Why Dawkins' weasel demonstrates mutations cannot produce a new functional gene

Jerry Bergman

Leading evolutionary biologist Richard Dawkins developed a computer simulation, called the 'weasel analogy', that has been used by many of his colleagues to argue in favour of molecules-to-man evolution. An examination of his weasel argument shows that it utterly fails to support the view that mutations can produce significant new genetic information. It fails for several reasons, including that it not only is extremely unrealistic, but also actually shows why Darwinism is impossible. Among the problems with the weasel analogy include the fact that a typical gene is far more complex than the example implies, and must be compatible with the rest of the genes for the organism to function.

From a line in Hamlet, Richard Dawkins borrowed the name, 'weasel analogy', for a computer program that he argues provides good evidence for atoms-to-humans evolution.[1] Dawkins chose the characters of his starting string of letters at random, using only capital letters to simplify things. He then copied his starting string 100 times and one, and only one, letter in each string was 'mutated'. The closest string to the target sequence METHINKS IT IS LIKE A WEASEL was selected and 'bred from' for the next generation.[2] The computer's choice of the best match was meant to illustrate natural selection. It required from 40 to 65 generations to produce the target set. This example was meant to simulate a living organism evolving.[3]

Dawkins' example is widely cited in the literature and some evolutionists argue that his computer analogies are clear evidence for Darwinism. Raymo argued that computer modeling is a reliable method to demonstrate evolution and

> 'as Dawkins suggests, personal incredulity is not a reliable guide to truth. What seemed unlikely to Darwin, and seems impossible to creationists, has been shown to be quite reasonable by high-speed computer modeling. Not only reasonable, but given the proven premises of random mutations and natural selection, virtually inevitable. Will successful computer simulations make any difference to creationist True Believers? Not likely.'[4]

An examination of Dawkins' weasel argument shows that it utterly fails to support the view that mutations can produce significant new genetic information. It fails because the example not only is extremely unrealistic and irrelevant to the real world, but also actually shows why Darwinism is impossible.[5] A major problem, Spetner notes, is that there are no lethal mutations in the model and every number can survive and 'reproduce'. As a result, there is no limit to beneficial changes. In Dawkins' model every change is in essence adaptive enough to be selected for the next generation, which represents a totally unrealistic set of circumstances.

Another of the many objections to the weasel analogy is that it assumes that all possible single base pair substitution, mutational changes in each base pair have close to equal probability. In fact, this is not the case: certain changes are far more likely than others.

Mutation hot spots

Studies of millions of mutations have been completed, and it is quite clear that mutations rarely occur in many or most gene base pairs, even though in some areas of the genome they are much more common.[6] Mutation studies have revealed that in some cases a large percentage of the known mutations are found in only a few codon numbers. For example, one study of human germline *p53* (a tumour suppressor gene) mutations found that of the approximately 400 codons mapped, only 37 mutations were at sites other than codon numbers 175, 245, 248 and 273.[7] The same results are also found in non-cancer genes, demonstrating that selection does not account for these hot spots in many (if not most) cases. One of the best-known mutation hot spots is the CG dinucleotide, in which mutations are about 12 times more common then in other dinucleotide sequences.

If a single nucleotide mutation tended to go, even only slightly, one way more frequently than any other mutation types, such as from C, G or T to A, instead of producing the sentence 'methinks it is like a weasel' Dawkins' example eventually would produce a series of As. Furthermore, only four letters exist in genes—not 26 as in Dawkins' example. If certain errors are more common, entropy (the degeneration of the code) increases more rapidly with four letters than if 26 letters are used. One reason is that the triplet code is partially resistant to mutation due to the degeneracy in the genetic code, because if a mutation occurs at the third base of a codon, the new code will often specify the same amino acid (eg. GGA, GGC, GGG, GGU all code for Glycine). Selection acting on the genome would in these cases not change either the process or the outcome.

Another type of 'hot spot' is found in DNA sequences that contain repetitive or short repeated sequences. The

same is true when editors scan a manuscript, spelling errors in words with multiple letters such as 'addresses', or 'assesses', often are missed, and misspellings such as 'accessses' or 'asssessors' are allowed to slip by.[8] One physical basis for the increased incidence of repeat DNA sequence mutations is the fact that symmetrical or repeated sequences allow base pairing to occur within the same strand when local DNA strands unwind to permit replication. This condition can interfere with both replication and repair enzymes, thereby increasing the chances of errors.

