Evolutionary Lock-in and the Origin of Modularity in RNA Structure

Lauren W. Ancel[*,†] **and Walter Fontana**[†]

[*]Department of Biology
Emory University
1510 Clifton Road
Atlanta, GA 30322 USA

ancel@charles.stanford.edu

[†]Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501 USA

walter@santafe.edu

Modularity is a hallmark of biological organization and a important source of evolutionary novelty[1–3]. Yet, the origin of modules remains a problem for evolutionary biology, even in the case of the most basic protein or RNA domains[4]. Biological modularity has been defined on many levels including genetic, morphological and developmental. The task of creating a general theory for the origins, ubiquity and function of modularity requires the synthesis of these perspectives.

The first step in understanding the causes and consequences of modularity is to define it. The second step is to formulate methods of detecting it. The third is to build models and design experiments to study its origins and evolutionary implications.

In this chapter we offer a rigorous definition of modularity in RNA. It is the partitioning of molecules into subunits that are simultaneously independent with respect to their thermodynamic environment, genetic context, and folding kinetics. On the variational level, the module consists of a stretch (or stretches) of contiguous ribonucleotides held together by a covalent backbone. On the functional level, the relevant interactions include the covalent bonds of adjacent ribonucleotides and hydrogen bonds between non-adjacent bases. These elements of secondary structure provide the scaffold for tertiary structure, which underlies the functionality of the molecule.

We furthermore offer a practical tool for identification of modules in this semi-empirical framework. The melting profile of a molecule is the set of minimum free energy shapes attained as the temperature increases from 0° to 100° C. Modules are exactly those subunits that dissolve discretely without perturbing the remaining structure as temperature increases. A modular molecule is one entirely made up of such subunits.

Finally we present a theory about the origins of modularity. Modularity may facilitate the evolution of more complex organisms through the combinatorics of modules. If we try to explain how modularity arises in the first place in terms of this evolutionary benefit, we run into a chicken-and-egg paradox. Modularity cannot produce a more sophisticated syntax of variation until it already exists.

Here we offer a more agnostic explanation for the evolution of modularity. Using a model of RNA folding, we show that modularity arises as a by-product when natural selection acts to reduce the plasticity of molecules by stabilizing their shapes. This is mediated by a statistical property of the RNA folding map: the more thermodynamically well-defined a shape, the more localized and less disruptive the effects of point mutations. We report two consequences of this relationship. First, under selection for thermodynamic stability, evolution grinds to a halt because of insufficient phenotypic variation. Second, the shapes trapped in this exploration catastrophe are highly modular. They consist of structural units that have become thermophysically, kinetically and genetically independent.

Under natural conditions, RNA molecules do not freeze in their minimum free energy shape (henceforth ground state), but exhibit a form of structural plasticity. Thermal fluctuations (corresponding to environmental noise) cause molecules to equilibrate among alternative low-energy shapes. We model the genotype-phenotype map from an RNA sequence to its repertoire of alternative secondary structures (henceforth

shapes) using an extension[5] of standard algorithms[6–8] which assist in the prediction of RNA secondary structure. For a given sequence, we compute all possible shapes having free energy within 3 kcal/mol of the ground state (at 37° C). We call this set of shapes the "plastic repertoire" of an RNA sequence (see Figure 1A for an illustration). The partition function[9] of a sequence is $Z = \sum_\alpha \exp(-\Delta G_\alpha / kT)$, where ΔG_α is the free energy of shape α and the sum runs over all possible shapes into which the sequence can fold. For any shape σ in the plastic repertoire of the sequence, the Boltzmann probability of σ, $p_\sigma = \exp(-\Delta G_\sigma / kT)/Z$, measures the relative stability of σ with respect to the entire repertoire. Assuming equilibration, p_σ is the amount of time the RNA molecule resides in shape σ. The ground shape is the most relevant shape, in the sense that the molecule spends the largest fraction of time in it.

Equipped with this computational model of RNA, we simulate an experimental protocol that evolves molecules to optimally bind a ligand[10]. We select sequences according to their similarity to a prespecified target shape[11]. In nature, the plasticity of an RNA sequence will presumably influence the overall binding constant of the molecule. At equilibrium, a fraction p_σ of a large number of identical sequences assumes shape σ and binds to ligand with the corresponding constant. In our model we proceed similarly by calculating for each shape σ in the plastic repertoire a selective value $f(\sigma)$ based on how well σ matches the target shape. The overall selective value, or fitness, r of the sequence is the average of the selective values of the shapes in its plastic repertoire each weighted by its occupancy time, $r = \sum_\sigma f(\sigma) p(\sigma)$. Point mutations provide the sole source of genetic variation in our simulations. This completes the model.

In order to identify the evolutionary implications of plasticity, we compare simulations of plastic RNA populations to simulations of non-plastic control populations. Sequences in the control populations rigidly fold into the ground state only. Consequently selection does not consider other low-energy shapes or the relative thermodynamic stability of the ground state.

