The proportion of polypeptide chains with native folds—part 5: experimental extraction from random sequences

Royal Truman

This is the second to last of a six part series which critiques the most widely cited papers claiming that a high proportion of random polypeptides naturally form native-like folds. Parts 5 and 6 might not seem like exciting reading, but they cover the ONLY known experiments we are aware of which look at random polypeptides. When Professor Robert Sauer from MIT was contacted by the author questioning some published claims, no effort was made to defend his own work. Instead, attention was drawn to the research by Professor Jack Szostak, which is discussed here.

We find the claim that one out of 10^{11} random polypeptides in free nature would produce proteins reliably folded to be unconvincing. The artificial polypeptides identified were much smaller than average-sized proteins, depend on the presence of zinc and ATP, and lacked the rich secondary structures characteristic of biological proteins.

What proportion of random polypeptide chains based on the twenty natural amino acids would fold into native-like folds? Ideally, one would examine experimentally a library of random sequences to answer this question. However, examining a vast number of random sequences 150 AAs (amino acids) long (the average size of a domain)[1] or 300 AA (the average size of a protein) is not feasible should indeed the proportion of those folding reliably be very small. Experimental methods such as CD (circular dichroism) spectra, NMR, denaturating spectra, X-ray crystallography, and so on, are very time-consuming.

To our knowledge only the series of experiments pioneered by Professor Szostak, winner of the 2009 Nobel Prize in Physiology or Medicine, have addressed this question empirically. During the last decade other researchers had made contributions to this original work, and some have concluded that a fairly high proportion of random polypeptides can fold stably. Lo Surdo states that

> "It had been previously suspected that the probability of a stable fold arising at random is extremely small (argument based on the complexity paradox that the disparity between the many potential protein sequences and the relatively few different structures known). However, we have confirmed that functionally directed in vitro evolution from random sequences can generate a novel fold with a tailored function. In addition, such folds may display the key features of naturally evolving proteins."[2]

We shall examine in this paper two key studies and two more in the next part of this series to see if this conclusion is warranted.

The proportion of native-like folded polypeptides is expected to decrease as chain size increases, due to undesirable interferences and decreasing solubility. For this reason, these experiments typically work with small chains of n = 80 AA (or shorter). The potential sequence space therefore covers $20^n = 20^{80}$, or 1 x 10^{104} alternative chains. Through clever search strategies, shown in figure 1, the researchers iteratively narrow the search space to promising regions in the sequence space.

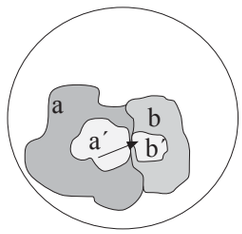


Figure 1. Strategy to estimate the proportion of random polypeptides with a native-like fold. Outer circle covers all possible sequences of length *n* amino acids (20^n possibilities), a portion of which is sampled (a). Of these some show indications of folding-related properties (a´) from which the best candidates are selected and mutated to generate new candidates (b´). The best sequences among a´ and b´ are members of the sampling space a + b. The best candidates in b´ are used to generate better variants c´ and so on. The estimated number of sequences of quality c´ compared to sampling size a + b + c gives an indication of the expected proportion in the total population of length n (the outer circle).

First study

A method to create large libraries of mRNA sequences lacking frameshifts, stop codons and internal translation initiating events has been developed[3] and was combined with the experimental technology described under 'Methods' below. Typically about 6 x 10^{12} different sequences are rapidly examined in these experiments based on their ability to bind to immobilized ATP in a separation

column. The researchers reasoned that strong and selective binding to ATP might be a relevant criteria to identify properly folded proteins. Noteworthy is that ATP binding proteins are found in all major enzyme classes.[4]

Since $10^{12} << 20^{80}$, the maximum number of alternative sequences 80 AA long, no duplicate sequences were likely to have been generated. After removing the sequences which did not bind to immobilized ATP, a subset of promising candidates were isolated (members *a´* in figure 1). Error-prone PCR amplification generated new, similar variants (region *b* in figure 1) likely not to have been present among those in sampling space *a*. From these the best candidates were isolated (sequences *b´*, figure 1). This process was repeated to select other polypeptides with yet stronger affinity for ATP, sequences *c´*, and so on. In this manner, promising portions of the overall search space can be efficiently identified.

The key insight is that sequences *b´* were the result of an intelligently directed search, and have a high probability of including sequences able to bind better to ATP than the best from *a´* could. They are members of a much larger pool of candidates, *a* + *b* (figure 1).

The search strategy assumes that sequences very different from those producing native-like folded proteins could be initially identified by natural selection and thereafter improved upon. This view is illustrated as assumption 'A' in figure 2. However, others[5] believe that chains significantly different from native folds are useless as an evolutionary starting point, since there are countless energy minima which have no relevance to the protein folding state (assumption 'B' in figure 2). This matter is not addressed in these series of studies, although clearly the authors simply assume 'A' to be true.

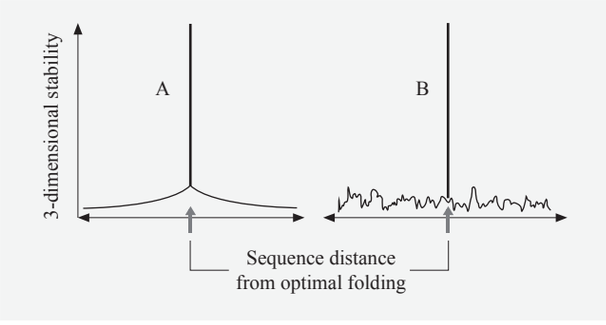


Figure 2. Sequences very different from that producing a native-like folded state might be selectable and steadily improved (assumption A); or distant sequences with properties suggesting folding-state potential might be evolutionary dead-ends, offering no opportunity to be selected to produce a native-like folded state. (Based on work from Axe[5]).

Now to the specifics. In the seminal paper, Keefe and Szostak[6] prepared a library of 6 x 10^{12} non-redundant random proteins 80 AA long, designed to avoid random stop codons so as to generate the maximum variety possible.[7] The random 80 AA region was flanked by short affinity tags to permit subsequent purification of the proteins, figure 3.