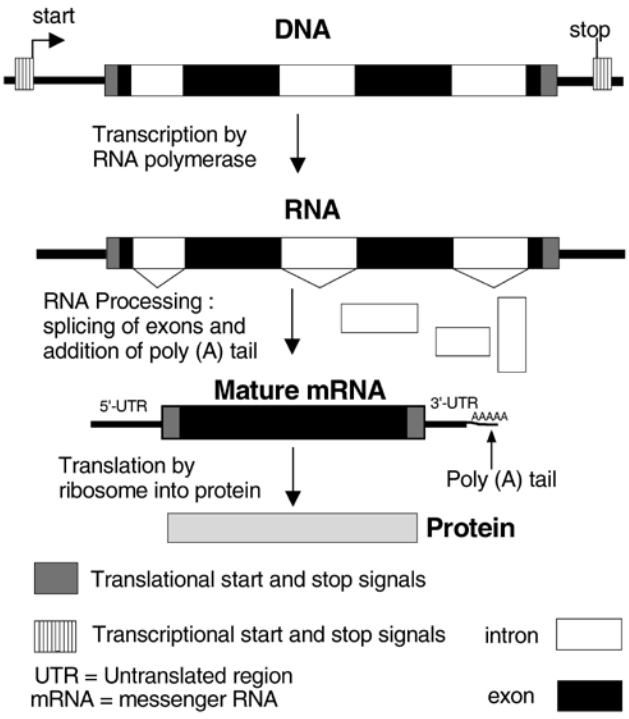
An example is the clotting factor IX gene, mutations in which cause the hemophilia B blood clotting disorder. Mutations in this gene occur up to 100 times more often at the 11 sites in the gene than compared to other sites having long CG dinucleotide repeats. Similarly, an inherited form of the bone-weakening condition, osteoporosis, usually is caused by an extra T that is inserted into a specific stretch of three Ts in the normal gene. The result is a tendency to produce nucleotide stuttering at this hot spot, or, in Dawkins' example, to convert weasel into weasssel.

In addition, Dawkins' portrayal of the genome is grossly oversimplified. To begin with, there is a huge disparity in size. Dawkins uses only a total of 23 letters, and the **smallest** genes are far larger than this. A relatively small gene such as somatostatin has 1,480 bp (base pairs) and the collagen gene is 18,000 bp long. Weinberg claims that the **average** human gene is 'several tens of thousands of DNA bases'.[9] The largest genes are up to 2,000,000 bp, such as the Duchene Muscular Dystrophy gene. The large size of many genes means one wrong mutation can wipe out hundreds or thousands of 'correct' mutations (particularly in the case of an insertion or deletion resulting in a frameshift mutation). Furthermore, the commonly occurring back mutations also often appear in the mutation hot spot areas. As a result, a few regions of the genome mutate back and forth with a tendency to accumulate certain bases.

Another problem for Dawkins is that promoter mutations, capping/polyadenylation mutations, chain termination mutations, mRNA splicing mutations, deletion mutations, insertion mutations, dynamic (triplet expansion) mutations and post-translation mutations are all at least as common as the base pair substitution mutations that Dawkins uses in his example. In virtually all cases, these mutations would wipe out Dawkins' putative evolving gene.

The dynamics of mutations is an area in which I have worked for over a decade. I have reviewed the literature fairly extensively, and have interviewed a number of other scientists who have spent much of their careers working in an area concerned with mutations (often primarily in cancer research). I have yet to locate one single verified example of a mutation that both **adds** information to the genome and produces an improvement which evolution requires.

The most commonly cited example is a person who is heterozygous for sickle cell disease (having both a normal and a sickle-type mutated hemoglobin gene, a condition in which the person is called a carrier) and becomes infected



Only portions of a eukaryotic gene code for a protein product (from Walkup).[32]

with malaria. In this situation, the defective blood cells prevent malarial pathogens from thriving in the victim as they ordinarily would, but they do not interfere with the oxygen-carrying capacity enough to cause serious problems in normal situations. This loss produces a damaged gene that is advantageous only in **one** situation, malarial infections. The same is true of antibiotic resistance, mutations which compromise gene expression (i.e. protein synthesis), yet allows survival in the face of antibiotics.

It is remotely possible that some beneficial mutations that have added information to the genome have occurred in history, and someday one or more might be documented. But to suppose, as Dawkins does, that the **entire** genome in **all** of the different **animal species** (estimated at up to **100 million**, each one of which contains an enormous amount of variety—e.g. an estimated 34,000 genes in humans) ultimately is the result of mutations is illogical. This 100 million number must be multiplied by a factor of 1,000 or more to produce the enormous intraspecies variety found in most species (such as the dog kind, for example).

A gene produced by mutations will not be selected unless it is advantageous, which would not be the case until a **complete** functional gene is produced. Actually, except for genes on the X chromosome, a **pair** of functional genes must evolve. A recessive mutation would be neutralized by the normal allele. However, even if the mutation occurs only once, and then is passed on to numerous descendants, it may be silent until mating occurs with an organism that also has the same mutated gene. Only then will the now recessive characteristic show itself.[10] This event, consider-

ing how rare beneficial mutations are, is very unlikely.

