Recombination During *In Vitro* Evolution

Niles Lehman,[1] **Peter J. Unrau**[2]

[1] Department of Chemistry, Portland State University, PO Box 751, Portland, OR 97207, USA
[2] Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada V5A 1S6

Abstract. Recombination, the swapping of large portions of genetic information between and among parental genotypes, can be applied to *in vitro* evolution experiments on functional nucleic acids. Both homologous and heterologous recombination can be achieved using standard laboratory techniques. In many cases, recombination can allow for the discovery of a ribozyme or DNAzyme phenotype that would not likely be encountered by reliance on point mutations alone. In addition, recombination can often aid in the discovery of global optima in sequence space and/or lessen the number of generations it would take to reach optima. Recombination is most efficiently used in combination with point mutations and applied after the first couple of rounds of selection but before high-fitness genotypes dominate the selection. The "recombination zone" describes that region of sequence space—defined by the residues that will ultimately participate in the function of the winning nucleic acid(s)—where recombination is expected to be the most beneficial in the search for high-fitness genotypes.

Key words: RNA — Ribozymes — Recombination — *In vitro* evolution

Author order determined by a single Bernoulli trial as implemented by RPS.

Correspondence to: Niles Lehman; *email:* niles@pdx.edu or Peter J. unrau; *email:* punrau@sfu.ca

Introduction

Evolution in a test tube is a powerful method to explore the catalytic capabilities of biological polymers such as nucleic acids. Invoking the principle of "irrational design," we can now uncover RNA, DNA, and even protein sequences with desired properties even though they may be extraordinarily rare in a randomized pool and/or have structure-function relationships that are entirely unpredicted beforehand. In essence, this technique is the laboratory equivalent of a blind man picking a golden needle out of an enormous (football-stadium-sized!) haystack. If that were not impressive enough, evolution *in vitro* gives us an almost unlimited ability to explore the operations of selection and evolution in a tightly controlled and finely parameterizable system. With it, we can set up and test basic evolutionary hypotheses in a fraction of the time that would be needed with systems that use living organisms. Often we can perform these tests with a precise change of a single variable such as temperature, time, or salt concentration, a task not easily made in many *in vivo* studies.

We have been taking advantage of *in vitro* evolution to discover new catalysts for about 15 years, and the successes have been remarkable at times, mundane at others. One example of a spectacular hit is the Diels-Alderase ribozyme: a catalytic motif of approximately 50 ribonucleotides that can create carbon-carbon bonds via a [4+2] cycloaddition reaction at rates thousands of times that of the uncatalyzed rate (Tarasow et al. 1997; Seelig and Jaschke, 1999; Keiper et al. 2004). Another is the

ligase ribozyme, a structure of 90–150 nucleotides that facilitates the attack of a 3′-OH on a phosphoanhydride bond and that is the fastest known ribozyme to date (Bartel and Szostak 1993; Bergman et al. 2000) Both of these examples are notable first in that they were plucked from incompletely sampled random pools of RNA sequences that have potential diversities exceeding 10^{30}, and second in that they catalyze chemical reactions uncommon or even unknown in natural ribozymes.

Despite these achievements, one can argue that we have been rather inefficient in our test tube experiments so far. In particular, we have not exploited all the tools that evolutionary biology has to offer, such that even more powerful catalysts remain undiscovered. As a discrete example, consider the Diels-Alderase ribozyme mentioned above. The initial discovery of such a ribozyme reported an 800-fold rate enhancement by the ribozyme over the uncatalyzed reaction (Tarasow et al. 1997). Subsequently, other efforts, using different initial pool sizes and variations on the selection strategy, have produced faster Diels-Alderases with rate enhancements approaching 20,000-fold (Keiper et al. 2004). Such studies show the benefit of exploring sequence space beyond a local optimum, with the payoff of several orders of magnitude in catalytic prowess.

In nature, the combination of mutation, drift, and selection combine to engender new phenotypes within an ever-changing ecological context. *In vitro* selection experiments provide one or perhaps two sorts of mutational events (typically initial pool randomization procedures augmented by polymerase-induced point mutations) and attempt to apply the strongest and quickest selective force feasible. Recombination, the swapping of large blocks of genetic information among genotypes, as a mutational event is far less employed in the test tube despite its widespread occurrence in the history of life (Felsenstein 1974; Maynard Smith and Szathmáry 1995; Lehman 2003). However, early accomplishments in protein selection have recognized the utility of recombination as a means to explore the vast polypeptide sequence space (cf. Stemmer 1994a; Zhao et al. 1998; Joern et al. 2002). A few studies have investigated recombination in nucleic acid selections to a limited extent (e.g., Biebricher and Luce 1992; Burke and Willis 1998; E.A. Curtis and D.P. Bartel, personal communication), but as a means of catalytic discovery it still remains largely untapped.