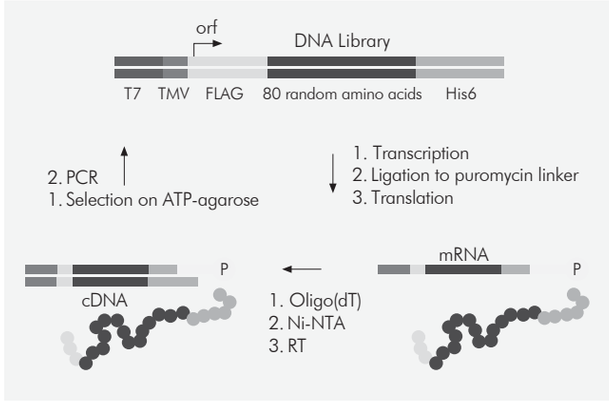


Figure 3. *In vitro* selection and amplification of mRNA-displayed proteins. The DNA library was constructed with the following elements: a T7 promoter sequence, which is not translated but is required by the T7 RNA polymerase[8] used to perform the translations; the following TMV (tobacco mosaic virus) sequence is also not translated and serves as a translation enhancer. An FLAG-tag follows the start codon, which consists of amino acids DYKDDDD to permit affinity chromatography.[9] The next block consists of 80 contiguous random non-stop codons, followed by a 6 histidine tag,[10] used for affinity purification.[11] The affinity tags permit the mRNA with frameshifts and internal starting events to be eliminated. The mRNA transcribed is ligated to a linker which includes 27 adenines and puromycin, which results in the fused mRNA-puromycin-protein product. Purification and reverse transcription generated a mixed RNA-DNA chain attached to puromycin-protein. The library was incubated with bound ATP-agarose and placed in a separation column. Fused proteins which don't bind to ATP were washed away, and then the proteins which did bind were eluted with free ATP.

The eluted fractions were amplified by polymerase chain reaction (PCR). The DNA produced was used to generate another library of mRNA-displayed proteins and the scheme repeated for a total of eight rounds (figure 3). During the first round about 0.1% of the proteins bound to ATP in the column, and after eight rounds of selection the amount had increased to 6.2%. The low level of ATP-binding of the best sequences isolated so far (6.2%) is probably due to the conformational heterogeneity of the sequences isolated[7]: the same sequence can fold in many non-native-like manners.

The authors cloned and sequenced 24 of the members after the eighth round and found that most could be grouped into four sequences (figure 4A). These were unrelated to each other and to any biological sequences known. Representatives of each of the four families (A–D) were examined. Between 5% and 15% of these fused mRNA-displayed proteins bind to immobilized ATP (see last column of figure 4A). Note that far more of the chain binds to ATP if still attached to its mRNA (second to last column of figure 4A), which was not discussed in the publication.

Figure 4. **A)** Consensus sequences of initially selected protein families after round 8; **B)** Sequence data of mutagenized re-selected proteins after round 18, free protein; **C)** Truncated versions of DNA-tagged protein '18-19'; **D)** Deletion analysis of clone 18-19 fused to MBP protein.[12]

This fact illustrates that organic molecules such as ATP can easily bind to many classes of bio-chemicals, not just to polypeptides.

In order to optimize binding to ATP, the improved library was mutagenized three times using mutagenic PCR amplification, with an average rate of 3.7% per amino acid for each round. After six additional rounds of amplification without mutagenesis and *in vitro* selection followed by elution with ATP, the proportion which binds to ATP in the column rose to 34%.

Of the 56 clones sequenced, all were derived from a member of the family B (figure 4B).

Comparing to the original family B revealed that the mutations generated had produced four which bound effectively to ATP (present more than 39 times in the 56 sequences examined), see second row from the top in figure 4b, labelled '18predom'. In addition, 16 other substitutions generated were also selectively enriched (present more than 4 times among the 56 examined), see second row from the top in figure 4b, labelled '18select'. This suggests that different amino acids distributed throughout the protein sequence might be interacting with various portions of ATP.

After selection in round 18, eight individual proteins were chosen randomly. As free protein, the proportion which binds in the column (and was then eluted with free ATP) varied from 5% to 40%.[12]

The four proteins that bound best to ATP were expressed in *E. coli* as fused proteins to maltose-binding protein (MBP), which is often done when the proteins are not soluble enough to work with conveniently. The clone which bound best was labelled '*18-19*' (figure 5) and had a dissociation constant of K_d = 100 nM for ATP at 4°C and 1:1 stoichiometry. However, gel filtration indicated that 65% of *18-19* is monomeric and the rest binds in higher-order aggregates. Only the monomers bond to ATP.

Unsurprisingly, selection for binding to ATP did not improve binding to various other bio-chemicals. Chemical analogs to ATP, such as CTP[13], GTP[14], UTP[15], and ITP[16] bound far less effectively to protein *18-19* (figure 5). Furthermore, removing one, two or three phosphate groups led to molecules which bound less effectively to ATP (figure 5). Removing either of the 2′ or 3′ hydroxyl groups also reduced the amount of binding.

The minimal region for ATP binding of clone *18-19* was explored by deleting portions of the protein. The fraction of the protein which bound to the ATP in the column, relative to *18-19*, is shown in the last column of figure 4C. Removing large sections from *18-19* improved the binding: those residues were apparently interfering with binding to ATP. A core domain of 45 amino acids was found to be sufficient to bind efficiently.[12] Significantly, the deletion shown in the fourth row from the top in figure 4C had a ten-fold deleterious effect on ATP binding. This drew attention to the fact that a **C**XX**C** motif, involving two cystein amino acids, had been destroyed.

In another experiment, variant *18-19* was fused to MBP protein and the dissociation constants, K_d, was measured after removing various fragments (figure 4D). Loss of affinity was observed for sections removed in the N-terminal portion (left-hand side of the sequences in figure 4D, note the last three rows). The authors propose that regions surrounding the important core which interacts with ATP might stabilize its structure or that additional amino acids might also interact with ATP.[12] This observation is consistent with the accepted view that domains require a significant number of residues to be present. The CATH database of domains gives an average domain length of 159.5 AAs and the three smallest have 13, 14, and 17 AAs (out of 128,688 entries). These three smallest domains consist of merely a single tiny alpha coil.[18]At the large end, the CATH database revealed 18 domains having over 500 AAs.[18]

The authors concluded, on the basis of their experiments that

> "The frequency with which ATP-binding proteins occur in sequence space can be estimated from the observed recovery of four such proteins from a non-redundant library of 6 x 10^{12} random sequences. On the basis of the average behaviour of the proteins isolated before mutagenesis, only about 10% of the potentially functional sequences present in the first round would be expected to generate correctly folded active proteins and thus survive to be amplified ...We therefore estimate that roughly 1 in 10^{11} of all random sequence proteins have ATP-binding activity comparable to the proteins isolated in this study."[17]

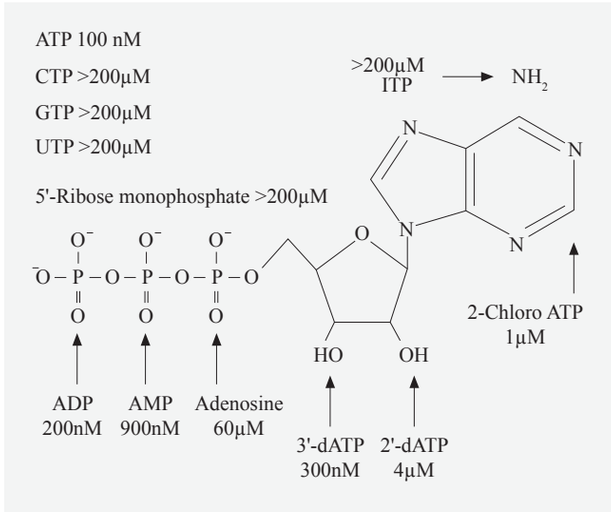


Figure 5. Dissociation constants, K_d, of ATP and similar analogs bound to clone *18-19*.[17] The innermost phosphate is designated by a, the middle one by b and the outermost one by g. The arrows show the group removed and the resulting K_d value.