If only one gene mutates, it can cause the organism problems for many reasons. For example, *p53* activates the transcription of several other proteins, and consequently even if only one defective *p53* protein is incorporated into the transcription tetrad required for transcription, the defective tetrad will prevent the transcription factor from functioning and will shut down the entire *p53* tumor suppressive system.[11]

Regulatory genes

Given these basic problems, the Dawkins' weasel example is actually a good illustration of why mutations **cannot** explain macroevolution. Dawkins' example also totally ignores the necessity for simultaneously evolving the many regulating factors involved in genetic transcription. Among these are promoters, including GpC islands and TATA boxes, and various additional upstream regulators such as enhancers, silencers, suppressors, and other controlling elements that help to regulate the amount of a gene product produced. Regulatory genes control what the cell does, when it does it, and how it does it—usually by encoding a protein that activates or inactivates specific genes.[12] Examples include tumor suppressor genes such as *p53* and proto-oncogenes such as *fos*, both of which are transcription factors.[13]

For a gene to be functional normally requires evolution of the gene and **all** of the required transcription and regulation machinery. The exception would be if a functional gene mutated and the altered gene was compatible with the existing transcription and regulation machinery. This assumes that the transcription and regulation machinery had previously evolved and was functional. Obviously, a gene would be useless to an organism until it also evolved the many necessary regulation factors, and then it would be useless (or worse) until the protein, mRNA, tRNA, or rRNA it produced was functional. In the meantime, if it possessed the regulation factors but was not functional, the proteins it produced would have a good likelihood of being toxic or damaging to the cell in some other way.

The correct promoter set must be built in with each gene to insure the proper amount of protein is produced. Both too much and too little as well as improper regulation can be lethal or cause disease. Promoters can both up-regulate (produce more) and down-regulate (produce less) of the protein, and can achieve this feat in complex ways. For example, the cholesterol ester hydrolase sequence position between nucleotide 1190 and nucleotide 1159 inhibits promoter activity in the presence of *mevalonate*. Conversely, promoter activity is partially restored by *squalestatin*. This example illustrates how genes are regulated according to various relevant factors.

Another type of control protein used in an estimated 53% of all genes is what are known as GpC islands. These are essential control factors for at least most housekeeping genes (i.e. those genes used to maintain normal cell functions). GpC islands (short stretches of DNA that can be methylated) are only one of many mechanisms that the cell uses to control gene expression. Over 75,000 GpC islands exist in the total human genome, which is now estimated to consist of approximately 34,000 genes. Even if a gene were functional, it would be useless until the proper control structures (such as GpC islands) also evolved. Was each of these 75,000 GpC islands produced randomly by the Dawkins method? The probability of this is so low as to be essentially zero.

Another critical gene control method that also must evolve is methylation. Methylation is a poorly understood mechanism that involves adding methyl (CH_3) groups to genes to silence them, or demethylation to reactivate them. This control is critical because in order for a cell to survive, the required genes must be strictly regulated. This means they must be switched on only when they are needed, and inactivated when they are not needed.

As an example of just how complex promoters are, the cholesterol ester hydrolase promoter has both a CG-box sequence (which can bind the positive transcription factor *Sp1* to drive transcription), and an inverted CCAAT box (which, in turn, can bind another positive transcription factor). To function, the gene must have the entire required set of structural and sequence characteristic so that it is compatible with other genes.

Promoters can be fairly long stretches of DNA. For example, the cholesterol ester hydrolase promoter contains 1,317 nucleotides. Of the 34,000 human genes, as far as we know, all contain promoters, each of which must be compatible with the others. Their protein products also must be compatible. A single, inappropriate interference can be lethal; so can two proteins that have different functions but which structurally are too similar in their critical areas.

