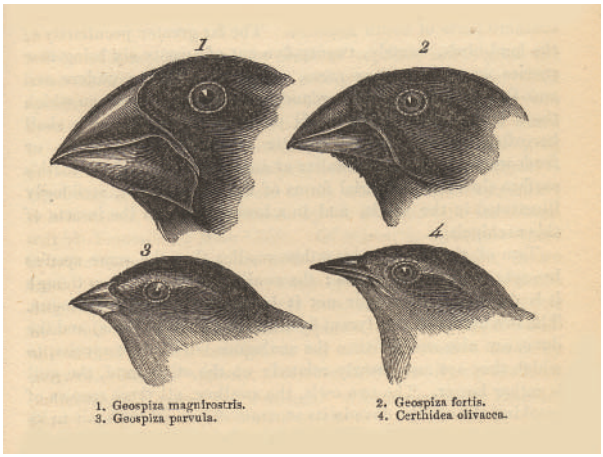
The onset of selection

Natural selection started to drive evolution as soon as molecular replication became possible.

Christian de Duve

Divined by the genius of Alfred Russell Wallace and Charles Darwin, the basic principles of evolution by natural selection are well known. First, there is genetic continuity, based on replication. Then, inevitably, comes variation. Finally, there is competition, leading to selection of the variants most apt to survive and proliferate under the prevailing conditions. The findings of modern biology have fully validated those principles, adding the fundamental fact that the causes of variation are strictly accidental and unintentional.

The key notion in this theory is replication. The rest follows obligatorily. Thus, in the origin of life, darwinian evolution must have started as soon as the first replicable molecules appeared. Here, I intend to draw attention to certain implications of this fact that are sometimes ignored or underestimated in discussions of the origin of life. I shall assume, in agreement with most workers in the field, that the first replicable molecules consisted of RNA.



The Galapagos finches inspired Darwin's theory of natural selection.

Direct versus indirect selection

The simplest manifestation of natural selection is the direct, molecular form — the object of many studies since it was first made to occur in the test tube by Sol Spiegelman in the 1960s. If RNA molecules are allowed to replicate *in vitro*, selection automatically screens out those mutant molecules that best combine stability and replicability — the molecular equivalent of darwinian survival and proliferation — under the adopted conditions.

By necessity, this kind of selection must have started with replication. In fact, the first product of molecular selection may well have been RNA itself. The mechanism whereby this substance arose is still unknown, but cannot possibly, unless guided by some prescient agency, have produced only authentic RNA molecules with the bases A, U, G and C as sole constituents. It is much more likely that such molecules were accompanied by other analogous assemblages and that they were selected out of this mixture and amplified, thanks to their ability to induce, by base pairing, the formation of complementary molecules that could in turn act similarly to reproduce the first ones.

Once initiated, such a process would have evolved naturally toward the production of what Manfred Eigen has called the Ur-Gen, the ancestor of all RNA molecules. This product would have arisen by molecular selection to form a 'quasi-species', consisting of a 'master sequence', optimized with respect to the prevailing environment, and of an ever-changing cohort of mutants arising through replication errors and other accidents.

The RNA molecules that initiated protein synthesis probably belonged to this early crop. It is widely believed that protein synthesis was launched by interactions between RNA and amino acid molecules, prefiguring the present role of transfer RNAs as carriers of the amino acids that are incorporated into proteins. Whereas the amino acids were present beforehand, the RNA partners of the primeval associations most likely arose as mutants of the Ur-Gen and were subsequently selected. This could have happened by a molecular mechanism. An RNA molecule bearing an amino acid could have adopted a more stable conformation. More importantly, for a long-lasting effect, it could have interacted more efficiently with the replication catalyst, thus furthering its own replication.

Other RNA molecules presumably also participated in the development of protein synthesis, for example by favouring the proper alignment of RNAs bearing amino acids, or by catalysing peptide bond formation, functions fulfilled today by messenger RNAs and ribosomal RNAs, respectively. Selection of such RNAs by a molecular mechanism cannot be ruled out, but is not readily visualized. In any case, a stage in the development of protein synthesis must have been reached where direct selection ceased to be the sole operating process and a new, indirect form of selection was initiated because of the growing complexity of the system.

