

SHORT REVIEW

The molecular underpinnings of genetic phenomena

N Lehman

Department of Chemistry, Portland State University, Portland, OR, USA

Epiphenomena are those processes that ostensibly have no precedent at lower levels of scientific organization. In this review, it is argued that many genetic processes, including ploidy, dominance, heritability, pleiotropy, epistasis, mutational load and recombination, all are at least analogous to biochemical events that were requisite features of the RNA world. Most, if not all, of these features of contemporary

whole organisms and populations may have their ultimate evolutionary roots in the chemical repertoire of catalytic RNAs. Some of these phenomena will eventually prove to be not only analogous but homologous to ribozyme activities.

Heredity (2008) **100**, 6–12; doi:10.1038/sj.hdy.6801053; published online 29 August 2007

Keywords: RNA; ploidy; pleiotropy; epistasis; heritability; recombination

Introduction

The history of life, if traced with perfect detail, would reveal a long and gradual expansion of networks of chemical reactions. If viewed in less detail, a series of plateaus would be seen, each representing a discrete advancement in complexity (Fontana and Schuster, 1998; Hazen *et al.*, 2007). Some of these steps could be described as emergent properties, or epiphenomena: biological processes that would not be predicted from knowledge of lower-level phenomena. Investigations into epiphenomena, especially epigenetics, are currently very popular. This is true in large part because the identification of a previously unappreciated level of complexity can lead to great excitement and a vast improvement in our understanding of biology. The role of DNA methylation in gene expression and transmission is an excellent example (cf. Robertson and Wolffe, 2000), as we now realize that there are mechanisms of heredity beyond the primary sequence of germline DNA.

Nevertheless, it is important to keep in mind that many genetic phenomena can in fact be traced back to familiar biochemical events. This is true even of many events often described as epiphenomena. The purpose of this review is to survey the wealth of recent data on *in vitro* studies of biopolymer function with the goal of detecting parallels between what molecules can do and what organisms can do. In some of these cases the two actions will eventually turn out not to be homologous, in the sense that they will not share a common evolutionary ancestry. But in some cases they will, and it will be enlightening to explore the generality of the molecular underpinnings of these genetic phenomena.

Below are discussed seven well-known genetic processes for which a clear molecular basis can be postulated. By ‘molecular basis’ it is meant that a chemical property of biopolymers, usually RNA, can be identified as a direct forbearer of the higher-order property observed in whole organisms. Again, an important aspect of this is the idea that many of these molecular events may have been the ultimate causal agents of the organismal events, as opposed to the alternative that the two processes are merely analogous.

Ploidy

One can begin the search for the molecular roots of genetics in an obvious location: the existence of more than one copy of genomic information in the cell, or ploidy. This is traceable to the existence of double-stranded nucleic acids, which in turn has its basis in the complementarity of nitrogenous bases. The discrete patterns of hydrogen-bond donors and acceptors on the Watson–Crick surface lead to a coding process by which information can be copied. This point is not as trivial as it sounds. Under the RNA world hypothesis (Gilbert, 1986), an RNA-like polymer was responsible for both genotypic and phenotypic functions in the most primitive of life forms. At issue for many years has been the mode of replication of such entities. Naturally, much attention has been focused on the concept of an RNA replicase, which would have been a catalytic RNA (ribozyme) capable of making a copy of an RNA strand in much the same way that contemporary DNA polymerases function. However, all known protein polymerases do not make exact copies, they make complementary copies. Thus, as noted by Eigen and Schuster (1977) and others, an RNA replicase would only be functional in a system where there exist + and – strands, requiring not one, but two copies of the genome. This would have been the ultimate origin of ploidy.

Correspondence: Dr N Lehman, Department of Chemistry, Portland State University, PO Box 751, Portland, OR 97207, USA.