High plasticity, that is, a large and diverse plastic repertoire, can be advantageous since multiple shapes, rather than just the ground state, contribute to the fitness of a sequence. In particular, a plastic sequence can partially offset a bad ground state with a good alternative shape in its repertoire. Yet plasticity is ultimately costly. The more shapes a molecule has in its plastic repertoire, the less time it spends in any one of them, including advantageous shapes. The evolutionary scenario under consideration – selection toward a constant target shape – must eventually favor the reduction of plasticity. The dynamics resemble a Simpson-Baldwin Effect in which organisms gain and then lose plasticity as they adapt to a novel environment.[12–15]. Figure 1A illustrates the plastic repertoire of a typical sequence present at the end of a plastic simulation. A comparison to a randomly chosen sequence folding into the same ground state reveals the staggering reduction in plasticity. The number of shapes in the plastic repertoire decreases 200-fold, the fraction of time spent in the ground state increases from 1.4% to 89%, and structural diversity within the 3 kcal band is nearly eliminated. Sequences found at the end of such simulations have well-defined ground

states that are extremely resilient to thermal fluctuations.

This reduction in plasticity produces a remarkable effect evident in the evolutionary trajectories of typical plastic and control populations (Figure 2A). Surprisingly, the population of plastic sequences evolves more slowly than the non-plastic control population, and quickly reaches an evolutionary dead-end. (Both evolve in a stepwise fashion with periods of phenotypic stasis punctuated by change toward the target shape [11].)

The loss of evolvability stems from a statistical correlation between the thermodynamic plasticity of an RNA molecule and the mutability of its ground state structure through point mutations[15]. In particular, thermodynamic robustness (or lack of plasticity) is positively correlated to mutational robustness. Mutations which stabilize the ground state of a molecule also serve to buffer the molecule against structural changes due to mutations. Figure 1B illustrates the mechanism underlying this correlation. A highly plastic molecule wiggles among multiple alternative shapes that are energetically close to one another. A point mutation can easily tip the energy landscape of the molecule in favor of an alternative shape (without destroying sequence compatibility with the original ground state). This occurs in the transition from (i) to (ii), where the alternative shape β becomes the new ground state. At the same time, a plastic molecule also offers opportunities to stabilize a ground state, as in the transition from (i) to (iii). Note, however, that along with this reinforcement comes the immunity of the ground state to a point mutation which affected it in the previous sequence context, as in (iii) to (iv) versus (i) to (ii).

This capability of genes (or in this case, ribonucleotides) to buffer other genes against the effects of mutations is known as epistasis. More generally, epistasis is the non-independence of loci. When the fitness consequence of a mutation at one site depends on the nucleotides present at other sites, then there is epistasis between the sites. In RNA we observe the evolution of epistatic buffering, where critical base pairs render the minimum free energy structure robust to most point mutations.

In summary, the plastic repertoire of an RNA molecule indicates how much and in which ways its ground state can be altered by mutation. We call this statistical alignment between the thermodynamic sensitivity of the ground state and its genetic mutability "plastogenetic congruence". We emphasize that this is not an assumption of our model, but a hitherto unknown statistical property of RNA folding algorithms that were developed independently of our evolutionary study. A similar correlation between the thermodynamic stability of the native conformation and its mutational robustness has been found in models of protein folding[16–18].

What are the consequences of plastogenetic congruence for our evolutionary model? Natural selection produces sequences with low plasticity (Figure 1A). Low plasticity sequences are, by virtue of plastogenetic congruence, highly buffered against the effects of mutation. A large fraction of all possible single point mutations on such sequences will preserve the ground state (Figure 2B). Futhermore, the low plasticity sequences not only reside mostly in their ground state, but spend the rest of their time in shapes that are structurally akin to it (Figure 1A). Again, by plastogenetic congruence, the

rare mutations that alter the ground state are likely to cause only slight structural changes. Thus, the phenotypic variability of the population, that is, the potential variation accessible through mutation, is dramatically curtailed. Plastogenetic congruence has steered the population into an evolutionary dead-end which we call "neutral confinement". This does not occur in the control populations (Figure 2A).

Low plasticity is achieved through increasing the thermodynamic independence of any one structural component from the remaining structure. At the same time, the effects (if any) of mutations become increasingly limited to local shape features. Wagner and Altenberg[2] argued that modularity evolves in organisms by such a decrease of pleiotropy. This process underlies the extreme modularity of ground states that we observe at neutral confinement. We distinguish here between modules defined in purely morphological (syntactical) terms and modules defined on the basis of thermophysical, kinetic and genetic autonomy. The former is a trivial notion in RNA, since it is implicit in the very definition of a secondary structure as a combination of loops and helices. We address the latter. (In the case of RNA secondary structures, these two notions are conveniently consistent with each other: the thermodynamic stabilization of the ground state as a whole can easily proceed through the stabilization of its component helices and loops.)

