, a structural component is modular because its integrity is impervious to other potentially interacting parts of the sequence. Further, what holds within a sequence may hold between sequences. This line of argument suggests that modular sequences should be resistant to forming (both thermodynamically and kinetically) joint secondary structures, and should instead remain frozen in their individual structures. Conversely, we claim that the need for interaction may then require parts of both sequences to be sufficiently plastic, and hence unmodular. This scenario could be classified under “varying environments”, where “environment” no longer means simply a physical parameter such as temperature, but includes other, possibly coevolving, RNA sequences. Work to substantiate these theories is in progress.

ACKNOWLEDGEMENTS

Thanks to our colleagues Ivo Hofacker, Peter Stadler and Stefan Wuchty for their contributions to the Vienna RNA package. We are grateful to Leo Buss, Marc Feldman, Christoph Flamm, Peter Godfrey-Smith, Ellen Goldberg, James Griesemer, Martijn Huynen, Erica Jen, Michael Lachmann, Laura Landweber, Mark Newman, Erik van Nimwegen, Peter Schuster and Andreas Wagner for valuable discussions, comments on the manuscript, and the patience to listen. Günter Wagner corrected several mistakes in a previous draft, and helped us avoid a few fallacies. Christoph Flamm kindly provided the folding kinetics of section 6. Matt Bell, 1999 REU summer intern at the Santa Fe Institute, shared valuable insight and back-up computations. The present work was supported by a Steinmetz Fellowship and a U.S. National Defense Science and Engineering Fellowship to L.A., by NIH grant GM28016 to M. W. Feldman, by the Keck Foundation, and by core grants to the Santa Fe Institute from the John D. and Catherine T. MacArthur Foundation, the
National Science Foundation, and the U.S. Department of Energy. W.F. and his research program are supported by Michael A. Grantham.

References