Here, we examine the potential of recombination as an aid in the search for new ribozymes and DNAzymes. We define recombination as a mutational process that swaps blocks of contiguous nucleotides between two genotypes. This is in contrast to point mutation, which changes the identities of single nucleotide positions. We stress that both events are of great utility during *in vitro* evolution experiments, but that under certain conditions point mutation alone may falter or fail unless accompanied by occasional recombinatorial events.

Types of Recombination

In principle, there are two types of recombination: homologous and heterologous (Fig. 1). Homologous recombination is the basis of equal crossover in genetics, and involves the exchange of portions of two sequences that share a recent common ancestor and thus are identical (or nearly so) in length and align with a high degree of sequence identity. Heterologous recombination, on the other hand, is the equivalent of unequal crossover, and results in a disproportionation event between two sequences that may be quite dissimilar in both length and sequence. Both types of recombination can occur *in vitro* and can thus have an impact on the evolutionary trajectory of the population of nucleic acids. These events can happen unintentionally, but recently some effort has been made to develop techniques by which the experimenter can modulate the recombination frequency during *in vitro* evolution.

Homologous Recombination

Homologous recombination occurs when two or more similar sequences shuffle their allelic differences. This happens in the lab more often than many people realize, as a consequence of the intrinsic propensity of DNA polymerases such as *Taq* and reverse transcriptase to jump erroneously from one template to another. This is actually a natural feature of viral population dynamics (e.g., Jetzt et al. 2000), but in the test tube, template jumping can lead to some rather unexpected combinations of input sequences, especially during the PCR (Meyerhans et al. 1990; Bradley and Hillis 1997; Judo et al. 1998). Fortunately, we can harness this phenomenon to introduce recombination artificially into an evolving lineage of nucleic acids *in vitro*. First described by Stemmer (1994a,b), a heterogeneous pool of DNA sequences can be treated with DNA exonuclease for a period of time sufficient to partially degrade most of the molecules into fragments. Fragments in a particular size range are then preferentially gel purified and then input into a "primerless" PCR. The *Taq* DNA polymerase will use any existing overlapping sequences to reconstruct full-length, but genotypically scrambled, molecules. This basic strategy was improved upon by Arnold, who employed extremely short extension times (e.g., 0–5 s) in the PCR to engender recombination without the need for DNase treatment or gel purification (Zhao et al. 1998). Both
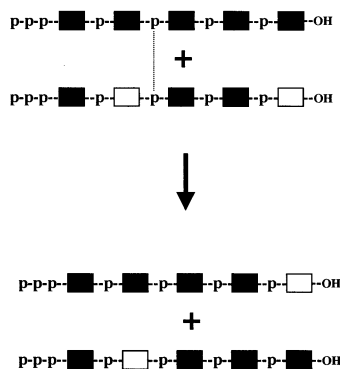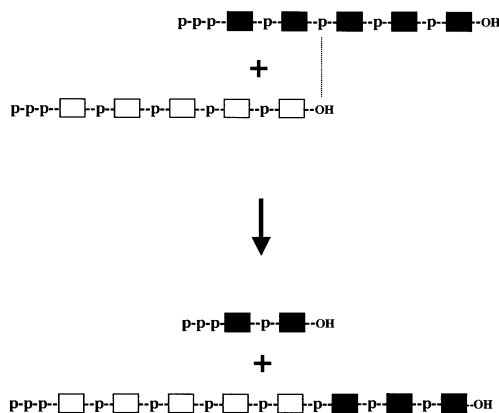
homologous recombination *heterologous recombination*

Fig. 1. A comparison of homologous and heterologous recombination. Homologous recombination (left) involves the reciprocal swapping of blocks of contiguous nucleotides between two or more very similar parental nucleic acids. Heterologous recombination (right) involves the unequal swapping of blocks of contiguous nucleotides between two or more parental nucleic acids that need not be similar in sequence or in length. The crossover points in each case are denoted by small dashed lines.

of these methods give essentially homologous recombination, because they will work only if the genotypes in the pool are above a certain threshold of sequence identity. Usually sequences must be identical at roughly 75% of their sites for these reconstruction methodologies to be efficient, and crossover sites tend to aggregate in regions where the input sequences match closely (Moore et al. 2001). To date, these techniques have been most often applied to protein molecular evolution (cf. Patten et al. 1997; Petrounia and Arnold 2000; Kurtzman et al. 2001; Castle et al. 2004), but their potential utility in nucleic acid selections is clearly enormous.