From a thermophysical perspective, low-plasticity shapes are composed of structural components that remain intact over large temperature regimes and that melt in distinct phase transitions as discrete units (Figure 3A). In particular, the melting of one unit does not disturb the other units (Figure 3B). This is in sharp contrast to the melting behavior of high-plasticity sequences with the same ground state (Figure 3A and 3C). From a kinetic perspective, these same units fold independently as suggested by a single folding funnel which dominates the energy landscape of a low-plasticity sequence (Figure 3D). A conformational energy landscape so organized prevents the occurrence of energetically trapped intermediates by guiding the folding events reliably and quickly towards the ground state. Again, this is in sharp contrast to the high-plasticity sequence whose energy landscape provides no guidance to the folding process (Figure 3E). Genetic autonomy is seen when sequence segments underlying these structural units are transposed from their original context into random contexts. Low-plasticity segments maintain their original shape with a much higher likelihood than the fragments of random sequences with the same shape. For example, the sequence segments underlying the shape features labeled a and b in Figure 3A maintain their original shape with probabilities 0.83 and 0.94, respectively, when flanked by random segments of half their size. These are much larger than the probabilities 0.017 and 0.015, respectively, for a random sequence with the same shape (sequence shown on the left of Figure 1A). There is computational evidence for such transposability in natural sequences[19].

While it is unclear how natural selection could generate modularity directly, there are many scenarios in which natural selection favors the reduction of plasticity. In our model, modularity arises as a necessary by-product of that reduction. It is the nature of modules to resist change, hence the process that produces modularity simultane-

ously leads populations into evolutionary dead ends. By enabling variation at a new syntactical level, however, modularity may provide an escape from the evolutionary trap that produced it in the first place.

We will turn next to the study of RNA structural evolution in a fluctuating environment. Evolutionary theory teaches us that phenotypic plasticity is favored under sufficiently heterogeneous conditions. When RNA molecules evolve under macroscopic fluctuations, for example multiple binding targets or changing temperatures, does natural selection produce plastic RNA molecules? If so, how does this impact the evolvability and modularity of the population?

References

1. Bonner, J. T. *The evolution of complexity.* Princeton University Press, Princeton, NJ, (1988).

2. Wagner, G. P. & Altenberg, L. Complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976 (1996).

3. Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).

4. Westhof, E., Masquida, B., & Jaeger, L. RNA tectonics: towards RNA design. *Folding & Design* **1**, R78–R88 (1996).

5. Wuchty, S., Fontana, W., Hofacker, I. L., & Schuster, P. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**, 145–165 (1999).

6. Nussinov, R. & Jacobson, A. B. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. USA* **77**(11), 6309–6313 (1980).

7. Waterman, M. S. Secondary structure of single - stranded nucleic acids. *Studies on foundations and combinatorics, Advances in mathematics supplementary studies, Academic Press N.Y.* **1**, 167 – 212 (1978).

8. Zuker, M. & Stiegler, P. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* **9**, 133–148 (1981).

9. McCaskill, J. S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**, 1105–1119 (1990).

10. Ellington, A. D. RNA selection. Aptamers achieve the desired recognition. *Current Biology* **4**, 427–429 (1994).

11. Fontana, W. & Schuster, P. Continuity in Evolution: On the Nature of Transitions. *Science* **280**, 1451–1455 (1998).

12. Baldwin, J. M. A new factor in evolution. *Am.Nat.* **30**, 441–451 (1896).

13. Simpson, G. G. The Baldwin effect. *Evolution* **7**, 110–117 (1953).

14. Ancel, L. W. A quantitative model of the Simpson-Baldwin effect. *J.theor.Biol.* **196**, 197–209 (1999).

15. Ancel, L. W. & Fontana, W. Plasticity, evolvability and modularity in RNA. *J. Exp. Zool. – Molecular and Developmental Evolution* **288**, 242–283 (2000).

16. Bornberg-Bauer, E. & Chan, H. S. Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. *Proc. Natl. Acad. Sci. USA* **96**, 10689–10694 (1999).

17. Vendruscolo, M., Maritan, A., & Banavar, J. R. Stability threshold as a selection principle for protein design. *Phys. Rev. Lett.* **78**, 3967–3970 (1997).

18. Bussemaker, H. J., Thirumalai, D., & Bhattacharjee, J. K. Thermodynamic stability of folded proteins against mutations. *Phys. Rev. Lett.* **79**, 3530–3533 (1997).

19. Wagner, A. & Stadler, P. F. Viral RNA and evolved mutational robustness. *J. Exp. Zool. – Molecular and Developmental Evolution* **285**, 119–127 (1999).

20. Huynen, M. A., Stadler, P. F., & Fontana, W. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA* **93**, 397–401 (1996).

21. Flamm, C., Fontana, W., Hofacker, I. L., & Schuster, P. RNA folding at elementary step resolution. *RNA* **6**, 325–338 (2000).

22. **Acknowledgement.** The research program of W.F. at The Santa Fe Institute is supported by Michael A. Grantham. Partial support came from the Keck Foundation and an NSF grant to the Santa Fe Institute, an NIH grant to M.W. Feldman at Stanford, a U.S. National Defense Science and Engineering Fellowship to L.A. and the Program in Theoretical Biology at the Institute for Advanced Study, Princeton, NJ.