E-mail: niles@pdx.edu

Received 29 June 2007; revised 2 August 2007; accepted 3 August 2007; published online 29 August 2007

Search for an RNA replicase in the laboratory has made some progress with *in vitro* evolution (Bartel and Szostak, 1993; Johnston *et al.*, 2001; Zaher and Unrau, 2007), but we still have a long way to go, in part because of the lack of processivity of replicase ribozymes discovered so far (≤ 20 nt), but also in part because a self-replicating RNA replicase would need to copy itself twice (+ \rightarrow - \rightarrow + dashed lines in Figure 1a) and/or be functional in both + and - stranded forms (Joyce and Orgel, 2006). Moreover, such an enzyme would depend on efficient separation of the product strand from the template strand, a chemical event that has a poor compatibility with the conditions under which replication takes place. Inspired by the power of the PCR, some have proposed tidal or thermal cycling as a means to overcome this problem (Lathe, 2003; Braun and Libchaber, 2004). Also, there is some precedence for both strand alternation and dual-strand functionality in the replicative pathways of some viruses, such as the human delta virus (HDV), which undergoes rolling-circle replication involving two complementary single-stranded versions of its genome. Interestingly, this pathway also involves a self-cleaving RNA motif, the HDV ribozyme, which comes in two forms, a genomic and an antigenomic, although these are not direct complements of each other. In any event, the advent of an RNA replicase allowed (or required) two copies of the genome. It is debatable whether this happened at the RNA stage or was concomitant with the first utilization of DNA (Breaker and Joyce, 1994).

Of course, the genomic duplication most commonly associated with ploidy is a subsequent copying of an entire group of nucleic acids into a noncomplementary set. This created a genomic redundancy that allowed evolutionary benefits such as gene divergence, dominance, certain modes of speciation and meiotic replication. Ploidy of this type can arise from a variety of mechanisms such as endosymbiosis, disruptions in the chromosomal segregation process or unequal crossover, discussed in more detail elsewhere (Lynch and Conery, 2000; Wolfe, 2001). Brosius (2003) has pointed out how retrotransposition could have led to many, if not all, of the complex genomes found today. Some of the aforementioned processes can be traced back to DNA splicing and rejoining catalytic activities such as those affiliated with recombination (discussed below), while others cannot, making ploidy a phenomenon with multifaceted molecular underpinnings.

Dominance

Dominance results when the phenotypic expression of one allele is disproportional to other identical-by-descent alleles in cases where the ploidy > 1 . Thus the molecular basis of dominance is tied to that of ploidy, although, because polyploidy does not in itself result in a dominance situation, there must be additional factors. Dominance comes into play when the phenotype of the products of one allele outcompete all others, and, at the finest level, is often manifest by stronger interactions between enzymes and their substrates (Figure 1b). Dominance can be also triggered when one gene product actively suppresses either the expression or the action of others, but again, this ultimately involves a molecular recognition event entailing noncovalent binding.

Turning again to the RNA world, substrate binding is a well-studied phenomenon, both for ribozyme and for RNAs that simply bind their substrates in a highly specific fashion (aptamers). We have been able to measure the binding constants between RNAs and their substrates in hundreds of cases. The K_m (Michaelis constant) or the K_d (dissociation constant) for RNAs can be quite impressive, in some cases plunging into the low μM range (Puglisi and Williamson, 1999; Bunka and Stockley, 2006). RNAs achieve tight and specific binding by employing most of the intermolecular interaction tools at their disposal, including hydrogen bonding, hydrophobic interactions and the use of cationic metal ligands. A classic example is the ATP aptamer originally discovered in the Szostak laboratory through selection from a random pool of RNA sequences (Sassanfar and Szostak, 1993). RNAs that possess a particular 12 or so nucleotides arranged in a highly asymmetrical internal loop can bind their substrate molecule with high affinity and can even discriminate against similar substrates such as GTP. Single mutations can affect the shapes of these loops enough to destroy the binding affinity, such that allelic variations could possess a dominance-recessive relationship if the phenotype were an eventual manifestation of the binding event.

