Where does the message begin?

SIR—The remarkable accomplishment of cloning and sequencing the gene responsible for the X-linked form of the phagocytic disorder chronic granulomatous disease (X-CGD) on the basis of its chromosomal location alone, without reference to any protein, has recently been reported in these pages by Royer–Pokora *et al.*[1]. With the amino acid sequence predicted from the gene sequence, it should now be possible to identify and characterize the previously unknown protein responsible for this disorder. However, it seems possible that the true amino terminus of the protein has not been identified.

Royer–Pokora *et al.* conclude that translation of the X-CGD messenger RNA probably initiates at the second in-frame AUG codon, at position 322, because it occurs in a sequence that more closely resembles the consensus sequence for functional initiator codons, defined by Kozak[2] on the basis of observed sequences, than does that at the first AUG at position 208. The primary difference between the two is that the AUG at position 322 has an A nucleotide three positions upstream, as do 79 per cent of functional initiation codons, whereas at the first AUG there is a U, which is hardly ever observed at this position[2]. On the other hand, 95 per cent of known messenger RNA molecules have translation initiated at the most proximal AUG codon[2], and the importance of the most proximal position has been demon-strated directly[3]. The basis of the nucleotide preference observed three nucleotides upstream is not known. Therefore, the choice of neither initiation codon is compelling from just these considerations.

The reason for suggesting that translation of the X-CGD messenger RNA initiates at the first AUG is that the 38 extra amino acid residues indicated by the messenger RNA sequence are not random. In particular, the first 20 residues closely resemble the signal peptides that have been implicated in the co-translational passage of proteins into the endoplasmic reticulum as the first step in targeting the protein for secretion from the cell, insertion into membranes, or sequestration into at least some cytoplasmic compartments[4-6]. The pertinent properties are that this sequence;

Met-Leu-Ile-Leu-Leu-Pro-Val-Cys-Arg-Asn-Leu-Leu-Ser-Phe-Leu-Arg-Gly-Ser-Ser-Ala-

is positively-charged, has predominantly hydrophobic amino acids from residues 2 to 15, and ends with five polar residues typical of signal peptide cleavage sites. The only unfavourable aspects of the sequence are the Arg and Asn residues at positions 9 and 10, respectively, within the hydrophobic segment, but such residues are observed at low frequencies in similar positions in known signal peptides[7]. The predictive scheme of Von Heijne[7] considers the frequencies with which the 12 amino acids proximal to the signal peptide cleavage site, plus the two distal, are observed in known signal sequences. The relative likelihood of the first 20 residues constituting a signal peptide, with cleavage after Ala 20, was calculated to be + 5.6, which is well within the range for known signal peptides and outside that found for the terminal amino-acid sequences of cytoplasmic proteins[7]. The first eight residues are not considered in this calculation but are very like those in known signal peptides.

The nature of the X-CGD disorder is consistent with the protein responsible being situated in granules of phagocytes[8]. Proteins designated for such cell sites are often synthesized with signal sequences and are directed to the appropriate cell compartment after entering the endoplasmic reticulum[9-11].

THOMAS E. CREIGHTON
Medical Research Council,
Laboratory of Molecular Biology,
Cambridge, CB2 2QH, UK

1. Royer–Pokora, B. *et al. Nature* **322**, 32–38 (1986).
2. Kozak, M. *Nucleic Acids Res.* **12**, 857–872 (1984).
3. Sherman, F., Stewart, S.W. & Schweingruber, A.M. *Cell* **20**, 215–222 (1980).
4. Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496–1500 (1980).
5. Perlman, D. & Halvorson, H.O. *J. molec. Biol.* **167**, 391–409 (1983).
6. Von Heijne, G. *J. molec. Biol.* **184**, 99–105 (1985).
7. Von Heijne, G. *Nucleic Acids Res.* **14**, 4683–4690 (1986).
8. Tauber, A.I., Borregaard, N., Simons, E. & Wright, J. *Medicine* **62**, 286–309 (1983).
9. Faust, P.L., Kornfeld, S. & Chirgwin, J.M. *Proc. natn. Acad. Sci. U.S.A* **82**, 4910–4914 (1985).
10. Shibahara, S. *et al. Nucleic Acids Res.* **14**, 2413–2417 (1986).
11. Lobe, C. *et al. Science* **232**, 858–861 (1986).

SIR—We appreciate the comments and suggestions from Creighton[1] and from Ashworth *et al.*[2] regarding our predictions for the protein product encoded by the gene responsible for X-linked chronic granulomatous disease[2].

Creighton suggests that a non-consensus ATG at position 208 of our cDNA sequence might be the correct translation initiation site (rather than the consensus ATG at position 322 we favoured) and that the additional N-terminal extension encodes a signal peptide based on a predictive scheme of von Heijne[4]. As stated in our paper, no certain choice between these ATGs can be made from cDNA sequence alone. Our analyses of the sequence, albeit without the recent approach of von Heijne, led us to favour the ATG at 322. Creighton's proposal needs to be very carefully considered, but in light of our recent finding (R. Brown and S.H.O., unpublished data) that an intron interrupts the alanine codon he suggests as a signal peptide cleavage site.

Whatever the final resolution, the uncertainty in ATG assignment does not materially affect experimental approaches to identification of the protein *in vitro*. For one, the size difference between the proteins predicted from initiation at either ATG (minus the signal peptide if ATG 208 were used) is small. In addition, we are obliged to examine all cellular fractions to localize the protein to a specific compartment. Finally, initial results with antisera directed against a portion of the predicted amino acid sequences identify an *in vivo* protein located in membrane-rich fractions of granulocytes (M. Dinauer and S.H.O. unpublished data). Expression of transferred cDNAs will permit definitive assignment of the initiator codon and the possible role of the signal peptide proposed by Creighton in the targeting of the protein within the cell.

Although the suggestion of Ashworth *et al.*[2] that a region near the C-terminus of the predicted protein resembles a haem-binding site of a cytochrome P450 is intriguing, the significance of the limited resemblance they note is unclear.

STUART H. ORKIN
Division of Hematology, Children's Hospital,
Howard Hughes Medical Institute,
Harvard Medical School,
Boston, Massachusetts 02115 USA

1. Creighton, T. *Nature* **324**, 21 (1980).
2. Ashworth, A., Shephard, E.A.& Phillips, I.R. *Nature* **322**. 599 (1986).
3. Royer Pokora, B. *et al. Nature* **322**, 32–38 (1986).
4. Von Heijne, G. *Nucleic Acids Res.* **14**, 4683–4690 (1986).

Occam and mankind's genetic bottleneck

SIR—In defending their support for and amplification of the view[1] that non-African human populations are descended from African populations via an emigrant stock that was of very small size, Jones and Rouhani[2] have exposed the weakness of their argument. It is true that such a bottleneck is an inevitable implication of their simple model. However, that model assumes that the β-globin non-coding sequences of DNA on which their argument is based are unaffected by selection, either directly or by linkage ('hitch-hiking'). This assumption is defended on two grounds. The first is that neutrality of non-coding sequences is implicit in most hypotheses of molecular evolution. But there is actually no evidence for this fashionable assumption. It is adopted merely because of the prejudices of those advancing the hypotheses, because it allows phylogenetic conclusions to be reached in molecular biology laboratories rather than at palaeontologists' benches, and possibly because it makes the models mathematically easier.

©1986 Nature Publishing Group