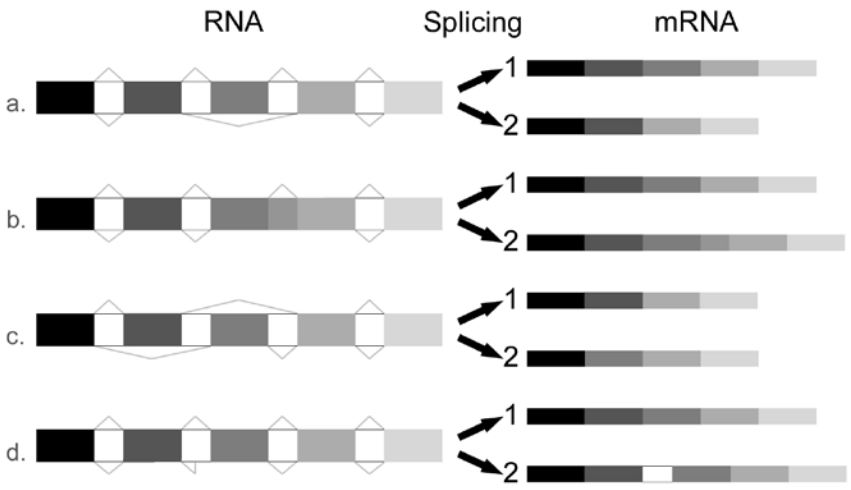
Every protein has to be compatible with other proteins, yet not interfere with their activities. The cholesterol ester hydrolase gene promoter has both structural design and sequence characteristics that are common to other well-known sterol responsive promoters. This is required for cholesterol ester hydrolase regulation to be compatible with, and functionally integrated into, the cell as a whole.

Packaging and processing concerns

When mRNA is transcribed from DNA, both the exon and the intron non-coding sequences must be transcribed into a macromolecule known as *heterogeneous nuclear RNA* (hnRNA), or immature RNA. The 5' end of the hnRNA must have the correct sequences so it can be capped by adding a chemically modified guanine (G) nucleotide. Then a poly (A) tail, which is composed of a series of 100–200 adenine bases constructed from adenosine monophosphate residues from ATP, is added to the 3' end. Last, address label and control end bases are added. These modifications help stabilize the RNA molecule and prevent it from being degraded when it reaches the cytoplasm.[14] Before the RNA

exits the nucleus, the introns must be removed by a precise cleavage-ligation reaction called splicing, thereby producing a functional mRNA.[15] The gene must contain all of the base pairs compatible with these operations in order for these modifications to occur.

Enzymes and small ribonucleoprotein structures exist in all *eukaryotic* cells to assist in removing introns. In most cases, type I introns are removed in the nucleus by a complex splicing machine composed of small nuclear ribonuclear proteins (SNRP). SNRPs are complex machines that consist of 60 to 300 small nuclear RNA (snRNA) nucleotides in an 'intimate alliance with a bouquet of proteins'.[16,17] About 14 types of snRNAs have been identified, but only four or five kinds (including type U1, U2, U4, U5 and U6) commonly are used in most cells. About a million SNRPs are found in each mammalian cell nucleus, but fewer SNRPs typically exist in most other eukaryotes.



RNA to mRNA splicing. Shaded areas are exons in the original RNA, white boxes are introns. In each case a single type of RNA transcript is spliced in two alternative ways to produce two distinct mRNAs. The grey triangles distinguish what is removed during the splicing. a) An exon is left out in the second splicing. b) In the first spicing, part of the exon is considered an intron and therefore is left out. c) In both splices, mutually exclusive exons are left out. d) Part of an intron is considered an exon and is included in the second splice (after Albets et al.).[33]

The process of intron removal involves a precise looping process controlled by a specific nucleotide sequence that abuts the exons.[18] Almost all known introns are identified by specific consensus sequences, including GT at the start or donor (3') end, and AG at the other end (called the acceptor (5') site), which help to identify introns for removal. The term 'consensus sequence' is employed because, although the sites often consist of more than just GT donor and AG acceptor bases, so far as is known these sequences are common to all eukaryotic organisms.[19] Consensus sequences are DNA segments that use similar base sequences in different genes within a single gene family or across many different species for a specific function. Alternative splicing is a process which variations in intron splicing can produce different polypeptides from a single gene. Alberts *et al.* claim that a 'substantial proportion of higher eukaryotes genes produce multiple proteins in this way'.[20]

Protein folding

Proper folding of the protein must occur in order for the protein to function. The three-dimensional shape of a protein is determined primarily by the polypeptide chain amino-acid sequence. Although many proteins can fold spontaneously into the correct shape in the cell cytoplasm, incorrect folding and destructive aggregation of proteins during their biogenesis, or under conditions of cellular stress, is prevented by molecular machines called chaperonins.[21]

The importance of chaperonins is illustrated by the chaperonin GroEL, a complex 57,000 AMU (atomic mass units) homo-oligomer that consists of 14 subunits and contains substrates that consist of several domains with ab-folds (which contain both α-helices and β-sheets with extensive hydrophobic surfaces). GroEL interacts with a specific set of polypeptides, including certain essential transcription/translation machinery components and specific types of metabolic enzymes. About one third of these proteins are unstable structurally, and repeatedly return to GroEL for conformational maintenance. Chaperonins must also have evolved in many cases to fold the protein properly; and without them, the protein will not work effectively or consistently. All of the structures that enable the above gene processing scenario to occur must exist in the genes for it to function properly.