The recent discovery (Winkler *et al.*, 2004) and structural elucidation (Kline and Ferré-D'Amaré, 2006) of the *glmS* riboswitch ribozyme is a good demonstration of how substrate binding by RNAs could directly affect the phenotype of a cell and potentially give rise to a dominance relationship upon polyploidy. The 5' untranslated region of the mRNA of a *Bacillus subtilis* gene that encodes the enzyme glutamine-fructose-6-phosphate amidotransferase folds into a motif that binds an effector molecule, glucosamine-6-phosphate (GlcN6P, a component of cell walls) with extremely high specificity. Upon doing so, the RNA turns into a self-cleaving ribozyme, effectively shutting down protein synthesis and altering the phenotype of the cell. These events are useful when GlcN6P concentrations are high, because the protein's role is to convert a precursor molecule into GlcN6P. A hypothetical gene duplication and subsequent mutation of any one of a dozen or so key nucleotides that disrupted the ability of the RNA to bind (or react with) GlcN6P would lead to a second allele that would be dominant over the wild type, as the enzyme would be produced in heterozygotes. This new allele would be disadvantageous in environments with limited resources as a consequence of the spurious production of the protein in the absence of its substrate, but could be selected for in environments with abundant resources where cell-wall production is in high demand.

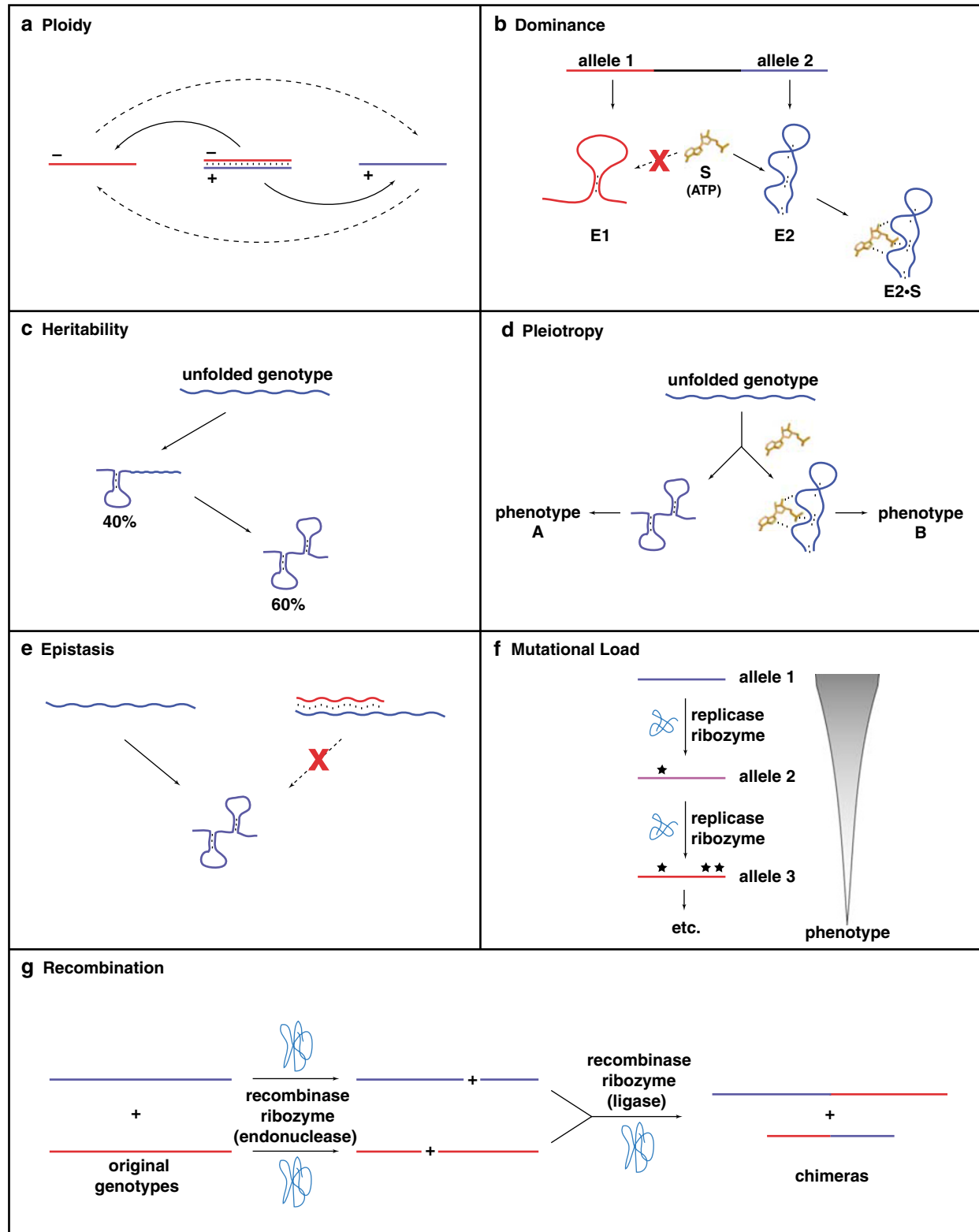
Heritability

Heritability is a measure of the predictability of the phenotype given a genotype (G), and deviations from perfect heritability arise when multiple developmental pathways are possible. Often the environment (E) in which development takes place influences the distribution and relative likelihoods of various phenotypes (G \times E interaction; cf. Lynch and Walsh, 1998 for review). The origin of development can be traced ultimately to the folding of RNA (Figure 1c), which is the finest level at

which we can observe a genotype expressing itself as a phenotype (cf. Ance and Fontana, 2000).

RNA function is dictated by its shape, which in turn is dictated by the base–base interactions that take place during folding. RNA folding is driven by fairly well-characterized thermodynamic processes, dominated by

the entropy of hydrogen-bond formation and the enthalpy of base stacking (for example, Mathews and Turner, 2006). Over evolutionary time, there has been selection pressure to minimize the range of possible folds from a given primary sequence (canalization), but the first RNAs—naked, acellular and without the benefit of a



