THE "INTRUDER" HYPOTHESIS AND SELECTION AGAINST ARGININE

by

Thomas H. Jukes

Space Sciences Laboratory
University of California, Berkeley, Calif. 94720

Received March 14,1974

SUMMARY

The "intruder" hypothesis for the presence of more arginine codons in the genetic code than are needed leads to the expectation for selection against arginine in protein synthesis. This selection is therefore a consequence of the intruder hypothesis rather than a substitute for it, as claimed by Wallis (1974).

Wallis (1974) has argued against the proposal (Jukes, 1973) that arginine may have been an "evolutionary intruder" into protein synthesis. In this proposal, I pointed out that the unusual basicity of arginine ($pK_3^1 = 12.5$), resulting from the distinctive properties of its guanido group, led to arginine being rejected to some extent in evolution. I noted that more lysine was used than arginine in most proteins, despite the fact that lysine has 2 codons and arginine has 6. The observations lead to two possibilities. Selection against arginine may have been a "rule of evolution" starting with the first formation of a primitive genetic code (which Wallis is apparently suggesting), or arginine may have entered the code as an "intruder", in which case, selection against arginine would be expected. Or, as Dr. Richard Holmquist has pointed out to me, selection against arginine might be a result of it becoming "obsolescent" rather than it being an "intruder." According to this concept, living systems would have evolved away from their former use of proteins that were high in arginine.

The intruder hypothesis is consonant with the idea that various genetic codes preceded the present code, which became universal as the result of a "frozen accident" (Ohno, 1973), in which a single organism, from which all existing organisms are derived, gained ascendancy over its competitors at

some point in time. The contrary hypothesis would be that the code evolved to reach an optimum composition, and that this composition was preserved during evolution. But this latter hypothesis is incompatible with the fact that the number of arginine codons is so large that selection against arginine takes place in protein synthesis. For this reason, I suggested the intruder hypothesis as superior to the optimum code concept in explaining the observed selection against arginine.

The arginine:lysine ratio of a protein is not correlated with its rate of evolution, as noted by Wallis (1974). This ratio will depend upon the extent to which arginine and lysine are needed or tolerated in the structure and special functions of the protein. Histones evolve very slowly because of strong restraints on their structure (Delange et al, 1969), but,because of their function in binding to DNA, they have a high content of basic amino acids. Other proteins, such as ferredoxins and caseins, are predominantly acidic, presumably for functional reasons. The only known protein that may be able to evolve quite freely side-by-side with a homologous protein that is obviously under greater restraint is the example of the variable and constant regions of immunoglobulins, which I cited (Jukes, 1973). The other examples in Wallis' Table 1 do not have "variable" and "constant" counterparts for comparison. His values, taken from Dayhoff's "Atlas", for their "rates of evolution",are open to question because of the paucity of data, except for immunoglobulins, hemoglobin, myoglobin and cytochrome c. Indeed, he gives a "rate of evolution" for trypsinogen, of which only one sequence has been determined, and he states that the constant region of immunoglobulins has evolved more rapidly than the variable region, which seems difficult to believe.

Wallis advances the following arguments for rejecting the "intruder" hypothesis: (i) the urea cycle is the most obviously important known role of arginine outside of protein synthesis; this is largely confined to (some) eukaryotes, but arginine is found in proteins of prokaryotes, thus suggesting

that the urea cycle appeared more recently than the use of arginine in protein synthesis. But prokaryotes of the type currently existing may or may not have preceded eukaryotes. A role for arginine apart from protein synthesis may have since disappeared, for the evolutionary process eliminates outmoded biological phenomena. (ii) Wallis says that the intruder hypothesis apparently assumes that:

> *"the initial function of arginine was not its role in protein synthesis, and that the advantages gained from its availability for this original function outweighed disadvantages which were encountered as it replaced ornithine in proteins."*

No such assumption was made; I suggested that arginine could have conveyed certain special advantages in proteins, but this role did not require that arginine should have as many as six codons. Therefore lysine was drawn upon, as a substitute for the displaced ornithine, to supply the requirement for a basic amino acid with properties less unusual than those of arginine.