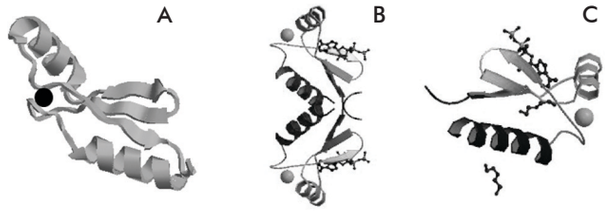


Figure 6. Crystal structure of clone *18-19* with the coordinates provided in the pdb database. The representation was created with RasTop 2.2[21] **A)** Using protein identifier 1UW1.[22] The small circle shows the location of the chelating zinc atom. **B)** Using protein identifier 1UW1. Notice that the researchers reported a dimeric structure in the crystal state. Two ADP molecules were found in the crystal structure. **C)** Using protein identifier 3DGN in the pdb database.[22] The crystal was formed in the presence of high concentration ADP to ensure its integration.

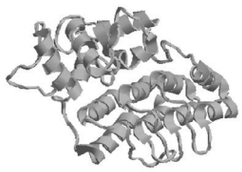


Figure 7. Crystal structure of 278-amino-acid domain ARG-GAP with which a small portion of clone *18-19* shows similarity. X-ray coordinates deposited in the pdb database under identifier 1DCQ.[22] Protein displayed with RasTop 2.2.[21]

Four years after this work[3] was published, the crystal structure of protein *18-19* was elucidated[2,19] (figures 6A–C) independently by two groups, and it was reported[20] to overlap quite well with a small portion of domain ARG-GAP of a biological protein (figure 7).

It must be emphasized that ADP, and not ATP, was found co-crystallized with clone *18-19*, even when the presence of ADP was rigorously prevented.

Second study

In a similar study[23] also perform in Dr Szostak's lab (see 'Methods' below), polypeptides able to bind to streptavidin[24] (figure 8) were identified from a semirandom library of 6.7 x 10^{12} members.

The library had been designed using short eleven-amino-acid cassettes which were then concatamerized together. By design, 44% of the incorporated cassettes encoded a peptide polar/non-polar pattern compatible to forming amphipathic alpha-helices and 45% to form beta-strands; 11% were not patterned. Similar experiments to that described above were performed[6] except that mutations were not deliberately generated. After seven rounds of selection for streptavidin and amplification with PCR, 20 different sequences were observed and analyzed. All were found to have frame-shifted, destroying the intended patterns which would lead to secondary structure. This experimental misfortune does provide us with another example of random sequences, however. The mutations and subsequence strong selection were due to the generation of one or more of the amino acid pattern *HPQ*,[27] which is known to bind strongly to streptavidin.

The clone with a single HPQ pattern which bound most strongly to streptavidin was selected for additional study. Removing more than half of the residues led to a 38-amino-acid species with slightly higher binding affinity, but additional deletions lowered the affinity.

The authors provide two possible explanations for these observations:

a. The flanking amino acids might stabilize the active conformation of the HPQ tripeptide motif.
b. Several distinct peptide elements could be interacting with different parts of the streptavidin.

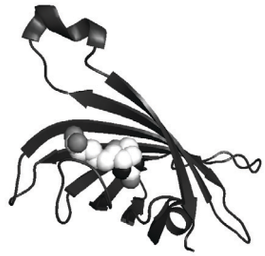


Figure 8. Secondary structure of a streptavidin monomer with bound biotin.[25] The secondary structure of a streptavidin monomer is composed of eight anti-parallel β-strands, which fold to give an anti-parallel beta barrel tertiary structure. A biotin[26] binding-site is located at one end of each β-barrel, which has a high affinity as well as a high avidity for biotin.[24]

Evaluation of the first study

The B-family proteins possess a pair of **C**XX**C** motifs (two **C**ysteins separated by two amino acids, 'X') (highlighted in figures 4A–D) and the presence of a single bound zinc, revealed by atomic absorption spectroscopy. The proteins with good ATP-binding properties, including the best-binding clone *18-19*, *require* zinc. The authors suggested that these proteins

> "… bind to ATP with a folded structure nucleated around, or stabilized by, a Zn^{2+} ion coordinated to the four invariant cysteines of the CXXC sequences."[17]

They reported that Mg^{2+} and other cations were not found to be suitable. Note that

> "Zinc was not added to the selection buffer, but is present at about 10 micromole mammalian blood, from which the reticulocyte lysate used for mRNA translation is prepared."[17]

The presence of zinc was accidental and no proteins were reported which bound well to ATP which lacked this metal.

As reiterated in a later paper,[28] the estimate of 1 out of 1011 involved sequences for which

> "All of the variants examined required high concentrations of free ATP in order to remain stably folded and soluble."[29]

They add,

> "… the ancestral and binding optimised proteins formed visible precipitates during the first 24 h and became completely insoluble after three days. Surprisingly, the core domain of the divergent protein 18-19 also formed a visible precipitate after 24 h and was almost completely insoluble by three days."[30]

In the process of preparing *18-19* for X-ray diffraction, large amounts were expressed in E. coli and then purified to homogeneity. Apparently the natural mixture of 18-19 conformations was not used:

> "Our structure determination targeted the functional core of ANBP [ANBP is an abbreviation for 'artificial nucleotide binding protein']. We aimed to optimise the production of a protein with a single conformation by screening different ANBP constructs, thus identifying N- or C-terminal residues unnecessary for either folding or ligand binding."[31]

The implication is that the secondary structure shown in the crystals generated seem not to represent the conformation formed in the solvated state at biologically relevant temperatures.

Only residues 7–73 could be identified in the electron density map, and residues 7–10 on the N-terminus were poorly defined.[32] The X-ray data revealed considerably smaller and less secondary structure than observed in CATH biological domains classified as Alpha Beta.[33] The crystals were developed at 100 K, a very low temperature, far below the freezing point of water. These facts suggest that under ambient conditions very little, if any, secondary structure is actually present. It is well known that the crystal state of a protein sometimes does not reflect its topology in solution.[34] Incidentally, the N-terminal alpha helix had been provided by amino acids present in the FLAG sequence, and therefore cannot be legitimately considered representative of a random sequence. Removing this alpha helix from figure 6B leaves still less secondary structure, considerably less than found in virtually all native proteins.