long adaptive history—would not have been as well behaved. Even today, some RNAs are notoriously poor folders in the test tube (Uhlenbeck, 1995). RNAs can exhibit multiple folded conformations either because they are kinetically slow to fold, and get trapped in metastable intermediates along the correct folding pathway, or because they sometimes embark on incorrect folding pathways leading to thermodynamically favored, but phenotypically impotent folded states. The behavior of the *Tetrahymena* group I self-splicing intron in solution is an example of the former (Pan and Woodson, 1998). It has been noted that as temperatures rise, the folding heterogeneity problem worsens, because smaller changes in free energy values will have more of an impact on the position of the equilibrium between conformations (Moulton *et al.*, 2000).

When RNAs fold variably there is no longer a 1:1 mapping between genotype and phenotype (Stadler and Stadler, 2006). This can retard the rate of evolutionary change by natural selection, just as the additive genetic variance component of total variation affects the response to selection in populations of whole organisms. A less-than-unity heritability value (*h*) equivalent can be seen in ligase ribozymes evolving *in vitro*, and the RNA folding variability under such conditions has been termed ‘molecular heritability’ (Schmitt and Lehman, 1999). In recent years a number of studies have pointed to variables that can heighten the probability that an RNA will fold into only one functional phenotype, and thereby increase its molecular heritability. In addition to lower temperatures, higher ionic strengths of the solution/cell facilitate proper folding by providing an electrostatic shield (Draper, 2004; Koculi *et al.*, 2007). Higher guanine–cytosine (GC) contents also seem to augment unimorphic folding, as evidenced by the high thermostability, excellent X-ray crystal resolution and rapid, cooperative and smooth folding behavior of the *Azoarcus* group I ribozyme, which is 71% GC (Kuo *et al.*, 1999; Rangan *et al.*, 2003; Adams *et al.*, 2004). In contrast, RNAs that are either GC poor, or are comprised of non-standard nucleotides that have tendencies to pair promiscuously, stack poorly or form few and/or extended hydrogen bonds (such as inosine (I) and 2,6-diaminopurine (D)) are less likely to possess robust phenotypes (Szathmáry, 1992; Schuster, 1994) and would have low molecular heritabilities. For example, a ligase ribozyme comprised of only two nucleotides, D and U, was selected *in vitro*, but its conformational heterogeneity contributed to its catalytic rate enhancement being more than 10-fold less than that of its parent ligase, which contained all four standard nucleotides (Reader and Joyce, 2002).

Pleiotropy

Pleiotropy is the phenomenon where a single gene influences multiple phenotypic traits. From a molecular point of view, it should be obvious that pleiotropy traces its roots to the same place as does heritability: alternative folding pathways of RNA (Figure 1d). The difference between the two is that in the case of pleiotropy, we expect that most or all of the folded conformations to be functional, in a positive sense. In other words, subunity heritability implies that the high-fitness phenotype is expressed (‘penetrates’) poorly, while pleiotropy implies that two high-fitness phenotypes are expressed under different conditions, perhaps in alternative environments.