An example of the interactions that occur in gene regulation is the addition of an extra dose of a gene. In many organisms this can have the paradoxical effect of blocking the gene's expression. The phenomenon, called post-transcriptional gene silencing (PTGS) occurs because the added gene causes destruction of the mRNA produced by both it and the corresponding cellular gene. The result is that the production of the gene's protein shuts down. The silencing mechanism evidently is produced by an enzyme called *RNA-dependent RNA polymerase* (RdRP), which copies one RNA from another, creating antisense fragments. Evidence for this mechanism includes the finding that RdRP levels rise when silencing occurs.[22]

In Dawkins' scenario, the gene would have to evolve to the point that it is functional, and only then could the many accessory structures described above (which allow appropriate transcription and gene/protein processing) be added and activated. If regular transcription of non-func-

tional proteins occurred, the cell would waste an enormous amount of energy producing and cutting up what had no function. This would have a detrimental effect in the cell, for example, if the excess proteins interfered with other cellular mechanisms. The cell is far too complex and interconnected to allow new genes to evolve without causing problems that often would be lethal.

The approximately one million protein molecules in each eukaryotic cell not only must be produced in the correct quantities at the correct time—neither too early nor too late—but they also must end up at the correct location in the cell. Recently, a Nobel Prize was awarded to German-born cell biologist Guenter Blobel, now at Rockefeller University in New York, for his discovery that each protein must carry an address tag. This address tag is written in amino acids and is coded by the protein's gene. This tag allows the cellular machinery both to identify and to route each protein type to its correct location in the cell. This complex address system is used in all plants, animals, and even by such allegedly primitive life as fungi.

Assuming a protein useful in some region in the cell could have evolved, the correct address label also would have to evolve simultaneously—an event that has a probability of close to zero. Without this tag, the protein never would get to where it was needed, and the wrong address label could carry the protein where it might be harmful or even lethal. The importance of this address system is illustrated by the discovery that address tag defects can cause diseases like cystic fibrosis, in which certain proteins do not arrive at the correct location in the cell (consequently, the cell membrane does not function properly). Address tag errors can even cause certain types of high blood cholesterol and kidney stones.

A mutated protein will most likely be recycled. As Wickner *et al.* note, most abnormal and incompletely or improperly synthesized proteins are degraded by adenosine triphosphate-dependent proteases.[23] Usually only normal and completely synthesized proteins are able to escape this cell quality control process.

Protein recycling

For a cell to function, it is critical that the protein be produced at the correct time in the correct amounts, and also be inactivated or broken down at the appropriate time. The new protein therefore must also co-evolve with a set of genes that can recycle it properly. All proteins have a half-life, which for some is a matter of minutes.

One example of the many proteins, which if improperly recycled can cause problems, is beta amyloid.[24] A defect in its recycling or processing causes it to build up in the brain, clumping together and influencing plaques to form that can damage brain tissue and can contribute to one form of Alzheimers disease. Evidently, the problem involves defects in the enzymes that recycle the protein.

As part of the recycling process, the amyloid precursor protein first is cut at two places by an enzyme pair called beta secretase and gamma secretase. Then, other enzymes cut these three protein sections up further until recyclable amino acids remain. In Alzheimer's patients the middle protein is not cut up, and as a result the next recycling step cannot occur, possibly due to a mutation of the enzyme.[25] As a result, it clumps together, producing the plaques that damage brain tissue and cause Alzheimers. The larger protein evidently is benign, and one possible approach to therapy is to inactivate the beta secretase and gamma secretase enzymes. The hoped-for result of this therapy is to prevent the accumulation of plaque, thus halting Alzheimers progression.

One potentially feasible way of inactivating of beta and gamma secretase is by inhibitors.[26] Unfortunately, not enough is known about the process to determine whether inhibiting these two enzymes will cause other problems, because these enzymes also may be used to recycle, modify, or process other proteins.

Furthermore, the long-term effects resulting from accumulating amyloid precursor protein likewise is unknown. This example clearly illustrates some of the potential problems that can result from the build up of various proteins. This shows that a new protein would be usless unless all the enzymes needed to recycle it evolved simultaneously. This is a highly unlikely and probably impossible event. Some evolutionists propose that an existing protein was mutated so that it could serve another function; thus all of the proteins need not have evolved from scratch, also an unlikely event.

Others speculate that preventing the production of beta amyloid itself by blocking the enzyme that controls its production (called presenilin) may prevent Alzheimers. This approach also may produce problems because presenilin may be involved in other biochemical reactions, and the proteins on which it acts may accumulate, thereby potentially causing problems. The effect of interfering with beta amyloid also is unknown.