Other means to induce homologous recombination exist. If crossovers at a specific nucleotide position are desired, then homologous double-stranded DNAs can be cut using a restriction enzyme, and the resulting pool of fragments can be religated together with DNA ligase (Zaher and Unrau 2005). This procedure can be carried out at the RNA level as well, using sequence-specific RNA endonucleases such as group I intron ribozymes or certain ribonuclease protein enzymes, and the products can be recombined using RNA ligase. The RNA approach is prone to generating heterologous recombination, however, because of the possibility that 5' fragments, say, could be ligated to one another. This problem can be lessened by the production of chemically unique 3' ends following cutting, or by the use of sequence-specific splints to guide proper religation. Lastly, if a relatively small number of nucleotide positions are candidates for recombination, one can synthesize degenerate DNA (or RNA) oligonucleotides containing all known or suspected nucleotides at the sites in question and perform a selection on the resulting pool. One can then recombine all the parental sequences at any desired frequency; if all sequences are to be given an equal opportunity to

recombine, then the synthetic oligo should contain degenerate bases in equal frequencies. This "slot machine wheel" approach has been successfully employed to apply recombination in the search for optimal kinase ribozymes (E.A. Curtis and D.P. Bartel, personal communication).

Heterologous Recombination

Heterologous recombination involves the reassortment of nucleic acid fragments between unrelated strands making possible the deletion, insertion, inversion, or translocation of blocks of sequence. This form of recombination has been hypothesized to play an important role in protein evolution. The Exon Shuffling model of Gilbert, Tonegawa, and Go (Gilbert 1978; Tonegawa, 1978; Go 1983) postulates that once small protein modules have been evolved, relatively little evolutionary effort would be required to reorder or shuffle these elements to create completely new protein folds. However, little experimental attention has been focused on this type of recombination. This may be a consequence of the dramatic rearrangements possible with this approach, which, in contrast to homologous recombination, can easily be imagined to disrupt rather than optimize a given function.

Nevertheless, the utility of heterologous recombination to isolate particular nucleic acid motifs or modules has been demonstrated in at least three cases. Burke and Willis (1998) split a randomized region of 150 nucleotides into two portions, an N70 pool and an N80 pool. The two halves each contained complementary fixed regions on either their 3' (N70) or 5' (N80) ends such that a composite double-stranded DNA could be created by overlap extension. Chimeric RNAs containing heads from one pool and tails from another combined the functions selected

from either pool and were shown to bind simultaneously two small molecules such as coenzyme A or adenosine. The main result here was that complex functionalities could readily be preserved through recombination. Later, Bittker and coworkers pioneered the use of scrambled DNA pools created from a low diversity library of input sequence. Such pools have been used in conjunction with *in vitro* selection to isolate the minimal motif of a DNA aptamer towards streptavidin (Bittker et al. 2002). Recently, this approach has been taken to recombine motifs within the catalytic core of a nucleotide synthase ribozyme (Wang and Unrau 2005). As with homologous recombination, DNA sequences are digested with DNase I, but with the intent of producing quite small fragments (~15 to ~75 bp). These fragments are polished with T4 DNA polymerase into blunt-ended constructs that are finally ligated back together using T4 DNA ligase. Because the ligation of blunt-ended fragments is sequence-independent, fragments are joined in all permutations and orientations by this approach, resulting in a diverse population even when starting from a single initial sequence. While the vast majority of these sequences are inactive, screening this pool for function using *in vitro* selection quickly leads to the identification of fragment combinations that are essential for activity.

It is also possible to perform recombination at the RNA level, obviating the need to scramble DNA templates. This was first demonstrated by Mörl and Schmelzer (1990), who were able to coax the yeast group II ribozyme to recombine 5′ "head" and 3′ "tail" portions of two large (250–625 nt) RNA substrates. Riley and Lehman (2003) generalized this strategy using the *Azoarcus* group I ribozyme in order to recombine smaller RNA oligonucleotides that contain the internal splicing guide sequence complement CAU immediately upstream of the crossover site. The *Azoarcus* ribozyme, operating at 60°C, can circumvent potential secondary structure obstacles and recombine two inert RNA motifs into fully functional ribozymes. In principle then, heterologous RNA recombination could be infused into many types of *in vitro* selection schemes without much additional effort (Jäschke and Helm 2003).