Figure 6A shows how the four cysteins binding to zinc hold the secondary structures together. However, this feature is not observed in biological proteins for which a similar zinc motif has been found (figure 7): the region held together by the zinc atom, while surely of biological value, is not responsible for the folding of the large domain. Indeed, the tiny zinc-chelated portion is far removed from the alpha coils and beta sheets, best seen by rotating the molecule using a protein viewer such as RasTop.

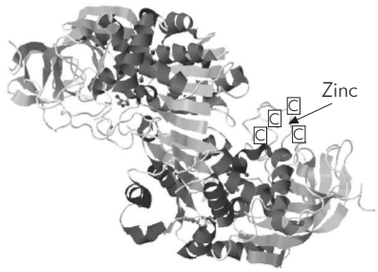


Figure 9. Human alpha alchohol dehydrogenase, showing zinc chelated to four cystein residues (C symbol in the figure). There is no evidence the zinc atom is producing the secondary structures. Figure generated using jmol from: www.rcsb.org/pdb/explore/jmol.do?structureId=1HT0&bionumber=1.

A well-known protein motif is known as the zinc finger.[35] These are very small structures found on about 500 proteins in human,[36] which usually consist of two cysteines located in a short strand or turn region followed by an α-helix which contains two His ligands. The key observation is that although zinc is used with these biological motifs, the zinc fingers are not responsible for the folded structure of the proteins in which they are found.

Figure 9 illustrates another example where zinc is chelated to four cysteins in a representative biological protein with rich secondary structure. Notice how the alpha coils and beta sheets are very distant from the chelation site, so that the native fold could not be driven by the presence of zinc.

However, in the synthetic protein described, a zinc atom seems to be entirely responsible for what little structure is present in the protein.

Analysis of the second study

We offer the following comments about this second study:

1. For the optimized 38-amino acid polypeptide, 13% did not bind even at high concentrations of streptavidin. Real proteins involved in enzymatic catalysis are highly specific and could not fail to bind to their intended ligands under these experimental conditions.
2. Removing more than half of the polypeptide led to slightly stronger binding, which reflects the potential for larger chains to offer steric interference to protein–ligand interactions. This reinforces the fact that as chain size for random polypeptides increases, the probability of deleterious interferences and insolubility increase.
3. Replacing the HPQ pattern with a similar HGA reduced the extent of binding by a factor of a thousand.[37] The special HPQ pattern was almost the whole cause for binding to streptavidin.[38] This reinforces the need to be very close to the final and correct sequence before natural selection can sense the gene's existence.

Evaluation of these studies

As reported in a later contribution, "Unfortunately, biophysical characterization of the selected ATP binding proteins proved impossible due to poor solubility."[39] For these reasons, laboratory tests necessary to characterize the secondary and tertiary structure, such as CD spectra and NMR, were not available when the first study was published.[6]

Analysis of the data published until 2004 suggests the following interpretation. The fortuitous contamination with zinc has led to a chelating site, thanks to the presence of CXXC motifs. Some of the isomers were thereby able to offer a small surface which can be moulded in the presence of ATP to permit some as of yet uncharacterized interactions. Notice how ATP only seems to interact with a single tiny beta sheet, see figure 6B. Selection for ATP did not generate a microenvironment on the protein which was also conducive to strong binding of ATP chemical analogs such as CTP, GTP, UTP, and ITP. This is not particularly surprising, given the selective protocol.

Removal of portions of the protein and improved binding upon mutating various positions need not indicate multiple interactions, the strength of which has been improved, but rather the *removal of disturbing amino* acids which can lead to insolubility, entanglement, intermolecular interactions and such features deleterious to ATP binding.

Often the structure of a protein in its dissolved condition under natural conditions and temperature is very different from the structure elucidated after crystallization using X-ray diffraction. The process of crystallizing clone *18-19*, at very cold temperatures, can select the particular conformations which are particularly stable. We suspect that the minimal secondary structure identified (figure 6B) is an artefact of the crystallization. Chelation with zinc forces several amino acids into close proximity, facilitating creation of alpha coils and beta sheets under biologically irrelevant conditions.

The limited data available supports this view. In the next part to this series, we draw attention to later work in which the authors discovered, unexpectedly, that CD spectroscopy at biologically relevant temperatures *denied the presence of secondary structure*. In addition, we must draw attention to the fact that in the absence of ATP suitable crystals were not formed. It appears that *not only zinc, but also ATP is responsible for moulding clone 18-19*, since otherwise it would not crystallize. The evidence is clear that this protein alone cannot produce a native-like fold under the above water-freezing temperatures. In a later paper we read,

> "Unlike many naturally occurring proteins, protein 18-19 requires high concentrations of free ligand in order to remain stably folded and soluble."[40]

This fact was not mentioned in the earlier work, nor quotes made subsequently of the estimated proportion of random peptides with native-like folds.

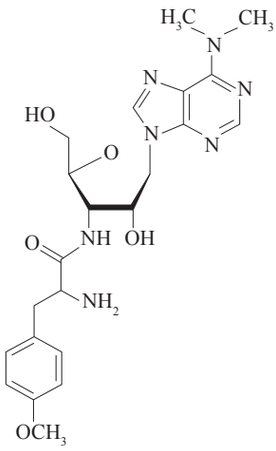
We welcome these kinds of studies to test the plausibility of our models, whether one believes in creation or evolution. We have some suggestions for those involved in this kind of work.

1. Justify the belief that highly specific binding to a ligand such as ATP, but not analogs, can permit an estimate of random proteins which produce a native-like fold. Why *must* a native-like fold be able to bind to this particular molecule? Might not additional ligands also bind strongly to other members of the sampled portion? *Why must a strong interaction at a judiciously constrained portion of the protein have any relevance to native protein folding*?
2. Generate a large number of variants similar to clone *18-19* and then determine whether there is evidence for secondary and native structure in the absence of ATP and zinc.
3. Repeat the studies based on selection for ATP or another ligand, using n = 40 … 200 amino acids, to determine whether the proportion (not absolute number) with a given level of binding follows a pattern. If larger random-sequence proteins become more chaotic, multimeric and insoluble, and prevent binding to a ligand, then native-like folds will be less likely.