Figure legends

Figure 1 — Loss of plasticity and plastogenetic congruence.

A: On the right: all six shapes in the plastic repertoire of the most frequent sequence after 10^7 replications in the plastic simulation depicted in Figure 2A. The few shapes in the plastic repertoire of the evolved sequence are structurally similar to each other and the ground state. On the left: a subset of the 1208 shapes in the plastic repertoire of a randomly chosen sequence with the same ground state shape. Dots stand for many shapes not displayed. The number to the left of a shape is its equilibrium probability. The vertical lines on the right measure 3 kcal/mol, and each shape points to a height proportional to its energetic distance from the minimum free energy.

B: The graph illustrates the mechanisms underlying plastogenetic congruence in a small sequence. Aside from a graphical depiction of shapes, we also use a string representation in which a dot stands for an unpaired position, and a pair of matching parentheses indicates positions that pair with one another. A highly plastic sequence (i) is shown in its ground state shape α together with a list of all shapes in its plastic repertoire (numbers indicate equilibrium probabilities). Single point mutations readily tip the energy balance in favor of another shape. For example, a point mutation from U to C in the loop of (i) makes its suboptimal shape β the new ground state in (ii). Generally, single point mutations tip a highly plastic sequence in favor of shapes already present in its plastic repertoire. Single point mutations can also act to reinforce the ground state. For example, a mutation from U to C (iii) generates a better stacking pair in the helix of α, and dramatically reduces the plasticity. As a consequence, the mutation which altered the ground state shape previously, (i) $\rightarrow$ (ii), no longer has a phenotypic effect, (iii) $\rightarrow$ (iv). Thermodynamic stabilization dramatically decreases variability – access to new structures through mutation.

Figure 2 — Evolutionary dynamics and neutral confinement.

A: A simulated population of RNA sequences undergoes mutation, replication and selection in a chemical flow reactor constrained to fluctuate around 1000 individuals. This process models a stochastic continuous time chemical reaction system[11,20]. We graph the average population distance from the target shape with respect to replication events rather than external time.
The selective value $f(\sigma)$ of a shape σ is defined in terms of the Hamming distance $d(\sigma, \tau)$ between the string representations (see Figure 1B caption) of σ and the target shape τ: $f(\sigma) = \frac{1}{0.01+d(\sigma,\tau)/n}$, where n is the sequence length (here $n = 76$). The fitness (or replication constant) of a plastic sequence is given as $r = \sum_{\sigma_i} f(\sigma_i)\frac{\exp(-\Delta G_{\sigma_i}/kT)}{Z}$, where the sum runs over all shapes σ_i with free energies ΔG_{σ_i} within 3 kcal/mol from the minimum free energy. The control population comprises sequences that are mapped simply to their minimum free energy shapes σ_0, and hence have fitness $r = f(\sigma_0)$. Replication accuracy per position is 0.999.

B: For each sequence species present in a given population, we compute its neutrality, that is, the fraction of single point mutants that preserve the ground state. The graph compares the distributions of neutralities for the plastic population (solid) and the non-plastic control (dotted) after 10^7 replications.

Figure 3 — Thermophysical and kinetic modularity.

A: The calculated melting behavior (specific heat versus temperature) as it would appear in a differential scanning calorimetry experiment. The modular (evolved) RNA molecule (solid line) melts in two sharp phase transitions. The dotted line depicts the melting behavior of a high-plasticity sequence.

B: The ground states of the modular (evolved) RNA molecule as a function of temperature (°C). The 37°C features have extended thermal stability, and melt individually at distinct temperatures while leaving other parts of the shape unaffected. In other words, the melting behavior is discrete.

C: The succession of ground states with rising temperature of a high-plasticity sequence with the same shape at 37°C as the evolved molecule in part B. The 37°C shape of the sequence is unstable upon temperature perturbations in both directions, and undergoes global rearrangements as the 37°C features destabilize with rising temperature.

D: Given a kinetic model of the folding process, the low energy portion of a molecule's conformational landscape can be represented as a tree[21]. The ordinate measures free energy, and the abscissa has no meaning. A leaf corresponds to a shape at a local energy minimum and the height of a branch point corresponds to the energy barrier between two local minima. Part D shows that the folding landscape of the evolved sequence is organized as a funnel leading directly to the ground state shape. Modules fold independently.

E: The energy landscape of the random sequence provides little or no guidance to the folding process, and results in frequent deadlocks at local minima.