Improving the Efficiency of Sequence Space Searches

Probability of "Winning"

The tools that nature has available to evolve linear nucleic acid polymers are limited. Point mutation and recombination provide the only mechanisms to introduce sequence variation if one ignores nucleotide modifications of the sort commonly found in tRNAs (e.g., pseudouridine). What then are the relative advantages of each form of variation, how might each be expected to play a role during the evolution of complex functional RNAs, and to what extent can we guarantee that a true global optima has been found?

These questions can perhaps be best addressed by considering the organization of sequence space. Points within this space represent particular sequences and the movement of points corresponds to evolution in response to selective pressures. Sequence space is infinite and can be conceptualized by ordering sequences into a series of concentric shells, like the layers of an onion, with each shell containing sequences of a particular length. The number of sequences up to a given length quickly becomes astronomical (there are $4^{n+1}/3$ sequences contained within a sphere of sequences having lengths $\leq n$; $n = 100$ gives $\sim 2.1 \times 10^{60}$ sequences, $n = 200$ gives $\sim 3.4 \times 10^{120}$). It is within this vast arena that the process of RNA evolution can be mathematically represented.

A logical focus is to explore the density of structured RNAs in sequence space. The vast majority of RNA sequences are expected to form structures, and it is the variety and distribution of these shapes that will determine the ability of evolution to find optimal solutions. Computer analyses suggest that the number of unique shapes found in a given shell of sequence space has an asymptotic upper bound of form $S(n) = 1.48n^{-3/2}1.85^n$ (Fontana 2002). This estimate indicates that the number of shapes increases exponentially as sequence length increases, but not nearly so fast as the number of sequences, which increases as 4^n. As a consequence, the total effective density of structured RNAs (defined here to be $S(n)/4^n$) decreases exponentially with increasing RNA length. In other words, while an outer shell contains more shapes, the number of these shapes relative to the total number of sequences in the shell (many of which have the same shape) drops, relative to that found in an inner shell.

These estimates of shape density in sequence space can be used to calculate the probability that point mutation can reach any other shape by following a random walk through sequence space. A point mutation moves within a shell to a nearest neighbor immediately adjacent to the starting sequence, while point deletions and insertions move inward and outward by one shell, respectively. As summarized in Fig. 2, point mutation can easily be used experimentally to explore all the shapes available from small sequences up to lengths of perhaps 60 to 65 nucleotides. At this point the number of possible shapes exceeds the number of mutants that can be constructed using current approaches ($\sim 10^{15}$), precluding an exhaustive sampling of shapes in a single conventional library. This argument is rigorously true so long as all shapes are created equally. Theoretically, the distribution of shapes is biased, with some
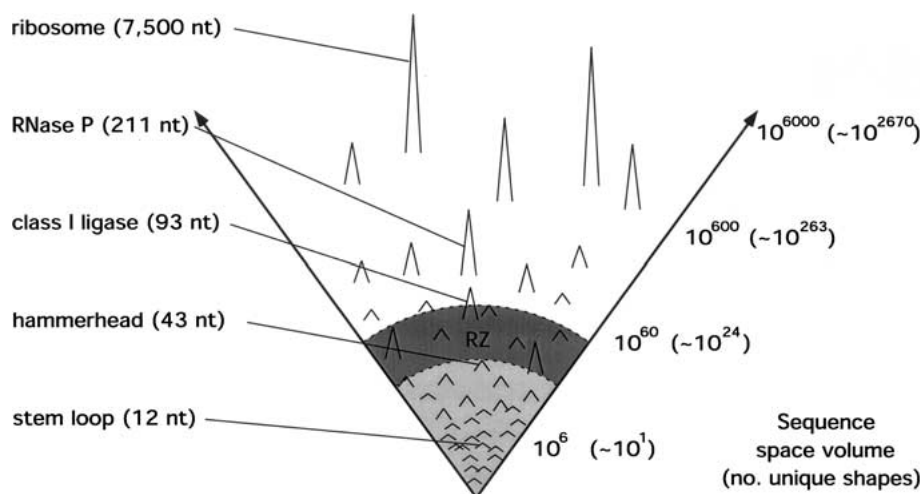
Fig. 2. A schematic of sequence space showing the interplay between motif size and the recombination zone for a number of naturally and artificially selected RNA motifs. Small motifs occur with high density, and relatively few mutations are required to shift one shape to another in this region (light gray shading). Few of these motifs are proficient at any given catalytic or other functional task but are easily sampled. Larger motifs are more functionally proficient but are difficult or impossible to sample by traditional *in vitro* selection. The recombination zone (dark gray shading marked RZ, Fig. 3) highlights a range where the discovery or optimization of RNA motifs could be assisted by recombination. The height of each motif peak indicates functional ability while the distance from the origin measures motif length on a log scale.

