

25th September, 1957

Dr. A. Lwoff,
Institut Pasteur,
25 Rue du Docteur Roux,
Paris XV^e.
France.

Dear Lwoff,

You are quite right. We have a skeleton in our cupboard. We have thought about the problem quite a bit, but have arrived at no satisfactory solution. The following remarks outline some of our ideas:

(1) We have done preliminary work on a quadruplet code. We can make a non-overlapping comma-less code using one DNA chain which makes sense, and the complementary DNA chain (when read backwards) is everywhere nonsense. The maximum for which we can code is more than 20 and less than 27 amino acids, but we don't know the exact number. We don't like this because it is inelegant.

(2) There are three distinct ways of "pairing" for a given ABCD code. These are A with B and C with D
A with C and B with D
A with D and B with C

If you take the first triplet code in our paper, and make the last of the above interchanges, and also read backwards (because the two DNA chains run in opposite directions) you will find that all of the 20 allowed triplets are turned into nonsense triplets e.g. ACB becomes DBC backwards; that is, CBD, which is nonsense. Unfortunately, however, one can get accidental bits of sense where these nonsense triplets (on the second DNA chain) overlap.

I tried to see if we could get away with forbidding those amino acid sequences which allowed two adjacent accidental triplets to be formed on the reciprocal chain, but I can prove vigorously that this rule forbids a sequence of three identical amino acids, and unfortunately such sequences are known.

(3) Though not stated in our paper one needs a code for "Start chain" and "stop chain". These can be any number of A's e.g. AAAAAA We are trying to develop this idea, but so far we have nothing worth repeating.

As to the suggestions in your letter, the data on the sequences of the insulin molecule suggest that one can change only one amino acid, whereas if each base of a base-pair controlled a different amino acid, a change in base-pair would, in most schemes, change two amino acids. Your second idea also leads to difficulties since by forming the triplets in this way you reduce very much the information a triplet can carry. In any case you would have the problem of reading it both ways. In fact this difficulty is inherent in the DNA structure and every code has⁵ faced it.

My own feeling at the moment is that of present ideas the triplet code is the best, and that it needs some trick modification or addition to get over the double-helix difficulty, rather than a radical change.

We had a most enjoyable visit from Francois and have been stimulated by him to develop a new idea about (phage-type) recombination. Briefly the idea is that recombination only takes place when there is a "not-base" in the structure i.e. a modified base which cannot pair. This seems a most promising postulate.

Yours sincerely,

F.H.C. Crick