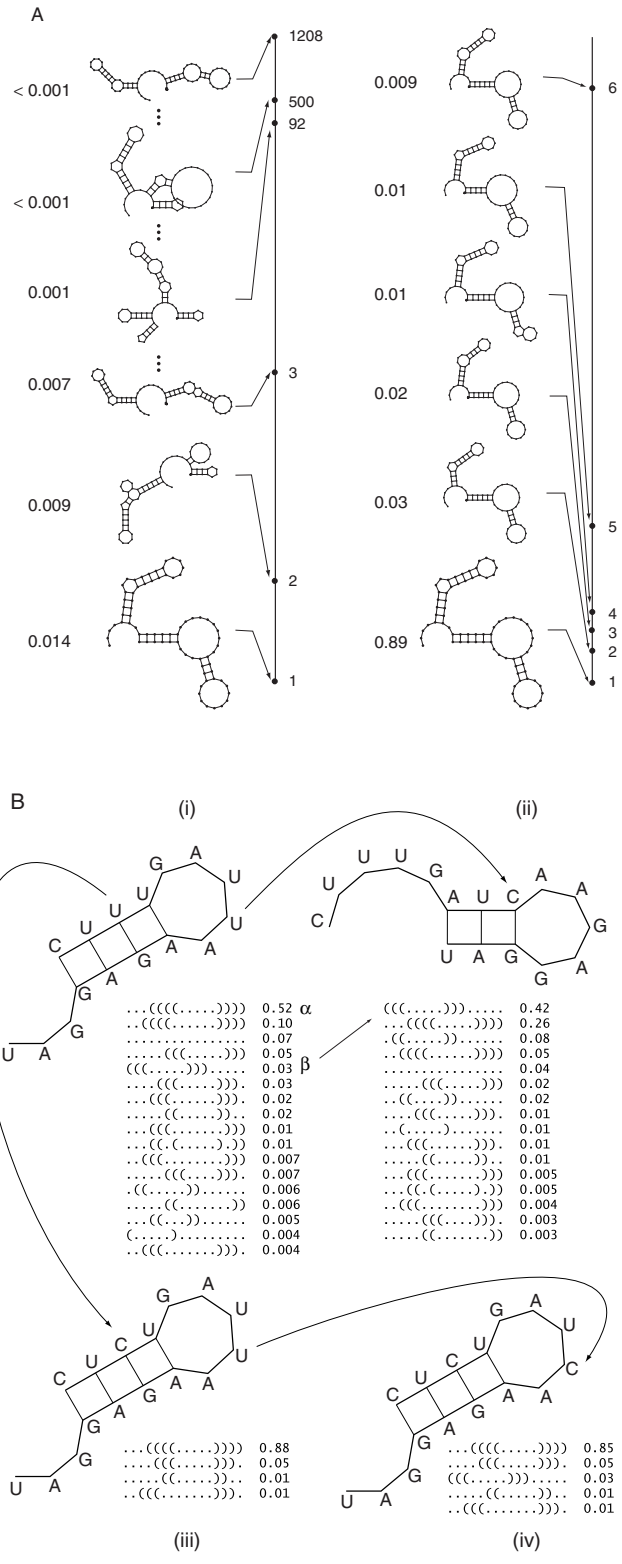
A
< 0.001
< 0.001
0.001
0.007
0.009
0.014
1208
500
92
3
2
1
0.009
0.01
0.01
0.02
0.03
0.89
6
5
4
3
2
1
B
(i)
(ii)
...((((.....)))) 0.52 α
..((((......)))) 0.10
.... 0.07
.....(((.....))) 0.05
(((.....)))..... 0.03 β
....(((.....))). 0.03
...(((.......))). 0.02
.....((.....)).. 0.02
...(((.......))). 0.01
...((.(.....).)) 0.01
..(((........))). 0.007
.....(((....))). 0.007
.((.....))...... 0.006
.....((.......)) 0.006
...((..))...... 0.005
(.....)......... 0.004
..(((.......))). 0.004
(((.....)))..... 0.42
...((((.....)))) 0.26
.((.....))...... 0.08
..((((......)))) 0.05
.... 0.04
.....(((.....))) 0.02
..((....))...... 0.02
....(((.....))). 0.01
..(.....)....... 0.01
...(((.......))). 0.01
.....((.....)).. 0.01
...(((.......))) 0.005
...((.(.....).)) 0.005
..(((........))) 0.004
.....(((....))). 0.003
.....((.......)) 0.003
(iii)
...((((.....)))) 0.88
....(((.....))). 0.05
.....((.....)).. 0.01
..(((.......))). 0.01
(iv)
...((((.....)))) 0.85
....(((.....))). 0.05
(((.....)))..... 0.03
.....((.....)).. 0.01
..(((.......))). 0.01