We offer some objections for the suggested proportion of random polypeptides leading to stable folding, 10^{-11}, for evolutionary modelling purposes.

a. The data suggests that ATP is necessary, but once bound to the crystallized clone *18-19* was immediately destroyed (hydrolyzed to ADP). Since ATP has negative charge, it is clear that it binds to the basic amino acids, as seen in *18-19* and other crystal structures. ATP has many hydrogen and oxygen atoms which are able to form numerous H-bonds. And adenine in ATP can form planar stacking interactions with aromatic residues. No wonder that it binds somewhere! Zinc can bind to cysteins and other residues[41,42] such as the side chains of aspartic acid, glutamic acid and histidine, leading to local structures able to bind bio-molecules. These facts illustrate the vast number of destructive possibilities available along an evolutionary path towards forming a properly folded protein. And using an example which indiscriminately destroys the eminently valuable energy carrier, ATP, is a questionable argument for how proteins could arise naturalistically. It appears rather to highlight constraints which must be overcome.
b. It would be extremely rare for a random polypeptide sequence to reliably chelate with zinc in a free ocean.

Seawater has a concentration of only about 30 ppb zinc[43] but thousands of times greater (about 75 ppm) in the earth's crust.[43] Others report an even lower concentration in seawater, such as 0.6–5 ppb, and rivers containing between 5 and 10 ppb zinc.[44] The higher concentration in rivers reinforces the fact that the source of seawater zinc was erosion of the crust, and supposedly three or four billion years ago the concentration of zinc in a primitive ocean would have been far lower still. In addition, zinc is normally not chemically freely available to chelate with polypeptides, but is found firmly associated with other base metals such as copper and lead in ores, and binds well with sulphides instead.[41,45]



Structure of puromycin.[49]

If metal chelation with random polypeptides had played a key role in the origin of proteins, then it would be reasonable for metals such as aluminium and iron, on average about a thousand times more abundant on Earth[46] than zinc, to dominate protein chemistry. By similar reasoning, silicon would be expected to play a far more important role than phosphorous, being three hundred times more abundant,[46] but the opposite is true.

In a standard reference book on proteins we read,

> "Surprisingly elements such as aluminium and silicon that are very abundant in the Earth's crust (8.1 and 25.7 percent by weight, respectively) do not occur in high concentration within cells. Aluminium is rarely, if ever, found as part of proteins whilst the role of silicon is confined to biomineralization where it is the core component of shells."[47]

This is the opposite of what a naturalistic origin for proteins predicts.

c. The average biological protein is four times larger than 80 AA, posing some difficulties.

• A 2006 study estimated that about 10% of human proteins (2,800) potentially bind zinc.[41] Few, if any, biological proteins are composed of domains, with about 80 AA of them *all* containing zinc. The 10^{-11} proportion proposed for stable folding was based on a very small domain, and we need to extrapolate to a reasonable-size protein by using the geometric mean of the possibilities. For example, if in the presence of zinc, 10^{-11} sequences were to produce a native-like fold, but without it the probability were only 10^{-15}, then the probability for a full protein with four domains might be more like $[(10^{-11}) \times (10^{-15}) \times (10^{-15}) \times (10^{-15})]^{1/4} = 1.0 \times 10^{-14}$ *per domain*, or $(1.0 \times 10^{-14})^4 = 1 \times 10^{-56}$ for the whole protein.

• Claiming that life could have initiated with much smaller proteins, like 80 AAs, than observed throughout nature today probably does not simplify matters for an evolutionist. Proteins must interact with several other biochemicals, including other proteins, to be able to perform useful processes. If a few hundred multifunctional proteins (of size 300 AA on average) are necessary for autonomous life,[48] this number must be several times greater if the numerous functions per protein need to be handled by multiple simpler versions. Furthermore, the functional regions of extant proteins are held together in suitable geometries, facilitating interaction with the multiple intended partners. This would not be true initially for simpler disconnected proteins.

We conclude that the proteins studied do not offer evidence of native-like folding and that the claim that one out of about 10^{11} random sequences 80 AA long form native-like folds is not well supported.

Materials and methods used in these studies

Puromycin (figure 10) is an antibiotic that mimics the aminoacyl end of tRNA and terminates translation of mRNA. Figure 11 shows how codons are used by a ribosome to identify the next amino acid to be added to a growing protein. During the last step the protein is located at the *P site* of the ribosome, and at that time puromycin enters the *A site* instead of the tRNA which corresponds to the next codon.

To further clarify, the chain-growing reaction which forms a new peptide bond between the amino acids at the A and P sites is shown in figure 12. In nature, puromycin

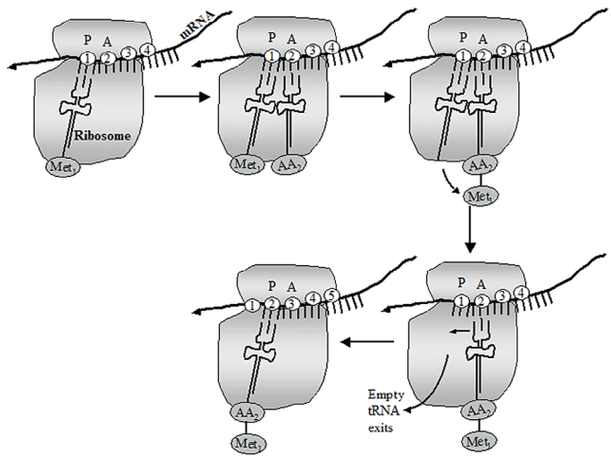


Figure 11. Sequence of steps carried out at a ribosome which lead to protein elongation.[50]

interferes at the A site by forming a stable amide linkage to the nascent protein using its free amino group.

In the methodology developed to create 'mRNA display proteins',[52] the puromycin is covalently attached to the mRNA of interest and remains so after the mRNA and its newly translated protein is released from the ribosome (figure 13). In this manner the mRNA and resulting protein remain physically linked.

Therefore, isolated proteins which bind to ATP in a column also provide their corresponding mRNA. Copies of those mRNAs can be made using PCR and they can also be deliberately mutated to generate similar variants in the hope of finding better proteins.

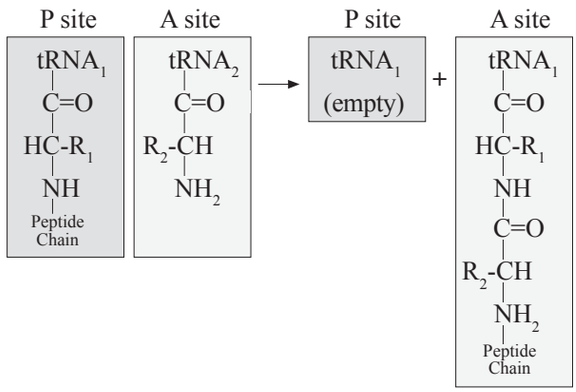


Figure 12. Reaction forming a peptide bond between the amino acids at the A and P sites of a ribosome.[51]