shapes being common and others being quite rare. As this distribution is expected to flatten with length (Fontana 2002), pools longer than 65 nt have been successfully used to isolate quite complicated motifs that would be rare or impossible to discover using shorter libraries (Bartel and Szostak 1993; Sabeti et al. 1997; Unrau and Bartel 1998).

Regardless of the pool used initially to isolate a winning sequence, *global* optimization depends sensitively on the number of residues required to specify its function. We can, therefore, define a recombination zone (RZ), a region of parameter space in which recombination would be expected to greatly aid in the exhaustive optimization of genotypic targets that ultimately have functional sequences within this range (Figs. 2, 3). For RNA populations formed by creating a completely random segment of RNA sequence (Fig. 3A), this zone would extend from approximately 25 to 65 nt and defines a space where the sampling of shapes quickly moves from complete saturation to bare coverage. An analogous RZ would be in effect both for homologous recombination applied to partially randomized RNA populations, where a wild-type sequence is mutagenized to a certain level (Fig. 3B). Both of these RZs occupy the region of sequence space just out of reach of the saturation level achievable in the laboratory. In Fig. 3B, for example, at values of k beyond the saturation boundary, there is a low probability that the mutations contained in the target sequence(s) exist in the initial pool or can accumulate during selection. In the RZ, there is still a high probability that significant subsets of the necessary mutations exist in individual molecules such that recombination can produce a highly optimized sequence with reasonable frequency. At still larger values of k, however, the homology among the sequences in the population drops below a value that would allow for the retention of information after recombination. Thus, the RZ extends in Fig. 3B only as far as the probability of "inbreeding" among homologous sequences containing different subsets of beneficial mutations exceeds the probability of "outbreeding" that either brings in too many deleterious mutations or disrupts adaptive suites of mutations. Similar logic can be applied to Fig. 3A, except that here recombination is primarily heterologous, and the RZ depends more heavily on the length of the randomized region as opposed to the number of mutations away from a wild-type contained in the eventual target.

Outside of this zone, heterologous recombination of RNA motif elements may have a significant advantage over any current selective technique to find new and complex shapes in sequence space by allowing searches for large modular RNAs. Large, highly structured RNAs are not common in modern metabolism but play a number of essential metabolic roles. RNaseP and the ribosome, as the two most prominent examples, lie far above the RZ that we propose could facilitate the evolution of smaller RNAs. In contrast to simpler motifs such as the hammerhead ribozyme (Uhlenbeck 1987; Carothers et al. 2004) and many of the moderately sized regulatory RNAs (Barrick et al. 2004) that appear to have motifs that lie within the RZ, these large RNA motifs are not likely to emerge spontaneously (Benner et al. 1989). This is not to say the components of these RNAs are not capable of evolution. Recently, in fact,
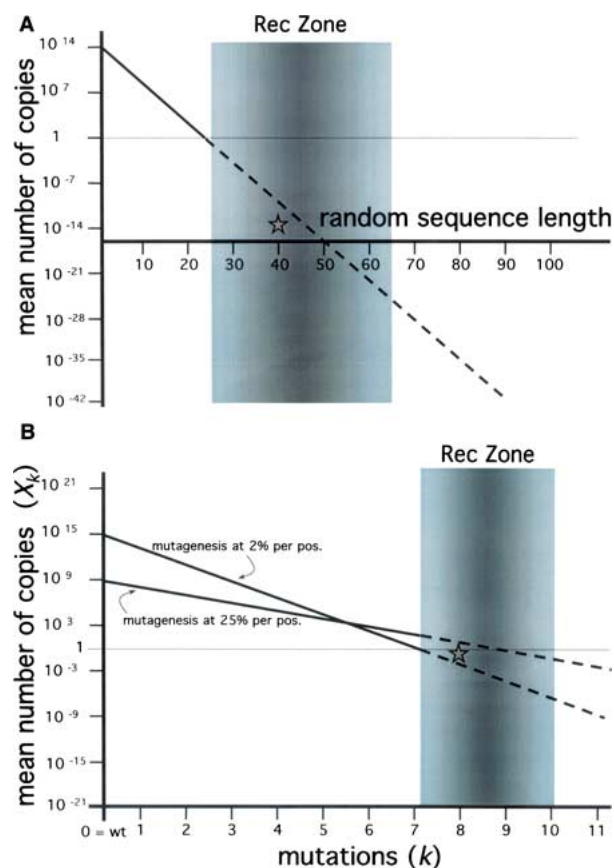
Fig. 3. The recombination zone (RZ), where recombination is most likely to augment significantly the search for a high-fitness genotype during evolution *in vitro*. Selections performed on both completely random nucleic acids pools (**A**) and randomized wild-type pools (**B**) can be aided by recombination. Heterologous recombination is used primarily in A, homologous recombination in B. The sloped lines indicate the theoretical relationship expected between measures of complexity (*x*-axis) and mean number of copies of a particular nucleotide sequence that is expected to exist in a 1 nmol initial pool (6×10^{14} molecules). The *y*-axes in both cases represents the expected number of each *unique sequence* in the pool (i.e., not the *total* number of sequences of a particular length or with a particular number of mutations). The dotted lines denote the saturation thresholds, i.e., below which one would not expect that all possible sequences would exist in the pool. The dashed portion of the sloped lines denotes the fact that 1-nmol pools are subsaturated in this region. The RZ is the region of these graphs just below the saturation thresholds: if a target sequence (star) lies in this region, then we would anticipate that recombinatory methods applied during *in vitro* evolution would greatly improve the probability of, or shorten the time to, encountering this sequence during the lineage. B, adapted from Breaker and Joyce (1994), depicts examples in which 50 randomized nucleotide positions in a molecule are considered.