While cases of RNAs that fold imprecisely toward a single functional phenotype are numerous, there are also cases where a single RNA can fold into multiple discrete, but active, shapes. Perhaps the most striking example is the 88 nt ribozyme engineered by Schultes and Bartel (2000). This RNA is the intersection sequence along a hypothetical mutational lineage between two distinct endpoints representing a ligase ribozyme and an HDV ribozyme. Consequently it can assume two very different folding shapes by adopting completely different sets of hydrogen-bonding partners among its nucleotides (different secondary structures). One shape has a modest amount of ligase catalytic activity in that it can enjoin a substrate oligonucleotide to its 5' end, while the other shape has a modest amount of self-cleavage activity. This RNA therefore is pleiotropic. While both activities can be simultaneously expressed at a low level under a given set of environmental conditions, the chemistries of the phenotypes dictate that the ligase activity would be favored at low temperatures while the cleavage activity would be favored at high temperatures.

Epistasis

In some sense, epistasis is the opposite of pleiotropy, in that here two genotypes interact to produce one phenotype, rather than one genotype being responsible for two phenotypes. The molecular basis for epistasis is intermolecular inhibition, and the most primitive form of this would have been the direct binding of one RNA onto another through complementary nucleotides (Figure 1e).

This type of interaction, when it shuts down expression of the target genotype, has been termed ‘antisense’. Antisense oligomers of nucleic acids have been examined as far back as the late 1980s as potential therapeutics such as anti-viral drugs (cf. Zhang *et al.*, 2007 for recent review). In principle, any RNA with a functional secondary structure can be inhibited by flooding the

Figure 1 The molecular underpinnings of genetic phenomena during and after the RNA world. (a) Ploidy results when single-stranded genomes must be replicated via complementary copies creating + and – strands (dashed arrows). At some point a double-stranded genome arose to preserve better the genetic information, and this could give rise to alternative single-stranded versions of the genome directly (solid arrows). (b) Dominance is a consequence of competition among phenotypes (folded RNAs, initially) for the ability to bind exogenous substrates, such as ATP. (c) Heritability is the probability distribution of different phenotypes from the same genotype, and was ultimately influenced by the likelihood that RNAs would complete their folding pathways in a particular environment. (d) Pleiotropy results when one RNA can take alternative folding pathways, each resulting a discrete and valuable phenotype. (e) Epistasis resulted from some RNA molecules binding others and affecting their folding. This can be described as antisense inhibition, and can be intermolecular (pictured) or intramolecular. (f) Mutational loads came with the advent of replicase ribozymes with finite (and likely high) mutation rates. Because most mutations were deleterious, the fitness would suffer each generation without some mechanism to restore activity. (g) Recombination is the swapping of large pieces of nucleic acids from two (or more) different sources and required ribozymes that had two chemical activities, strand scission and strand ligation. Recombination can proceed by a homologous mechanism (equal crossover) or, more likely, by a heterologous mechanism (pictured).

system with a high concentration of a perfectly complementary oligo to an essential secondary structure element. It may also be possible to block translation of structureless mRNAs by the same mechanism. Recent practical advances in antisense therapies have included the use of chemically modified nucleic acids, which can lengthen their lifetimes before they reach their targets (for example, 2'-O-methyl RNAs) or increase their target binding affinities (for example, 'locked' nucleic acids). More intricate ways to exploit antisense include the targeting of ribozymes or deoxyribozymes toward complementary targets, as these molecules actually have the potential to cleave the phosphoester backbones of their targets.

In the same vein, the most impressive parallel between the epistasis and antisense concepts is RNA interference. Recent reviews of this topic are many (for example, Peterson *et al.*, 2006), and include the description by Kim and Rossi (2007) of how genes can be silenced by small interfering RNAs (siRNAs), which initiate the silencing pathway via an antisense interaction between themselves (~21 nt) and target RNA sequences. Although RNA interference is ultimately manifest with the aid of protein RNase enzymes such as dicer, the process is ultimately dependent on a very specific base-pairing interaction between a short and a long RNA. It is becoming clear that regulation of gene expression by siRNAs may have a central role in biology and a correspondingly long evolutionary history—one that probably traces back to RNA–RNA interactions in the RNA world as the first instances of epistasis.