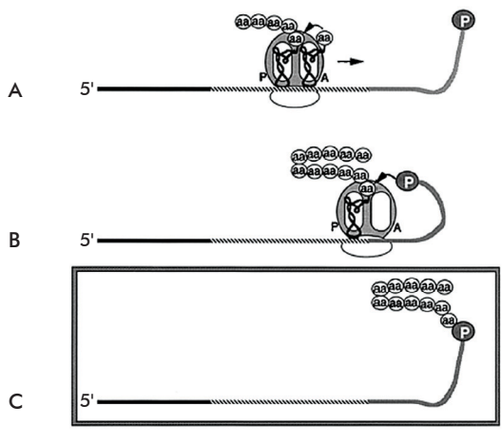


Figure 13. Proposed mechanism for the RNA-peptide formation on a ribosome. **A)** Translation occurs towards the attached puromycin, labelled '*P*'. **B)** When the ribosome reaches the end of the mRNA it comes in contact with the deliberately attached stretch of 27-adenines which causes translation to stall. The puromycin now has time to enter the *A* site of the ribosome and accepts the nascent peptide. The exact mechanism is not yet known.[53] **C)** Once formed, the mRNA-peptide fusion is purified from the mRNA which lack puromycin and from free peptides using affinity chromatography. Slightly under 1% of the input mRNA is converted to the fusion product,[54] producing a library of about 10^{12} fusions in a 10-ml translation reaction.[53]

References

1. For example, the average size of a globular domain according to the CATH database is 153 residues according to: Shen, M-y., Davis, F.P. and Sali, A., The optimal size of a globular protein domain: A simple sphere-packing model, *Chemical Physics Letters* **405**:224–228, 2005.
2. Lo Surdo, P., Walsh, M.A. and Sollazzo, M., A novel ADP- and zinc-binding fold from function-directed *in vitro* evolution, *Nat. Struct. Mol. Biol.* **11**:382–383, 2004. See p. 383.
3. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W., Constructing High Complexity Synthetic Libraries of Long ORFs Using *in Vitro* Selection, *J. Mol. Biol.* **297**:309–319, 2000.
4. Smith, M.D., Rosenow, M.A., Wang, M., Allen, J.P., Szostak, J.W. and Chaput, J.C., Structural Insights into the Evolution of a Non-Biological Protein: Importance of Surface Residues in Protein Fold Optimization, *PLoS ONE* **2(5)**:e467, 2007. doi:10.1371/journal.polne.0000467. Pointed out in p. e467.
5. Axe, D.D., Estimating the Prevalence of Protein Sequences Adopting Functional Enzyme Folds, *J. Mol. Biol.* **341**:1295, 2004.
6. Keefe, A.D. and Szostak J.W., Functional proteins from a random-sequence library, *Nature* **410**:715–718, 2001.
7. Keefe and Szostak, ref. 6, p. 715.
8. en.wikipedia.org/wiki/T7_RNA_polymerase.

   T7 RNA Polymerase is an RNA polymerase that catalyzes the formation of RNA in the 5′→ 3′ direction. T7 polymerase is extremely promoter-specific and only transcribes bacteriophage T7 DNA or DNA cloned downstream of a T7 promoter.
9. en.wikipedia.org/wiki/FLAG-tag. The tag contains eight amino acids: DYKDDDDK, with the left-hand end towards the N-terminal and the right-hand end, the C- end of the protein.
10. en.wikipedia.org/wiki/Polyhistidine-tag.
11. Cho *et al.*, ref. 3, p. 310.
12. Keefe and Szostak, ref. 6, p. 716.
13. CTP: cytidine triphosphate. For the structure, consult any biochemistry or cell biology textbook.
14. GTP: guanosine triphosphate. For the structure, consult any biochemistry or cell biology textbook.
15. UTP: uridine triphosphate, For the structure, consult any biochemistry or cell biology textbook.
16. ITP: inosine triphosphate. For the structure, consult any biochemistry or cell biology textbook.
17. Keefe and Szostak, ref. 6, p. 717.
18. CATH domains were downloaded from www.cathdb.info/domain/1vpcA00. (Version 3.3.0 status 24 June 2009). Entries analyzed with Microsoft Excel.
19. Simmons, C.R., Stomel, J.M., McConnell, M.D., Smith, D.A., Watkins, J.L., Allen, J.P. and Chaput, J.C., A Synthetic Protein Selected for Ligand Binding Affinity Mediates ATP Hydrolysis, *ACS Chemical Biology* **4**(8):649–659, 2009.
20. Krishna, S.S. and Grishin, N.B., Structurally Analogous Proteins Do Exist! *Structure* **13**:1125–1127, 2004.
21. www.geneinfinity.org/rastop/.
22. www.rcsb.org/pdb/explore/images.do?structureId=1UW1.
23. Wilson, W., Keefe, A.D. and Szostak, J.W., The use of mRNA display to select high-affinity protein-binding peptides, *PNAS* **98**(7):3750–3755, 2001.

24. en.wikipedia.org/wiki/Streptavidin#cite_note-0.

    "Streptavidin is a 60000 dalton protein purified from the bacterium *Streptomyces avidinii*. Streptavidin homo-tetramers have an extraordinarily high affinity for biotin (also known as vitamin B7). With a dissociation constant (Kd) on the order of ≈10–14 mol/L, the binding of biotin to streptavidin is one of the strongest non-covalent interactions known in nature. Streptavidin is used extensively in molecular biology and bionanotechnology."

25. upload.wikimedia.org/wikipedia/commons/5/5a/Streptavidin.png.

26. en.wikipedia.org/wiki/Biotin.

    "Biotin is necessary for cell growth, the production of fatty acids, and the metabolism of fats and amino acids. It plays a role in the citric acid cycle, which is the process by which biochemical energy is generated during aerobic respiration. Biotin not only assists in various metabolic reactions but also helps to transfer carbon dioxide. Biotin may also be helpful in maintaining a steady blood sugar level.

    Biotin deficiency is rare, as, in general, intestinal bacteria produce an excess of the body's recommended daily requirement."

27. H = Histidine; P = Proline; Q = Glutamine.

28. Mansy, S.S., Zhang, J., Kümmerle,R., Nilsson, M., Chou, J.J., Szostak, J.W. and Chaput, J.C., Structure and Evolutionary Analysis of a Non-biological ATP-binding Protein, *J. Mol. Biol.* **371**:501–513, 2007.

29. Mansy *et al.*, ref. 28, p. 502.

30. Mansy *et al.*, ref. 28, p. 504.

31. Lo Surdo, ref. 2, p. 382.

32. As determined from the comments included in the structure data file, www.rcsb.org/pdb/explore/images.do?structureId=1UW1.

33. www.cathdb.info/cathnode/3.

34. Garbuzynskiy, S.O., Melnik, B.S., Lobanov, M.Y., Finkelstein, A.V. and Galzitskaya, O.V., Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray and NMR-resolved protein structure, *Proteins: Struct. Funct. Genet.* **60**:139–147, 2005.

35. en.wikipedia.org/wiki/Zinc_finger.