the recognition domain of RNaseP was shown to exist in two functionally equivalent, but structurally distinct motifs (Westhof and Massire 2004; Krasilnikov et al. 2004). It may, therefore, not be unreasonable to postulate that RNA motifs that exist within the RZ could be used to build in a modular fashion more complicated RNAs that lie outside of the RZ. If, for example, in an early organism many reactions were catalyzed by moderately sized RNAs, the evolution of larger RNAs with more complicated

functionalities could be achieved quite naturally by heterologous recombination. This modular RNA argument bears many parallels with the ideas of Gilbert and Tonegawa, and experimentally appears much easier to explore than with protein motifs.

Time to Winning or Optimization (Number of Generations to Convergence)

A second issue of relevance to the practical aspects of *in vitro* evolution is how much effort is required to find, or to optimize, a successful nucleic acid catalyst. These experiments are often costly, and though more facile than in previous years, a selection can be labor intensive and take months, or even years, to come to fruition. In fact, though this is difficult to substantiate, it is our impression that a large fraction of *in vitro* evolution experiments fail entirely. Thus, finding an acceptable answer quickly, in say 6–12 rounds of selection and amplification, could save thousands of research dollars and dozens of person-hours.

In many cases, recombination can speed the discovery of a desired catalyst (Kauffman and Macready 1995). In Fig. 3, for example, the target sequence space can lie below the RZ, can lie within the RZ, or can lie above the RZ, As we discuss above, if the target is within or above the RZ, then point mutation can outright fail to encounter the target, in which case recombination obviously speeds up the investigator's task.

However, recombination can also provide time benefits in the margins of parameter space just below and just above the RZ. In the former case, a relatively low error mutant that exhibits activity will be present in early rounds of selection but be rare. For example, imagine a four-error variant present in a given generation's population at a frequency (*p*) of one in one million. The time (in generations) it will take for this frequency to rise to detection level depends on the selective advantage *s* of that variant over worse or even inactive catalysts—not all selection criteria are 100% stringent—as predicted by standard population genetics theory by the following expression (cf. Kimura 1983, p. 122):

$$dp/dt = sp(1 - p)$$

To rise to a detection level of 5%, a mutant at a 10^{-6} frequency would require about 26 generations if it had a selective advantage of 50% over all other molecules in the population, and about 13 generations if its selective advantage were 100%. (The selective advantage here is the fraction more offspring that the mutant will leave in the next generation compared to the population's average.) These rough calculations assume that the desired mutant is not also arising via mutation from, say, three-error mu-

tants. With relatively high recombination frequencies (5–40%), however, the creation of molecules with the requisite four mutations from appropriate 1-, 2-, and 3-error mutants will be significant.