Mutational load

The accumulation of slightly deleterious mutations in a genome, and the consequent mutational load is a clear case of where the proximal cause today and the ultimate cause in the RNA world have similar sources: the error rates of polymerases responsible for replicating the genomes. Today, these mutations are typically the result of a finite probability that DNA polymerase will misincorporate a nucleotide during template-directed polymerization, a value that ranges between 10^{-3} and 10^{-10} mistakes per nucleotide per replication event (Kunkel and Bebenek, 2000). Spontaneous chemical changes in germline DNA, such as cytosine deamination, can also contribute to mutations even with perfect polymerase fidelity.

In the RNA world, the culprit would have been the RNA replicase ribozyme as described above (Figure 1f). The first ribozyme with replicase activity developed in the laboratory had a fidelity of about 96.7% (Johnston *et al.*, 2001). Though quite impressive for our initial attempt to get one RNA to replicate another, this accuracy in a replicase operating in the absence of proteins would have failed to maintain the information content of any genomes it was responsible for replicating. This is a consequence of Eigen's error threshold, which posits that error rates must be progressively lower to maintain the evolutionarily acquired fitness of progressively larger genomes (cf. Eigen, 2002 for a recent review). The error rate of the replicase selected by Johnston *et al.* (2001) was about three times too high to maintain a genome consisting only of the replicase (165 nt). A recent significant improvement on this replicase selected by Zaher and Unrau (2007) is not only more

processive but somewhat more accurate, such that a laboratory demonstration of a sufficiently accurate RNA replicase ribozyme to get below the threshold is certainly foreseeable.

Eigen's error threshold is an alternative mathematical description of a process well known to population biologists: the synergy of the accumulation of deleterious mutations and the drop in population size leading to a 'mutational meltdown' (Lynch *et al.*, 1993). RNA populations in the precellular world would have been at risk not only for loss of information if their replication was not accurate enough, but also for outright extinction if their population sizes dropped too low. Recently empirical evidence shows that small RNA populations evolving in a test tube (with the aid of protein polymerases) can quickly go extinct through mutational accumulation, but that larger population sizes and beneficial mutations can help protect them (Soll *et al.*, 2007). It is apparent that for a number of reasons early selection pressure for improved replicase fidelity—by ribozymes or proteins—would have been severe.

Recombination

Recombination is the swapping of large blocks of contiguous nucleic acids between two sources to enhance genetic diversity, to repair damaged information or to assimilate exogenous sequences for a variety of uses. Historically, recombination is most likely an ancient phenomenon, predating its higher-level manifestation in sexual reproduction by billions of years (Cavalier-Smith, 2002; Lehman, 2003). Recombination owes its antiquity to the fact that from a chemical viewpoint, this process is the breaking and subsequent rapid reforming of phosphodiester bonds (Figure 1g), events that are not only of great utility to nascent life, but ones that can occur with little or no energy input. Unlike template-directed polymerization, recombination can occur with a free-energy change of essentially zero, and recombinase enzymes can facilitate this process mainly by bringing the free 5' and 3' ends of nucleic acids into close proximity.

Recombination of RNA strands in a preprotein environment may have been necessary to ameliorate the mutational load imposed by sloppy replicases. Ribozymes are clearly capable of RNA recombination, which is not surprising because the most common activities of known catalytic RNAs from natural sources involve RNA strand scission and/or ligation. The self-splicing of group I introns out of primary rRNA or tRNA transcripts is effectively a recombination event when all of the product RNA sequences are considered. In fact, our group has engineered group I introns to be general RNA recombinases, and they can even self-assemble from smaller RNA fragments through recombination events, suggesting a key role for recombination in the origins of life (Riley and Lehman, 2003; Hayden and Lehman, 2006). RNA breakage and religation has proven useful to add or remove specific blocks of sequences from molecules such as defective mRNA genes (Watanabe and Sullenger, 2000; Johnson *et al.*, 2005).