(iii) Wallis states that:

> *"the six arginine codons are served by more than one type of tRNA, and it would be remarkable if arginine had succeeded in capturing all six from ornithine."*

Six tRNAs are not needed for six codons. "Wobble pairing" would enable three anticodons (UCG, GCG and UCU) to suffice for the six arginine codons. Different tRNAs with different anticodons for the same amino acid can be charged by a single tRNA aminoacyl ligase (Blank and Söll, 1971). If all the postulated ornithine tRNAs had the same affinity site for the ligase system, it would be possible for arginine to "capture" all of them simultaneously.

Wallis then lists "alternative explanations" for the relative scarcity of arginine. He says:

> *"most residues in most proteins do appear to be strongly conserved, either they show no changes during evolution or they show only conservative changes."*

There are only three families of proteins for which enough sequences are available so that the question can be examined to any extent; the cytochromes c, the globins and the immunoglobulins. All three have many sites at which

there are substitutions of several different amino acid residues, and often these are not necessarily "conservative changes." Indeed, as more sequences come to light, it becomes evident that many so-called "conservative" sites can undergo numerous changes, and that some "invariant" residues are actually subject to replacement. It is also incorrect to imply, as Wallis does, that the amino acid composition of proteins does not reflect the nature of the genetic code; indeed, in the next paragraph Wallis contradicts this statement himself. He then states that "at least three possible bases for selection against arginine can be proposed as alternatives to selection resulting from the replacement of ornithine by arginine." These could be the consequences of selection following the replacement of ornithine by arginine. One of the three is the possible role of arginine codons in the control of protein synthesis. The evidence listed for this could also apply to one or more of the codons for other amino acids which have several different codons. Wallis states, "it is interesting that prevailing evidence from bacteriophage RNA sequences and from hemoglobin mutants provides many instances of the use of arginine codons of the group CGX, but no definite evidence for use of AGG or AGA." However, AGG and AGA (AGR) codons have been found in viral RNA sequences aligned with polypeptide sequences of viruses (Contreras et al, 1973). Hemoglobin mutants show the use of arginine codons of the group CGN (CGX) (King and Jukes, 1969), but some of these mutations can be interpreted either in terms of CGN or AGR. Examples are mutations between arginine and glycine and between arginine and serine. There are only two mutations where AGR codons can be definitely established; these are between arginine and threonine or lysine. The first of these occurs in E. coli and the second in tobacco mosaic virus proteins. If "tRNA recognizing AGA and AGG represents only a tiny fraction of the total unfractionated" arginine tRNA in E. coli, this could result from selection against arginine codons.

In Wallis' final paragraph he says "selection against arginine may have resulted in selection for lysine", which is what I said (Jukes, 1973). He

also says that lysine is "frequently found in modified form" in proteins. However, this is comparatively rare.

In summary, the article by Wallis discusses various mechanisms for selection against arginine, a procedure which would be the consequence of the "intruder" hypothesis or of arginine becoming "obsolescent", as suggested by Holmquist.

ACKNOWLEDGMENT

This work is supported by NASA grant NGR 05-003-460 to the University of California, Berkeley. I thank Dr. Richard Holmquist for suggestions.

REFERENCES

Blank, H. U. and Söll, D. (1971) J. Biol. Chem., 246:4947.
Contreras, R., Ysebaert, M., Min Jou, W. and Fiers, W. (1973) Nature New Biology, 241:99.
Delange, R. J., Fambrough, D. M., Smith, E. L. and Bonner, J. (1969) J. Biol. Chem., 244:319.
Jukes, T. H. (1973) Biochem. Biophys. Res. Commun., 53:709.
King, J. L. and Jukes, T. H. (1969) Science, 164:788.
Ohno, S. (1973) Nature, 244:259.
Wallis, M. (1974) Biochem. Biophys. Res. Commun., 56:711.