In this form of selection — which dominates darwinian evolution — genes are selected not because of what they are, but because of what they or their products do, which in the beginning must have been mainly to catalyse a chemical reaction useful to the gene's replication. Barring rare exceptions (such as self-replication), the criterion of usefulness requires the reproduction of the gene to be linked to that of an entity that derives an advantage from the new reaction allowed by the gene. This condition almost mandates the existence of primitive cells, or 'protocells', able to grow, multiply and compete with other protocells for available resources.

It follows that cellularization must have occurred very early in the development of life, probably no later than the inception of protein synthesis. Most of the catalytic RNAs (ribozymes) involved in the so-called 'RNA world' and all the first protein enzymes must have been 'invented' by protocells capable of participating in darwinian competition and of deriving a selective advantage from the catalyst. An important implication of this conclusion is that the early protocells must have been sufficiently individualized to be able to engage in competition and benefit from selection. This point is relevant to the proposal, advocated by Carl Woese, W. Ford Doolittle and others, that the first cells, up to the last universal common ancestor (LUCA) of all known living beings, formed a collective of ill-defined entities that freely exchanged and shared genes by horizontal transfer. This phenomenon may, indeed, have been more important in the early days of the development of life than it is today — which is not inconsiderable — but it cannot have been so extensive and frequent as to blur the distinctions between individual lineages, suppress competition and impede selection.

Mutual selection

Selection is usually visualized as a one-way process, in which a shifting collection of evolving entities is subject to screening by the environment. But the process is often mutual, with the environment being itself screened by the evolving entities. This reciprocity is asymmetric, involving an active and a passive

R. T. PRITCHETT IN *JOURNAL OF RESEARCHES*, C. DARWIN, 1845

partner. The active partner is a replicable entity subject to genetic variation and natural selection, whereas the passive partner is no more than a component of an existing pool, from which it is simply singled out.

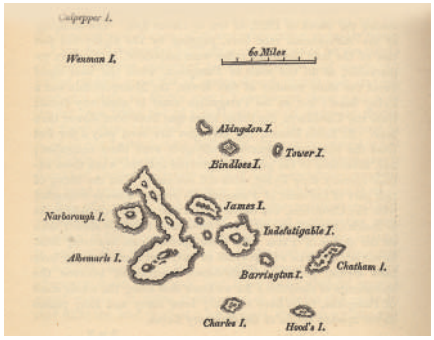
Illustrating such an occurrence is the mechanism previously outlined, whereby certain RNA molecules (the active partner) may have been selected by virtue of their ability to combine with amino acids (the passive partner). If, as seems likely, there was any chemical specificity in the interactions involved, the selected RNA molecules must have acted reciprocally to recruit out of the prebiotic pool the first amino acid molecules that came to be used for protein synthesis — thereby launching the selection of the amino acids, including their chiral forms, that universally serve for the construction of proteins.

The development of metabolism offers a second possible instance of mutual selection. Here, the active partners were either RNA ribozymes or RNA-encoded protein enzymes. The passive partners belonged to the pre-existing chemistry (protometabolism), which must have supplied the catalysts with one or more substrates to act on and with one or more outlets for the product(s) of the catalysed reactions — otherwise, the catalysts could not have been selected. Without substrate or outlet, a catalyst cannot be useful; it may even be harmful. This relationship, which is rarely appreciated, implies that some of the key features of present-day metabolism must have been prefigured in protometabolism; the two were congruent. A corollary is that the catalysts acted reciprocally to select out of protometabolism reactions (or, at least reactants) that came to be taken up into metabolism. In other words, certain components of early metabolism were selected from what was perforce a 'dirty' protometabolism by the catalysts they themselves served to select.

Optimizing selection

A common opinion is that the course of evolution, being dependent on chance occurrences, is utterly contingent, unpredictable and non-reproducible. This is not necessarily so; chance does not exclude inevitability. It all depends on the number of opportunities provided for an event to happen, relative to the event's probability. Even a seven-digit lottery number has a 99.9% likelihood of coming out if 69 million draws are held. Lotteries don't operate that way, of course, but the evolutionary lottery often does, thanks to an extensive, if not exhaustive, exploration of the possibilities open to a system.