Once genomes evolved to utilize primarily DNA, protein recombinases would have taken over the breakage and religation functions from ribozymes, but many of the fingerprints of the primitive mechanisms exist. Spliceosomal-mediated mRNA intron removal is still

highly dependent on RNAs for (at least) substrate alignment, and the chemistry of the reaction bears a strong resemblance to that of group II self-splicing introns. 'RNA-free' protein enzymes such as RecA and those with key serine (for example, $\gamma\delta$ resolvase) and tyrosine (for example, Cre recombinase) catalytic residues, use similar strand-cleavage mechanisms, but do not require energy input such as ATP (Sarkis *et al.*, 2001). Although these proteins have lost the nucleotide–nucleotide complementarity feature that self-splicing RNA introns use to guide precise splice-site choice, high-resolution structural data such as that for the $\gamma\delta$ resolvase (Li *et al.*, 2005) are showing us how such proteins, as do restriction endonucleases, depend on short recombination signals in the DNA that are recognized by a restricted set of amino acid–nucleotide hydrogen-bonding interactions.

Summary

Ignoring for a moment the perils of blatant reductionism, we can begin to draw biochemically based parallels between molecular and organismal genetic phenomena. These comparisons should help us look for patterns in nature that will lead to the discovery of shared evolutionary provenances. With ever improving phylogenetic reconstruction tools, we hope be able to trace critical genetic phenomena back to their roots in the putative RNA world. Some of the connections described here, such as mutational load and recombination, will probably turn out to be ultimately based on homology. Others, such as ploidy and epistasis, may have been reinvented many times and will thus be based on analogy.

Acknowledgements

We thank A Burton and E Hayden for useful discussions in the preparation of this manuscript and the NSF for a grant (DEB-0315286) for financial support.

References

Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**: 45–50.

Ancel LW, Fontana W (2000). Plasticity, evolvability, and modularity in RNA. *J Experim Zool* **288**: 242–283.

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

Braun D, Libchaber A (2004). Thermal force approach to molecular evolution. *Phys Biol* **1**: P1–P8.

Breaker RR, Joyce GF (1994). A DNA enzyme that cleaves RNA. *Chem Biol* **1**: 223–229.

Brosius J (2003). Gene duplication and other evolutionary strategies: from the RNA world to the future. *J Struct Funct Genomics* **3**: 1–17.

Bunka DHJ, Stockley PG (2006). Aptamers come of age—at last. *Nat Rev Microbiol* **4**: 588–596.

Cavaliere-Smith T (2002). Origins of the machinery of recombination and sex. *Heredity* **88**: 125–141.

Draper DE (2004). A guide to ions and RNA structure. *RNA* **10**: 335–343.

Eigen M (2002). Error catastrophe and antiviral strategy. *Proc Natl Acad Sci USA* **99**: 13374–13376.

Eigen M, Schuster P (1977). The hypercycle. *Naturwissenschaften* **64**: 541–565.

Fontana W, Schuster P (1998). Continuity in evolution: on the nature of transitions. *Science* **280**: 1451–1455.

Gilbert W (1986). The RNA world. *Nature* **319**: 618.

Hayden EJ, Lehman N (2006). Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol* **13**: 909–918.

Hazen RM, Griffen PL, Carothers JM, Szostak JW (2007). Functional information and the emergence of biocomplexity. *Proc Natl Acad Sci USA* **104**: 8574–8581.

Johnson AK, Sinha J, Testa SM (2005). *Trans* insertion-splicing: ribozyme-catalyzed insertion of targeted sequences into RNAs. *Biochemistry* **44**: 10702–10710.

Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**: 1319–1325.

Joyce GF, Orgel LE (2006). Progress toward understanding the origin of the RNA world. In: Gesteland RF, Cech TR, Atkins JF (eds). *The RNA World*, 3rd edn. Cold Spring Harbor Press: Cold Spring Harbor. pp 23–56.

Kim DH, Rossi JJ (2007). Strategies for silencing human disease using RNA interference. *Nat Rev Genet* **8**: 173–184.

Kline DJ, Ferré-D'Amaré AR (2006). Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* **313**: 1752–1756.

Koculi E, Hyeon C, Thirumalai D, Woodson SA (2007). Charge density of divalent metal cations determines RNA stability. *J Am Chem Soc* **129**: 2676–2682.