An example of how critical the timely, efficient destruction of a protein is would be the 'destruction box' DNA sequence that is part of certain cell cycle regulator genes. The destruction box is a short peptide motif consisting of approximately ten amino acids near the N-terminus, something that is required for programmed proteolysis and consequently for the completion of the cell cycle.

The specific and rapid destruction of the proteins cyclin A and cyclin B during mitosis is critical for the cell cycle to function, and malfunction often is lethal.[27] Although the destruction box is **necessary** for cyclin A degradation, it is not **sufficient**. The protein kinase activity of p34cdc2 is inactivated when the mitotic cyclin to which it is bound becomes degraded. The amino (N)-terminus of mitotic cyclins includes a conserved destruction box sequence that is essential for the protein's degradation.[28] Mutations that interfere with p34cdc2 binding also interfere with cyclin destruction, indicating that p34cdc2 binding is required for

Base pairs are read in sets of three (codons). Inserting a base pair will cause a frame-shift, *resulting in different codons and thus different amino acids. In the example above, a U-base was added (shaded) and thereby changing the amino acid sequence (and the resultant protein).*

cell cycle-regulation.

DNA packing

Another means used to control genes involves the DNA packaging mechanism that produces chromosomes. Chromatin performs the 'most remarkable feat of packaging' known to science.[29] The nucleus of each human cell historically contains 2 meters of DNA, which is tightly coiled into chromosomes. In its most 'condensed' state, the DNA is so tightly bundled that it is inaccessible to gene-activating enzymes. To code protein, stretches of the appropriate chromosome are uncondensed, exposing the proper genes so that they are available for expression. This requires a control and signaling mechanism. But in sections known as constitutive heterochromatin, DNA is condensed permanently and therefore is safely packaged to prevent transcription. The regulation mechanism—which allows unpacking to occur only at the proper time—also must provide for the gene to function when it is needed and to be properly repackaged when it is not.

To evaluate the probability of Dawkins' model, Spetner ran a more realistic simulation and found Dawkins would have been at his computer for a long time to achieve his results. Dawkins' simulation would have been more realistic had he used a genome of 500 symbols instead of only 28, and a mutation rate of 10^{-10} instead of 0.04. Had he done that, Spetner estimates that Dawkins would have needed about seventy billion generations to obtain the mutations needed to produce his phrase. He would have needed many more to allow the selection to spread through the population. Dawkins also ignores the high likelihood that a beneficial mutation will be wiped out by genetic drift long before it takes over the population.[30]

Conclusion

There are numerous reasons why Dawkins' example is at best an illustration of why mutations **cannot** function as the major or even a minor means of creating genes and ultimately new basic kinds of organisms.[31] This brief review covers but a few of the more salient reasons. It is important that creationists focus on questions involving molecular biology because this area is central to the whole question of Darwinism. Although other mechanisms have been proposed to contribute to evolution, at its heart is the need for new information by mutations. Therefore, the critical analysis of Dawkins' 'simulation' does not provide support for the belief in the sufficiency of mutations.

Glossary

3' end — (3 prime) the carbon-3 end of DNA which is the start of the gene transcription end. In other words, the DNA template is read in a 3' to a 5' direction, a direction said to be upstream.

5' end —(5 prime) the carbon-5 is the opposite of the carbon-3 end of DNA, the end of the gene base pairs.

Address label — an amino acid chain used to help direct proteins to where they belong in the cell for further processing, storage, or function.

α-helices — a string of amino acids that fold to form a coiled, spring shaped structure which is used as a unit that is assembled with other units to become part of a protein.

Allele — one of the two or more forms of a gene that codes for a specific trait.

AMU — (atomic mass unit) a measure of the mass of atoms equal to about one hydrogen nucleus. The new standard uses one-twelfth of a carbon 12 isotope. Also called a Dalton.

Back mutation — a mutation which reverses a previous mutation, in essence, restoring the gene to its previous normal functional state.

β-sheets — also called a pleated sheet. A structure made out of amino acids that forms a structure that resembles corrugated cardboard.

Capping– A methylated G nucleotide that is added to the **5' end** of the RNA to prevent RNA fragments such as **introns** from binding to the **mRNA**, and also to prevent it from degrading.

CAAT box — also called a cat box, is a set of nucleotides which serves as a transcription regulator, specially as a **promoter** to up regulate protein.

CG-box sequence (also called CG island) — is a set of nucleotides that surround **promoters** of housekeeping genes. Housekeeping genes code for many proteins essential for cell life.

Chain termination — the code used to end a gene sequence and cause dissociation of the newly synthesized RNA, the RNA polymerase, and the accessory proteins from the transcribed DNA.

Codon — the sequence of three adjacent nucleotides that code for a specific amino acid. An example is the three cytosines code for proline.