An easily quantifiable case is represented by the sequence spaces of nucleic acids and proteins, that is, the number of distinct sequences of a given length that are possible. A simple calculation shows that, for molecules of the size found in living organisms today, the sequence spaces are unimaginably immense, and the region occupied within the spaces by biomolecules is vanishingly small. In the eyes of the small but vocal minority of scientists who defend 'intelligent design', life could not have reached this position by chance alone; it must have been 'guided'. Another less controversial view is that the sequence spaces contain enough life-compatible 'islands' for one such island to be reached with a fair degree of probability by a random walk, but by a pathway that is utterly contingent, unpredictable and non-reproducible, in agreement with the prevailing doctrine.



The Galapagos Islands, as drawn in 1835.

Such views rest on false premises. Life did not start with the kind of large nucleic-acid and protein molecules it uses today. Molecules of that size (and even much smaller ones) would, as demonstrated by Eigen, inevitably have degenerated upon repeated replication, due to the cumulative errors caused by the necessarily low fidelity of the primitive RNA-replicating systems. For this reason, the first replicable RNA molecules and their protein translation products must have been very short, and their subsequent lengthening — probably by splicing of pre-existing RNA stretches — was imperatively dependent on the development of RNA-replicating systems of greater fidelity. Protein molecules thus most likely reached their present size by a stepwise process, 'bootstrapped' at each stage by some improvement in the fidelity of RNA replication, presumably made possible by the appearance of longer, more complex RNA or protein molecules to serve as catalysts.

An important implication of the necessarily stepwise course of RNA and protein lengthening is that exploration of the corresponding sequence spaces was itself stepwise. It may have started with a small enough number of possibilities (short enough stretches) to allow selection to produce the molecules best adapted to the prevailing conditions, by way of the protocells. The number of such molecules could, in turn, have been small enough to allow selective optimization at the next step, and so on. In other words, selection could, at each size level, have reduced the number of molecules available for splicing down to a value compatible with exhaustive or near-exhaustive exploration of combinatorial possibilities, thus approaching optimization at each stage. It follows that nascent life could have reached the infinitesimally minute area it occupies in sequence spaces by a succession of selective bottlenecks, so that the pathway ended up being close to obligatory, at least in its main lines, under the prevailing environmental conditions.

Another instance of optimization may be represented by the genetic code. According to computer-simulation experiments by Stephen Freeland and colleagues, the code seems to be to be more efficient than average — by approximately six orders of magnitude — in minimizing the changes in hydrophobicity, and thus presumably, the harmful consequences resulting from point mutations leading to the replacement of one amino acid by another. If this quality of the genetic code should indeed be a product of natural selection, it would represent a particularly impressive case of optimizing selection, as it implies that primitive protocells had the opportunity to 'experiment' with a large number of distinct genetic codes, the selective criterion being the progeny's long-term ability to withstand the consequences of point mutations. The conditions that would have allowed this kind of experimentation raise challenging questions.

Even later evolution could have enjoyed less freedom than is often claimed, as indicated by phenomena such as drug resistance, mimicry and convergent evolution. It seems that, in many cases, evolving systems had the opportunity to put to the test a vast enough array of relevant mutations to evolve near-optimal, reproducible solutions to environmentally imposed survival problems.

In conclusion, darwinian selection started operating very early in the origin of life and probably played a major role in the shaping of the first living cells.

Christian de Duve is at the ICP, B-1200 Brussels, Belgium, and at the Rockefeller University, New York, New York 10021, USA.

R. T. PRITCHETT IN *JOURNAL OF RESEARCHES*, C. DARWIN, 1845

FURTHER READING

de Duve, C. *Life Evolving* (Oxford Univ. Press, 2002).
de Duve, C. *Singularities* (Cambridge Univ. Press, 2005).
Eigen, M. & Winkler-Oswatitsch, R. *Naturwissenschaften* **68,** 282–292 (1981).
Eigen, M. & Schuster, P. *Naturwissenschaften* **64,** 541–565 (1977).
Woese, C. R. *Proc. Natl Acad. Sci. USA* **95,** 6854–6859 (1998).
Doolittle, W. *Curr. Opin. Struct. Biol.* **10,** 355–358 (2000).
Freeland, S. J. *et al. Orig. Life Evol. Biosph.* **33,** 457–477 (2003).
Conway Morris, S. *Life's Solution* (Cambridge Univ. Press, 2003).

Acknowledgements
I am greatly indebted to Gerald Joyce for valuable comments on an early draft of this esay. Neil Patterson has contributed insightful editorial suggestions.