Kunkel TA, Bebenek K (2000). DNA replication fidelity. *Annu Rev Biochem* **69**: 497–529.

Kuo LY, Davidson LA, Pico S (1999). Characterization of the *Azoarcus* ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme. *Biochim Biophys Acta* **1489**: 281–292.

Lathe R (2003). Fast tidal cycling and the origin of life. *Icarus* **168**: 18–22.

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

Li W, Kamtekar S, Xiong Y, Sarkis GJ, Grindley NDF, Steitz TA (2005). Structure of a synaptic $\gamma\delta$ resolvase tetramer covalently linked to two cleaved DNAs. *Science* **309**: 1210–1215.

Lynch M, Burger R, Butcher D, Gabriel W (1993). The mutational meltdown in asexual populations. *J Hered* **84**: 339–344.

Lynch M, Conery JS (2000). The evolutionary fate and consequences of duplicated genes. *Science* **290**: 1151–1155.

Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer: Sunderland, MA.

Mathews DH, Turner DH (2006). Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol* **16**: 270–278.

Moulton V, Gardner PP, Pointon RF, Creamer LK, Jameson GB, Penny D (2000). RNA folding argues against a hot-start origin of life. *J Mol Evol* **51**: 416–421.

Pan J, Woodson SA (1998). Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J Mol Biol* **280**: 597–609.

Peterson CP, Doench JG, Grishok A, Sharp PA (2006). The biology of short RNAs. In: Gesteland RF, Cech TR, Atkins JF (eds). *The RNA World*, 3rd edn. Cold Spring Harbor Press: Cold Spring Harbor. pp 535–565.

Puglisi JD, Williamson JR (1999). RNA interaction with small ligands and peptides. In: Gesteland RF, Cech TR, Atkins JF (eds). *The RNA World*, 2nd edn. Cold Spring Harbor Press: Cold Spring Harbor. pp 403–425.

Rangan P, Masquida B, Westhof E, Woodson SA (2003). Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc Natl Acad Sci USA* **100**: 1574–1579.

Reader JS, Joyce GF (2002). A ribozyme composed of only two different nucleotides. *Nature* **420**: 841–844.

- Riley CA, Lehman N (2003). Generalized RNA-directed recombination of RNA. *Chem Biol* **10**: 1233–1243.
- Robertson KD, Wolffe AP (2000). DNA methylation in health and disease. *Nat Rev Genet* **1**: 11–19.
- Sarkis GJ, Murley LL, Leschziner AE, Boocock MR, Stark WM, Grindley ND (2001). A model for the $\gamma\delta$ resolvase synaptic complex. *Mol Cell* **8**: 623–631.
- Sassanfar M, Szostak JW (1993). An RNA motif that binds ATP. *Nature* **364**: 550–553.
- Schmitt T, Lehman N (1999). Non-unity molecular heritability demonstrated by continuous evolution *in vitro*. *Chem Biol* **6**: 857–869.
- Schultes EA, Bartel DP (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**: 448–452.
- Schuster P (1994). Extended molecular evolutionary biology: artificial life bridging the gap between chemistry and biology. *Artif Life* **1**: 39–60.
- Soll SJ, Diaz Arenas C, Lehman N (2007). Accumulation of deleterious mutations in abiotic populations of RNA. *Genetics* **175**: 267–275.
- Stadler PF, Stadler BMR (2006). Genotype–phenotype maps. *Biol Theory* **1**: 268–279.
- Szathmáry E (1992). What is the optimum size for the genetic alphabet? *Proc Natl Acad Sci USA* **89**: 2614–2618.
- Uhlenbeck OC (1995). Keeping RNA happy. *RNA* **1**: 1–4.
- Watanabe T, Sullenger BA (2000). Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proc Natl Acad Sci USA* **97**: 8490–8494.
- Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004). Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**: 281–286.
- Wolfe KH (2001). Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* **2**: 333–341.
- Zaher HS, Unrau PJ (2007). Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**: 1017–1026.
- Zhang C, Pei J, Kumar D, Sakabe I, Boudreau HE, Gokhale PC *et al.* (2007). Antisense oligonucleotides: target validation and development of systemically delivered therapeutic nanoparticles. *Methods Mol Biol* **361**: 163–185.