36. Whitford, D., *Proteins: Structure and Function*, John Wiley & Sons, Hoboken, NJ, p. 262, 2008.

37. Wilson *et al.*, ref. 23, p. 3754.

38. Wilson *et al.*, ref. 23, p. 3755: "Structures of HPQ-containing peptides bound to streptavidin have been determined by x-ray crystallography. The HPQ motif inserts itself into the biotin-binding cleft and forms hydrogen bonds and hydrophobic interactions with streptavidin."

39. Chaput, J.C. and Szostak, J.W., Evolutionary Optimization of a Nonbiological ATP Binding Protein for Improved Folding Stability, *Chemistry & Biology* **11**:865–874, 2004.

40. Smith *et al.*, ref. 4, p. 2.

41. en.wikipedia.org/wiki/Zinc.

42. Whitford, ref. 36, p. 25.

43. Emsley, J., '*Zinc*', *Nature's Building Blocks: An A–Z Guide to the Elements*, Oxford University Press, Oxford, pp. 499–505, 2001.

44. www.lenntech.com/periodic/water/zinc/zinc-and-water.htm.

45. Fukai, R. and Huynh-Ngog, L., Chemical Forms of Zinc in Sea Water, *J. Oceanographical Society of Japan*, **31**:179–191, 1975.

46. en.wikipedia.org/wiki/Abundance_of_elements_in_Earth%27s_crust.

47. Whitford, ref. 36, p. 7.

48. Gil, R., Sabater-Muñoz, B., Latorre, A., Silva, F.J. and Moya, A., Extreme genome reduction in Buchnera spp.: Toward the minimal genome needed for symbiotic life, *PNAS* **99**(7):4454–4458, 2002.

49. en.wikipedia.org/wiki/Puromycin. "Puromycin is an aminonucleoside antibiotic, derived from the *Streptomyces alboniger* bacterium, that causes premature chain termination during translation taking place in the ribosome. Part of the molecule resembles the 3′ end of the aminoacylated tRNA. It enters the A site and transfers to the growing chain, causing premature chain release. The exact mechanism of action is unknown at this time, but, the 3′ position contains an amide linkage instead of the normal ester linkage of tRNA, the amide bond makes the molecule much more resistant to hydrolysis and thus causes the ribosome to become stopped."

50. Wolfe, S.L., *Introduction to Cell and Molecular Biology*, Wadsworth Publishing Company, Belmont, California, p. 466, 1995.

51. Wolfe, ref. 50, p. 467.

52. Roberts, R.W. and Szostak, J.W., RNA-peptide fusions for the *in vitro* selection of peptides and proteins, *Proc. Natl. Acad. Sci. USA* **94**:12297, 1997.

53. Roberts and Szostak, ref. 52, p. 12301.

54. Roberts and Szostak, ref. 52, p. 12300.

Royal Truman has bachelor's degrees in chemistry and in computer science from State University of New York; an MBA from the University of Michigan (Ann Arbor); a Ph.D. in organic chemistry from Michigan State University; and a two-year post-graduate 'Fortbildung' in bioinformatic from the Universities of Mannheim and Heidelberg. He works in Germany for a European-based multinational.

The land Dyaks of Borneo claim they came from a fish, while the Kayans claim they came from a tree.[26]

Practically every animal in creation, including the apes, can claim to be someone's ancestor! Perhaps the most bizarre account is that of the Samoans, who believe that at first two men were developed out of two grubs who had been transformed from the remains of a rotting convolvulus. It happened that one man died and the great god, Tangaloa, changed the body into a woman and brought her to life and these two humans subsequently became the parents of mankind.[27] The Greeks were represented by the account of Empedocles (5th century BC), who has shapeless lumps of earth and water thrown into the subterranean fires to form monsters that were gradually eliminated until the existing species of animals and humans remained.[28] Note the almost total absence of concordance among the evolutionary accounts in comparison with the remarkably good concordance among the creation accounts. Surely, any unbiased jury faced with this kind of evidence would declare that God's creation account is likely to be closer to the truth. Nevertheless, today's anthropology has chosen to reject the book of Genesis in favour of Darwin's *Origin of Species*.

21st-century medicine faces the creation account

In 2011 Dr Parvez Haris,[29] head of the Environmental Health group at De Montfort University (Leicester, UK), was called upon to conduct a public health survey of the Sikor or clay sold in the ethnic pharmacies of the UK. The consumers are invariably pregnant women who may eat as much as 500 g of baked clay per day. The declared object being to ensure a healthy child. Haris' study showed high levels of arsenic, lead and cadmium in the Sikor samples, representing a serious health risk to both mothers and children.

The practice of eating baked clay is known as geophagy; it is very ancient and widespread in many parts of the world. It is particularly common in rural Africa and Asia, such as India and Bangladesh. Could this practice be related to God's account of creating Adam from clay of the ground—given as 'dust of the ground' in Genesis 2:7?

References

1. Toynbee, A., *A Study of History*, Oxford University Press, vol. 1, p. 424, 1934.
2. Perry, W.H., *The Children of the Sun*, Methuen and Co., London, 1923.
3. Smith, G.E., *Human History*, W.W. Norton, NewYork, 1929.
4. Frazer, J.G., *Folklore in the Old Testament*, Macmillan, London, 3 vols., 1918.
5. Frazer, ref. 4, vol.1, p. 29.
6. Frazer, ref. 4, vol.1, p. 12.
7. Frazer, ref. 4, vol.1, p. 18.
8. Frazer, ref. 4, vol.1, p. 24.
9. Oldfather, C.H. (Ed. and trans.), *Diodorus* Siculus, William Heinemann, London, vol. 1, book 1, 1933; footnote in 1960 ed.
10. Griffiths, J.G. (Ed. and trans.), *Plutarch's De Iside et Osiride*, University of Wales, pp. 235 & 551, 1920. For more see: Griffiths, J.G., Human sacrifices in Egypt; the classical evidence; in: *Annales du Service des Antiquités d'Egypte* (ASAE), pp. 409–423, 1948.
11. Hastings, J. (Ed.), *Encyclopedia of Religion and Ethics*, T. and T. Clark, Edinburgh, vol. 6, p. 845, 1913; pp. 40–867 are devoted to the anthropology of human sacrifice.
12. Frazer, ref. 4, vol.1, p. 6.
13. Frazer, ref. 4, vol.1, p. 14.
14. Frazer, ref. 4, vol.1, p. 15.
15. Frazer, ref. 4, vol.1, p. 22.
16. Frazer, ref. 4, vol.1, p. 23.
17. Frazer, ref. 4, vol.1, p. 28.
18. Frazer, ref. 4, vol.1, p. 8.
19. Frazer, ref. 4, vol.1, p. 9.
20. Frazer, ref. 4, vol.1, p. 10.
21. Frazer, ref. 4, vol.1, p. 11.
22. Frazer, ref. 4, vol.1, p. 25.
23. Darwin, C., *The Descent of Man*, John Murray, p. 29, 1871.
24. Frazer, ref. 4, vol.1, p. 30.
25. Frazer, ref. 4, vol.1, p. 31.
26. Frazer, ref. 4, vol.1, p. 34.
27. Frazer, ref. 4, vol.1, p. 40.
28. Frazer, ref. 4, vol.1, p. 44.
29. Haris, P., Health Matters, www.dmu.ac.uk/news_events/news/current/110131_Sikor.jsp, 14 February 2011.