Control end bases — bases that have several functions including to help regulate gene length and protect the end of an RNA section.

Chromatin — a coiled string of duplex DNA wrapped around histones, the normal state of DNA which is folded up into chromosomes just before cell division.

Dynamic mutation (causing a triplet repeat expansion disorder) — a mutation that causes the addition of one or more sets of three (a **codon**) to the gene. Also called a strutting mutation.

Enhancers — a set of DNA sequences that upregulate gene transcription, thus resulting in the production of a greater amount of protein. Many enhancers contain a unique patchwork of nucleotide motifs.

Exon — the coding DNA sequences which are transcribed into RNA and then into protein. The gene is **spliced** together after the introns are removed.

Frame-shift mutation — a mutation that causes a shift in the reading frame, i.e. a base is added, throwing off the triplet pattern. An example is AUGGUU becomes AU**C**GGUU if the base C is added after the first U. The result is instead of reading methionine (AUG) and valine (GUU), it is read threonine (ACU) and glycine (GGU) and the same type of changes are caused to occur downstream.

GpC islands — are short stretches of DNA that can be methylated (and thereby controlled) and serve as essential gene regulatory factors for most housekeeping genes (i.e. those genes used to maintain normal cell functions).

Heterozygous — a diploid cell or individual that contains two or more different **alleles** at one or more positions on the chromosome.

Heterochromatin — condensed **chromatin** that consists largely of DNA which is inactive in RNA synthesis.

Homo-oligomer — a short polymer of nucleotides that is identical with a specific nucleotide section.

Intron — a sequence in genes that is non-coding but is excised as part of processing before the mRNA exits the nuclei. The functions of introns are now the subject of intense research. So far dozens of uses have been confirmed or researched.

Mevalonate — salt of a dihydroxy acid, mevalonic acid, used in organic synthesis. $HO_2CCH_2C(OH)CH_3CH_2CH_2OH$

mRNA — messenger RNA is an RNA molecule that specifies the amino acid used to make protein.

N-terminus — the NH_2 end of a polypeptide or protein.

Peptide motifs — amino acid structural elements that fold in a very similar manner in many different proteins and are called super secondary structures or motifs. Examples include helix-turn-helix and calcium binding motif. This large family of secondary structures is used to construct the protein end product.

Polyadenylation — the adding of a sequence of from 50 to 200 AMP (one of four nucleotides of RNA) residues to form what is called a poly(A)tail. They are added to the 3' end of most eukaryotic mRNA's and serve several functions including to enable proper mRNA transport to occur, and also to regulate turnover and translation.

Post-translation — modification of protein after translation occurs, i.e. when the mRNA code is converted into amino acid chains.

Promoter — the area on duplex DNA where RNA polymerase attaches to initiate transcription. This patchwork of specific base pair sequences called motifs functions as a regulatory element.

Proto-oncogenes — a gene that controls normal cell proliferation and in some cases cell differentiation. When damaged it contributes to cancer, and is then called an oncogene.

Recessive mutation — a mutation that is not expressed in the phenotype if the defect is on only one allele and the other allele is normal, and therefore the normal gene can usually produce enough functional protein.

rRNA — a type of RNA that is used to construct ribosomes.

Silencers — a transcript control factor which blocks the transcription of DNA.

Splicing — the process of removing introns and reconnecting the gene exons and repairing the side chains of the remaining nucleotides (the exons).

Squalestatin — a protein which can undo or repair certain kinds of DNA damage.

Suppressor — a protein regulator produced by a suppressor gene that serves to down regulate protein production.

TATA boxes — an almost universal nucleotide sequence used in eukaryotes about 25 base pairs upstream from where transcription begins. This TA-rich region is where transcription factors bind, and is part of the means used to regulate the production of proteins.

Transcription tetrad — four protein units which work as a unit to regulate transcription. If one is defective, the entire unit will not function. An example is the *p53* tumor suppressor gene.

Transcription factor — a protein involved in regulating RNA transcription. They play a central role in regulating gene expression which in turn is critical in regulating a cell's many functions including its metabolism, growth, division and production of cytokines and hormones. Transcription of most eukaryotic genes is usually under control of several transcription factors, the average being around 5.

tRNA — is an abbreviation for transfer ribonucleic acid. It is a complex structure used to transfer the proper amino

acid to its correct place on the amino acid chain as it is being constructed.

Acknowledgments

I wish to thank Bert Thompson, Robert Kofahl, John Woodmorappe and Wayne Frair for their valuable insight and feedback on an earlier draft of this paper.