Naturally, the converse loss of high-fitness genotypes by outbreeding would also occur. But if recombination were performed only after the selection had been applied for at least a couple of generations, then the constellation of mutations that confer high fitness would be well established in the population, and relatively few outbreedings should occur. Be aware, however, that recombination may lose effectiveness if applied too late in the selection lineage, in part because once functional shapes are present with high frequency, then recombination is more likely to disrupt co-adaptive allelic combinations than it is to produce new, higher fitness ones. During later generations then, fine tuning by point mutations is safer. Moreover, there are theoretical and empirical evidences for a negative correlation between recombination rates and fitness (Hadany and Beker 2003); this is exemplified by many microorganisms that induce sexual forms of reproduction only when environmental stress has reduced their fitness below a local optimum.

It is of particular interest that homologous recombination should speed up the advent of genotypes that are the result of the additive effects of independent mutations. Again this can be seen from the four-mutation example. If each of the four mutations themselves incurs a fitness benefit, then after 2–4 generations there should be a large fraction of the population that possesses at least one of these mutations. Recombining the population at that point would engender a good number of higher-error mutants, which would then rise to populational prominence quickly as a consequence of their selective advantage. Without recombination, one would have to wait for the higher-error mutants to arise via point mutation from lower-error mutants and/or for rare high-error mutants to out-compete the multitude of low-error mutants. This non-recombinatory scenario was observed in the case of the selection of Ca(II)-competent *Tetrahymena* ribozymes, where even after 12 generations, the "winning" seven-error mutant of the wild-type molecule represented only 14% of the population, which was dominated by 3–5 error sub-optimal genotypes (Lehman and Joyce 1993).

By the same logic, recombination would also be expected to shorten the number of generations required to encounter extremely high-error variants lying on the edge of the RZ furthest from the *y*-axes in Fig. 3. These are molecules that are not likely to be in the initial population at all, and hope of their encounter in a given experiment could rest solely on the recombination of mutations into a high-error sequence. However, if too many mutations are required, especially in the case where the phenotypic effects of individual mutations are not additive, then recombination could actually be a hindrance by causing the breakup of beneficial mutational combinations. On the other end of the spectrum, if 2 or 3 mutations are needed, then recombination probably would be of no benefit. These cases represent situations too far from the RZ.

Conclusions

The application of recombination to *in vitro* selections for functional nucleic acids can be beneficial both in terms of the probability of success and the time to convergence on an optimal solution. There is a theoretical basis for this view, and there currently are a variety of techniques that can be used in the lab to introduce recombination into a lineage in a controlled way. Recombination is in particular likely to be advantageous when searching for a complex solution that lies just outside of the experimenter's ability to search exhaustively the initial pool of variable sequences, or when disparate structural motifs must be brought together to form a functional whole. The recombination zone describes that region of sequence space, defined by the residues that will ultimately participate in the function of the winning nucleic acid(s), where recombination is expected to be the most beneficial. Of course, in practice, the RZ can be determined only *a postiori* because only then would the identity of the winning sequences become known. However, our prediction is that many selections in the future would benefit from the judicious application of recombination.

References

Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc Natl Acad Sci USA 101:6421–6426

Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. Science 261:1411–1418

Benner SA, Ellington AD, Tauer A (1989) Modern metabolism as a palimpsest of the RNA world. Proc Natl Acad Sci USA 86:7054–7058

Bergman NH, Johnston WK, Bartel DP (2000) Kinetic framework for ligation by an efficient RNA ligase ribozyme. Biochemistry 39:3115–3123

Biebricher CK, Luce R (1992) *In vitro* recombination and terminal elongation of RNA by Qβ replicase. EMBO J 11:5129–5135

252

Bittker JA, Le BV, Liu DR (2002) Nucleic acid evolution and minimization by nonhomologous random recombination. Nat Biotech 20:1024–1029

Bradley RD, Hillis DM (1997) Letter to the Editor: Recombination DNA sequences generated by PCR amplification. Mol Biol Evol 14:592–593

Breaker RR, Joyce GF (1994) Inventing and improving ribozyme function: Rational design versus iterative selection methods. Trends Biotechnol 12:268–275

Burke DH, Willis JH (1998) Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. RNA 4:1165–1175

Carothers JM, Oestreich SC, Davis JH, Szostak JW (2004) Informational complexity and functional activity of RNA structures. J Am Chem Soc 126:5130–5137

