Elucidation of the amino acid sequence of the Insulin B-chain (community effort)

Sanger & Tuppy (1951)

The first problem of Problem Set 2 reads:

> **PS2.1.** Use only the results of Sanger and Tuppy (1951) [Biochem J 49:463-481] to deduce as much of the structure of insulin you can. Do this as if it were a geometric proof, appealing to lines within the tables (axioms) and truths you derive from them (theorems). For example:
>
> | | Assertion | Justification |
> |---|---|---|
> | **A.** | Thr-Pro* | Table 6, Line 8 |
> | **B.** | Thr-(Ala,Lys,Pro)¶ | Table 9, Line 6 |
> | **C.** | Only one Pro | Table 14 |
> | **D.** | Thr-Pro-(Ala,Lys) | **A**+**B**+**C**† |
>
> * Meaning "The dipeptide N-Thr-Pro-C lies somewhere in the insulin polypeptide chain". The form N-XxxYyy-C means that the amino acids are read from amino end to carboxyl end.
>
> ¶ Meaning "A tetrapeptide somewhere in insulin begins N-Thr and is immediately followed by Ala, Lys, and Pro in some unknown order"
>
> † Meaning "The assertion on this line follows from the assertions on lines A, B, and C"

This essentially asks you to take all the data from Sanger & Tuppy (1951) and recreate their chain of reasoning that led to the sequence of insulin (to the extent they got there). This is entirely do-able, but takes a while. Big scientific problems don't take a while – they take forever, for an individual – so they're generally solved by communities. Even though **PS4.1** isn't too big of a scientific problem, we can still approach it in that spirit.

Accordingly, I've divided you into working groups, giving each person an area of expertise. Each person is associated with an amino acid. That person should consider all experiments from Sanger & Tuppy (1951) that bear on the amino acid, deducing all that's possible from the results, in the form shown in Problem 1, above. When you are satisfied by a non-trivial deduction (the last line of your chain of reasoning), publish it, by posting it to the community bulletin board.

You'll be able to reach increasingly sophisticated deductions (perhaps even the entire sequence of the insulin B-chain!) as you combine your results with those of others in your group. You might also make some use of results from other groups, but beware! There might be all sorts of garbage posted to the community bulletin board – not intentionally, of course, but people do make mistakes. If you accept the results of others uncritically, you may accumulate mistakes and reach erroneous conclusions yourself.

How can you avoid contaminating your growing chain of deductions?