Figure 1

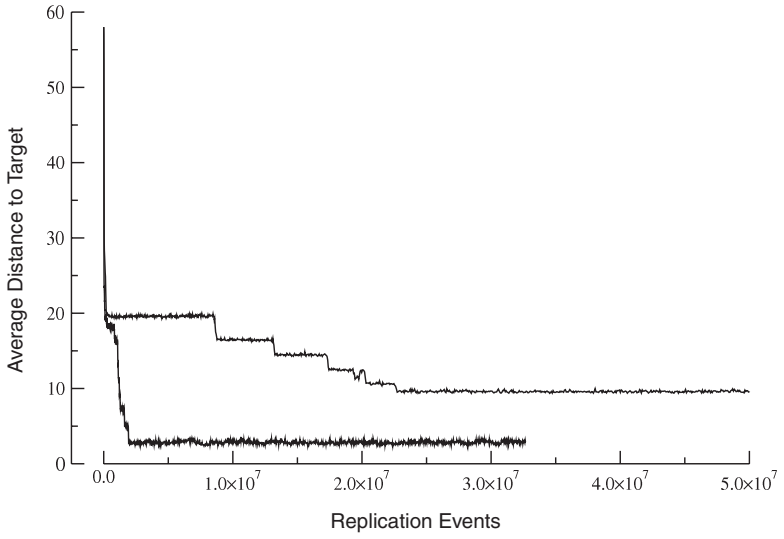

Figure 2A

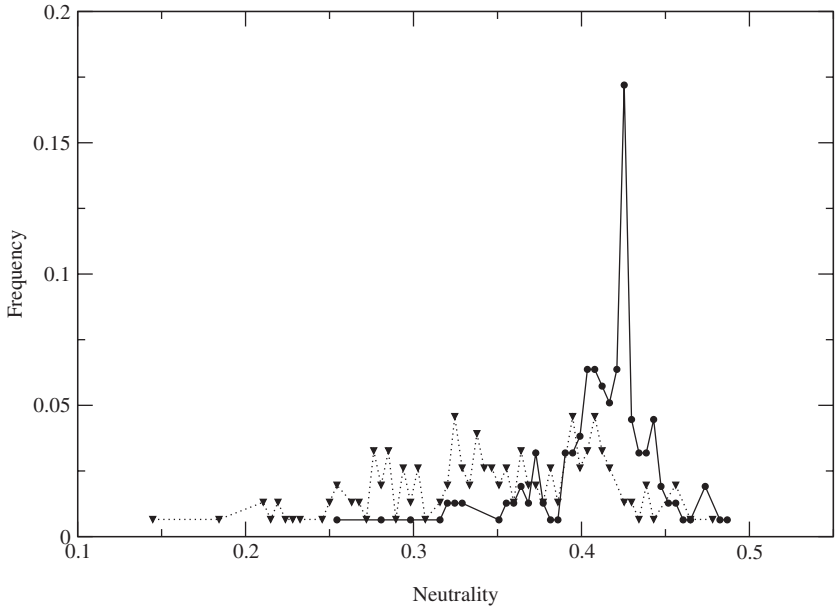

Figure 2B

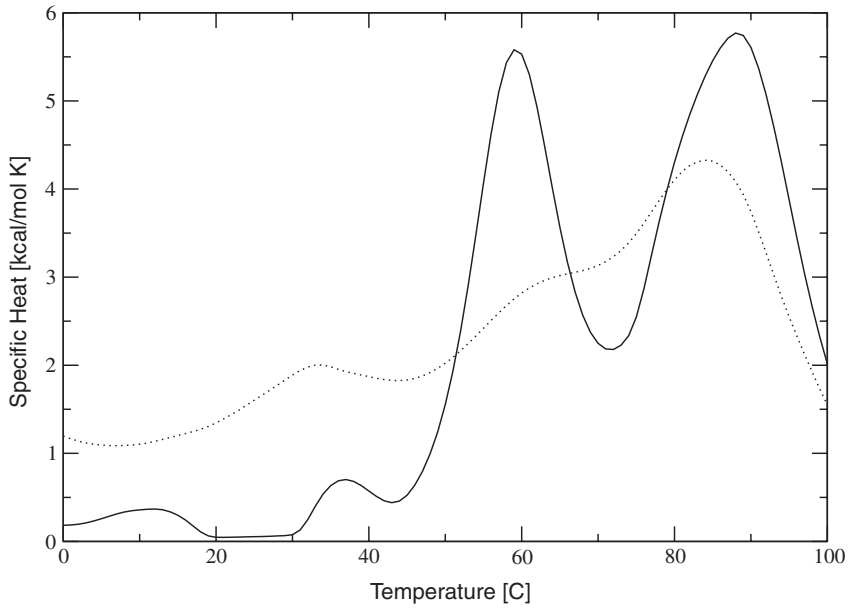

Figure 3A

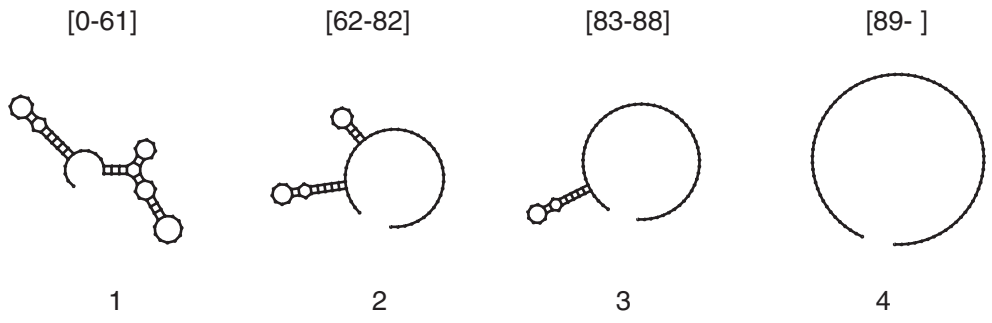

Figure 3B

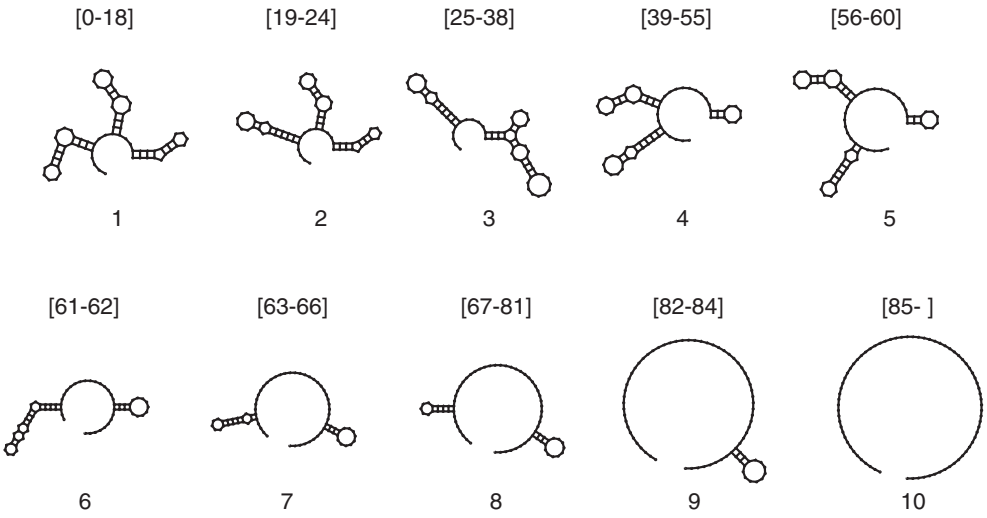

Figure 3C

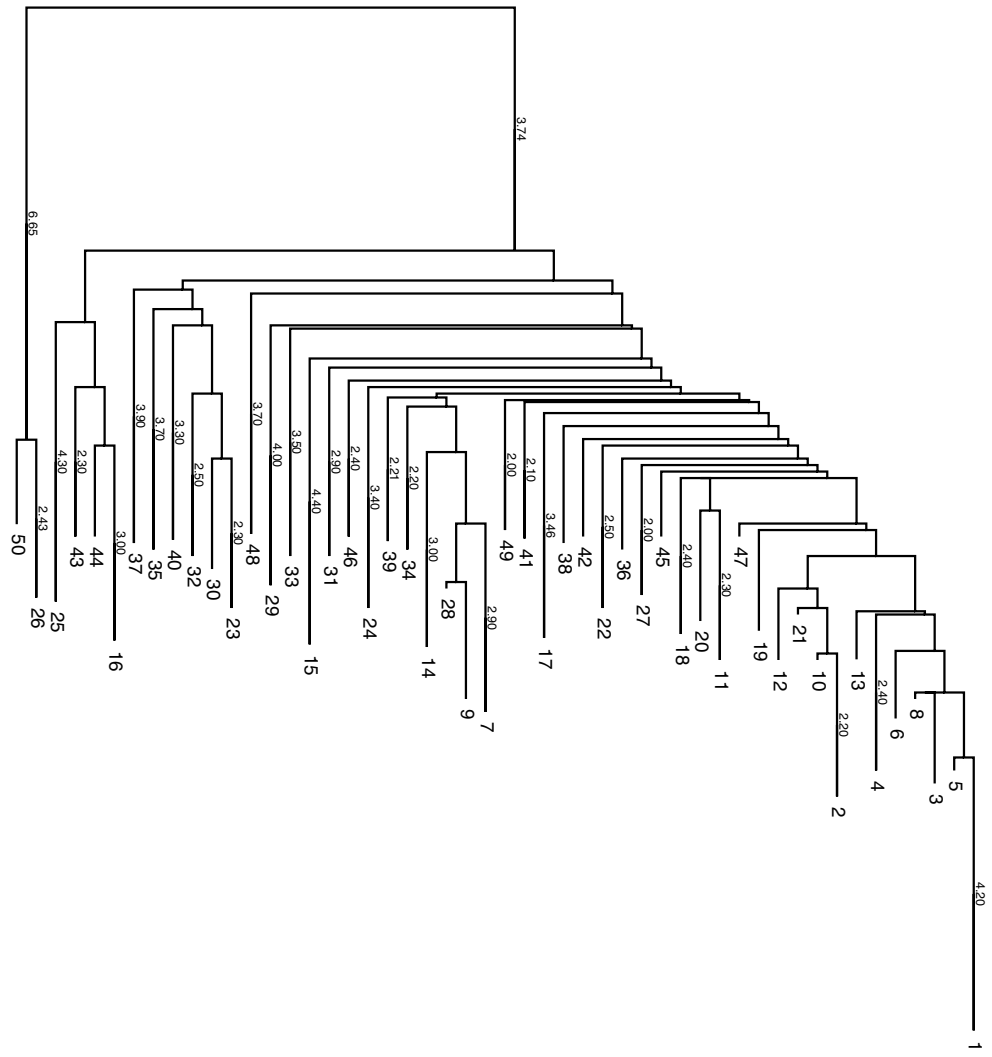
3.74
6.65
50
26
25
43
44
16
37
35
40
32
30
23
48
29
33
15
31
46
24
39
34
14
28
9
7
17
49
41
38
42
22
36
27
45
18
20
11
47
19
12
21
10
2
13
4
6
8
3
5
1
2.43
4.30
2.30
3.00
3.90
3.70
3.30
2.50
2.30
3.70
4.00
3.50
4.40
2.90
2.40
3.40
2.21
2.20
3.00
2.90
2.00
2.10
3.46
2.50
2.00
2.40
2.30
2.20
2.40
4.20