Castle LA, Siehl DL, Gorton R, Patten PA, Chen YH, Bertain S, Cho HJ, Duck N, Wong J, Liu D, Lassner MW (2004) Discovery and directed evolution of a glyphosate tolerance gene. Science 304:1151–1154

Felsenstein J (1974) The evolutionary advantage of recombination. Genetics 78:737–756

Fontana W (2002) Modelling "evo-devo" with RNA. BioEssays 24:1164–1177

Gilbert W (1978) Why genes in pieces. Nature 271:501

Go M (1983) Modular structural units, exons, and function in chicken lysozyme. Proc Natl Acad Sci USA 80:1964–1968

Hadany L, Beker T (2003) Fitness-associated recombination on rugged adaptive landscapes. J Evol Biol 16:862–870

Jäschke A, Helm M (2003) RNA sex. Chem Biol 10:1148–1150

Jetz AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP (2000) High rates of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 74:1234–1240

Joern JM, Meinhold P, Arnold F (2002) Analysis of shuffled gene libraries. J Mol Biol 316:643–656

Judo MSB, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res 26:1819–1825

Kauffman SA, Macready WG (1995) Search strategies for applied molecular evolution. J Theor Biol 173:427–440

Keiper S, Bebenroth D, Seelig B, Westhof E, Jaschke A (2004) Architecture of a Diels-Alderase ribozyme with a preformed catalytic pocket. Chem Biol 11:1217–27

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Krasilnikov AS, Xiao Y, Pan T, Mondragon A (2004) Basis for structural diversity in homologous RNAs. Science 306:62–63

Kurtzman AL, Govindarajan S, Vahle K, Jones JT, Heinrichs V, Patten PA (2001) Advances in directed protein evolution by recursive genetic recombination: applications to therapeutic proteins. Curr Opin Biotechnol 12:361–370

Lehman N (2003) A case for the extreme antiquity of recombination. J Mol Evol 56:770–777

Lehman N, Joyce GF (1993) In vitro evolution: Analysis of a lineage of ribozymes. Curr Biol 3:723–734

Maynard Smith J, Szathmáry E (1995) The major transitions in evolution. WH Freeman and Co., New York

Meyerhans A, Vartanian J-P, Wain-Hobson S (1990) DNA recombination during PCR. Nucleic Acids Res 18:1687–1691

Moore GL, Maranas CD, Lutz S, Benkovic SJ (2001) Predicting crossover generation in DNA shuffling. Proc Natl Acad Sci USA 98:3266–3231

Mörl M, Schmelzer C (1990) Group II intron RNA-catalyzed recombination of RNA in vitro. Nucleic Acids Res 18:6545–6551

Patten PA, Howard RJ, Stemmer WP (1997) Applications of DNA shuffling to pharmaceuticals and vaccines. Curr Opin Biotechnol 8:724–733

Petrounia IP, Arnold FH (2000) Designed evolution of enzymatic properties. Curr Opin Biotechnol 11:325–330

Riley CA, Lehman N (2003) Generalized RNA-directed recombination of RNA. Chem Biol 10:1233–1243

Sabeti PC, Unrau PJ, Bartel DP (1997) Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. Chem Biol 4:767–774

Seelig B, Jäschke A (1999) A small catalytic RNA motif with Diels-Alderase activity. Chem Biol 6:167–176

Stemmer WPC (1994a) DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. Proc Natl Acad Sci USA 91:10747–10751

Stemmer WPC (1994b) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391

Tarasow TM, Tarasow SL, Eaton BE (1997) RNA-catalysed carbon-carbon bond formation. Nature 389:54–57

Tonegawa S, Maxam AM, Tizard R, Bernard O, Gilbert W (1978) Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc Natl Acad Sci USA 75:1485–1489

Uhlenbeck OC (1987) A small catalytic oligoribonucleotide. Nature 328:596–600

Unrau PJ, Bartel DP (1998) RNA-catalysed nucleotide synthesis. Nature 395:260–263

Wang QS, Unrau PJ (2005) Ribozyme motif structure revealed by nonhomologous recombination and evolution. RNA 11:404–411

Westhof E, Massire C (2004) Structural biology. Evolution of RNA architecture. Science 306:62–63

Zaher HS, Unrau PJ (2005) Nucleic acid library construction using synthetic DNA constructs. Methods Mol Biol 288:359–378

Zhao H, Giver L, Shao Z, Affolter A, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 16:258–261