References

1. Dawkins, R., *The Blind Watchmaker*, Norton, New York, 1986.
2. Dewdeny, A.K., A blind watchmaker surveys the land of biomorphs, *Scientific American* **258**(2):128–131, 1988.
3. Spetner, L., *Not by Chance*, Judaica Press, New York, pp. 167–168**,** 1997.
4. Raymo, C., *Skeptics and True Believers*, Walker, New York, p. 152, 1998.
5. Truman, R., Dawkins' weasel revisited, *CEN Tech. J.* **12**(3):358–360, 1998.
6. Jorde, L., Carey, J. and White, R., *Medical Genetics*, Mosby, St. Louis, 1997.
7. Vogelstein, B. and Kinzler, K., *The Genetic Basis of Human Cancer*, McGraw Hill, New York, p. 398, 1998.
8. Lewis, R., *Human Genetics*, McGraw Hill, New York, pp. 187–188, 2001.
9. Weinberg, R., *One Renegade Cell; How Cancer Begins*, Basic Books, New York, p. 11, 1998.
10. Asimov, I., *The Wellspring of Life*, Abelard-schuman, New York, p. 146, 1960.
11. Zhang, Z., Lantry, L.E., Wang, Y., Bergman, G., Lubet, R.A. and You, M., Susceptibility to 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone (NNK)-induced Lung Adenomas in Germline *p53* Mutated Mice, paper submitted for publication.
12. Natarajan, R., Ghosh, S. and Grogan, W. McL., Molecular cloning of the promoter for rat hepatic neuro cholesterol ester hydrolase: evidence for transcriptional regulation by sterols, *Biochemical and Biophysical Research Communications* **243**:349–355, 1998.
13. Ritter, P., *Biochemistry; A Foundation*, Brooks/Cole Pub., New York, pp. 792–795, 1996.
14. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D., *Molecular Biology of the Cell*, 3rd Edition, Garland Publishing Inc., New York and London, pp. 369, 456, 458–466, 1994.
15. Alberts *et al.*, Ref. 14, pp. 373–375 and Ritter, Ref. 13, p. 659.
16. Patrusky, B., The intron story, *Mosaic* **23**(3):22–43, 1992.
17. Ritter, Ref. 13, p. 710.
18. Downie, S.R., Olmstead, R.G. and Zurawski, G., Six independent losses of the chloroplast DNA rp12 intron in dicotyledons; molecular and phylogenetic implications, *Evolution* **45**(5):1245–1259, 1991.
19. Jorde *et al.*, Ref. 6, pp. 21, 32.
20. Alberts, *et al.*, Ref. 14, p. 454.
21. Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F. and Hartl, F.U., Identification of *in vivo* substrates of the chaperonin GroEL, *Nature* **402**:147–148, 1999.
22. Strauss, E., Candidate 'gene silencers' found, *Science* **286**:886, 1999.
23. Wickner, S., Maurizi, M. and Gottesman, S., Post-translational quality control: folding, refolding, and degrading proteins, *Science* **286**:1888–1893, 1999; p. 1889.
24. Selkor, D.J., Alzheimer's Disease: genes, proteins, and therapy, *Physiological Reviews* **81**(2):741–766, 2001.
25. Pinnix *et al*., A Noval gamma-secretase assay based on detection of the putative C-terminal fragment-gamma of amyloid beta protein precursor, *J. Biological Chemistry* **276**(1):481–487, 2001.
26. Robners, *et al*., The regulation of amyloid precursor protein metabolism by cholinergic mechanisms and neurotrophin receptor signaling, *Progress in Neurobiology* **56**:541–569. 1998.
27. Stewart, E., Kobayashi, H., Harrison, D. and Hunt, T., Destruction of *Xenopus* cyclins A and B2, but not B1, requires binding to p34cdc2, *Embo Journal* **13**(3):584–594, 1994.
28. van der Velden, H.M. and Lohka, M.J., Cell cycle-regulated degradation of Xenopus cyclin B2 requires binding p34cdc2, *Molecular Biology of the Cell* **5**(7):713–724, 1994.
29. Ochert, A., Transposons; so-called junk DNA proves its worth, *Discover* **20**(12):59–66, 1999; p. 64.
30. Spetner, Ref. 3, p. 170.
31. Read, B., 'Dawkins' weasel revisited', Letters to the editor, *CEN Tech. J.* **13**(2):73–75, 1999.
32. Walkup, L.K., Junk DNA: evolutionary discards or God's tools? *CEN Tech, J.* **14**(2):18-30, 2000.
33. Alberts, *et al.*, Ref. 14, p. 454.

Jerry Bergman has seven degrees, including in biology, psychology, and evaluation and research, from Wayne State University (Detroit), Bowling Green State University and other colleges. He has taught at Bowling Green State University (Ohio) and at the University of Toledo. He is now a professor of science at Northwest College, Archbold (Ohio), and is working on a third Ph.D., this one in molecular biology.