Figure 3D

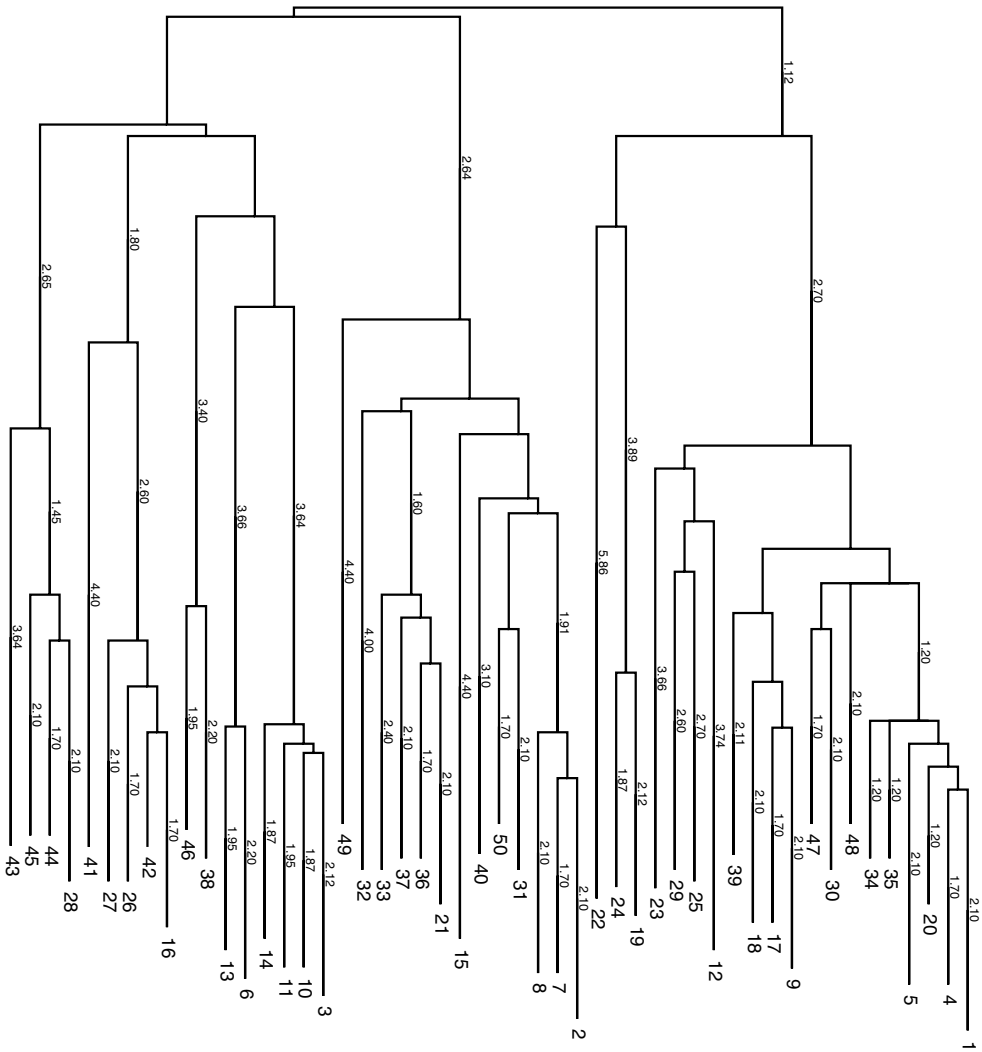
1.12
2.64
2.70
1.80
2.65
3.40
3.89
2.60
1.45
3.66
3.64
1.60
5.86
4.40
1.91
1.20
4.40
4.00
3.64
3.10
2.10
2.10
2.70
2.60
3.74
2.10
1.70
2.10
1.20
1.20
1.70
2.10
2.10
1.95
2.20
1.70
2.10
2.40
2.10
1.70
2.10
1.87
2.12
2.10
1.70
2.10
1.70
2.10
2.10
1.70
2.10
1.95
2.20
1.87
1.95
1.87
2.12
49
32
33
37
36
21
15
40
50
31
8
7
2
22
24
19
23
29
25
12
39
18
17
9
30
47
48
34
35
5
20
4
1
43
45
44
28
41
27
26
42
16
46
38
13
6
14
11
10
3

Figure 3E