Ian Taylor graduated in metallurgical engineering in the UK, emigrated and was employed by the research laboratories of the Aluminum Company of Canada. He obtained patents for industrial processes. Saved in 1974, he spent five years in daily Christian television production, where he became producer/writer for the Crossroads science documentary series. He is the author of *In the Minds of Men*, now in its sixth updated edition (2008). In 1996 he joined the former Bible-Science Association, and became the voice of their Creation Moments daily radio program; He is still the radio voice of the program, now heard throughout North America.

Instructions to Authors

Journal of Creation is dedicated to upholding the authority of the 66 books of the Bible, especially in the area of origins. All our editors adhere to the *Creation Ministries International* (CMI) Statement of Faith and most papers will be designed to support this. Rarely, other papers may be accepted for publication on merit so that certain issues can be clarified. The views expressed in the papers are those of the authors and not necessarily those of CMI.

Because the scope of this journal is broad, we welcome articles on any topic that is relevant to biblical creation. This includes the sciences such as geology, chemistry, biology, astronomy, etc., and also archaeology, theology, hermeneutics, biblical history, geography, linguistics, ethics and society, philosophy, law, etc. Potential authors should familiarise themselves with the journal and its position and style before submitting articles. Authors should also indicate if their manuscript has been submitted elsewhere, address previous articles on the topic, and ensure the work of others is properly acknowledged.

Word length: Shorter articles are more likely to be read so preference will be given to publishing them. All articles, including letters, may be edited for brevity and clarity. Perspectives: 1,000–2,000 words; Book reviews: 1,500–3,000 words, but please enquire first. You may be asked to scan the cover of the book; Letters to the editor: 1,000 words. We will publish critical letters on articles already published, but a reply will usually follow the criticism; Papers, Overviews, Countering the Critics, Viewpoints and Essays: <5,000 words.

Articles must be written clearly and concisely and should be focused on only one topic/subject. The most readable articles are those with an average sentence length of about 17 words, with one idea per sentence. Do not use too many big or extra words such as "in spite of the fact that" when "although" would do. Please use first person sparingly and do not use "this present writer", which is verbose false modesty and ambiguous. Care with spelling is crucial, and British spelling generally applies. Personal invective or similar comments against others are not acceptable.

Specialist technical terms and abbreviations should be defined the first time they are used. If many technical terms are used, these should be included in a glossary. SI units should be used in scientific papers. Words in non-Latin alphabets (e.g. Hebrew, Greek and Russian) must use Unicode characters and be accompanied by a transliteration, also in Unicode characters (Unicode should avoid errors when the file is transferred to a publishing program or HTML).

Abstract: All articles except Perspectives, Letters and Book Reviews should be preceded by an Abstract, which should not exceed **200 words** and must be without abbreviations and reference citations. The Abstract should summarise the key points of the paper and be comprehensible to readers before they have read the paper.

References should be indicated in the main text by superscript numbers in sequence and then listed in numerical order at the end of the text (End Notes). Full details of all references are required, including all authors and their initials, the full title of the paper or book, the full title of the journal or its accepted abbreviation, the volume number, the page number(s), the editor(s) of the book or proceedings (if applicable), the publishers and place of publication (in the case of a book), and the year of publication. If a citation is repeated, then the same superscript number should be used (a cross reference).

Quotes must be verbatim, with omissions clearly shown by ellipsis (…). Even erroneous grammar and spelling in the original must be reproduced, and indicated by [*sic*]. Any additions or explanations within a quote need to be placed in square brackets [ ]. Primary sources are preferred, but if a secondary source must be used, the reference format must be [primary source], as cited in [secondary source]. For internet URLs the date last downloaded should be included.

Tables should be embedded in a Microsoft Word document, with brief, self-explanatory titles, and captions if an explanation is essential to understanding the table. Tables should be numbered according to their sequence in the text and all should be referred to in the text, e.g. (see table 1).

Graphics should be supplied electronically on CD or by email to journal@creation.info as 300dpi *.TIF, *.EPS, *.PSD, *.AI, *.JPG, *.PDF or *.BMP files, not embedded into a document file. They can be sent in colour or greyscale, but will appear only in greyscale in the journal. Graphics should have a minimum size of 80 mm x 80 mm when printed (preferably 1024 x 768 pixels). Photographs are acceptable provided they have good contrast and intensity, and are submitted as sharp, glossy copies or as 35 mm slides. Computer print-outs are not acceptable.

Figures should be numbered according to their sequence in the text. References should be made in the text to each figure. In planning images and figures, the page format and column widths of the journal should be kept in mind and should allow for the possibility of reduction. Each illustration should have a self-explanatory caption. The captions should be collected together at the end of the article document file. If graphics are not provided with a manuscript, authors may be asked to submit suggestions and possible sources of non-restricted material.

Copyright ownership for each and every graphic must be specified and any relevant permissions and or credit requirements given in writing. If you **are not** the copyright owner, please supply written permission (preferably with an original signature) from the copyright owner. If you **are** the copyright owner, your submission of a graphic will be taken to grant us permission to use the image in the journal and also any other publications or on our website, unless you specify otherwise.

Biography: Papers, Overviews, Countering the Critics, Forums, Viewpoints, Research Notes and Essays should include a biography of no more than 100 words for each author. It should include the authors' qualifications and experience, and involvement in creationist work.

Please also:

- send documents in Rich Text (*.rtf) or Word (*.doc) format, not Publisher (*.pub) format. Excel (*.xls) is also acceptable for complicated tables.
- type left-justified without hyphenation except for compound words
- input text continuously; only insert hard returns at the end of sections, lists, etc.
- use single spaces after each sentence within a paragraph
- use Times New Roman 10.5 font and use the word processing formatting features to indicate bold, italics, Greek, Hebrew, maths, superscript and subscript characters
- submit complex mathematical formulas or equations using MathType according to the guidelines in our "Equations layout guide", which can be obtained by contacting us
- insert captions for graphics at the end of your manuscript
- type references in the correct order and style of the journal using the word processing "Endnotes" and "Cross-references" features to link references to the text correctly.

Photographs and/or CDs may be sent to:

journal@creation.info

The Editor,
Journal of Creation
Creation Ministries International
P.O. Box 4545
Eight Mile Plains QLD 4113
AUSTRALIA

NOTE: Papers prepared according to these instructions are more likely to be considered.