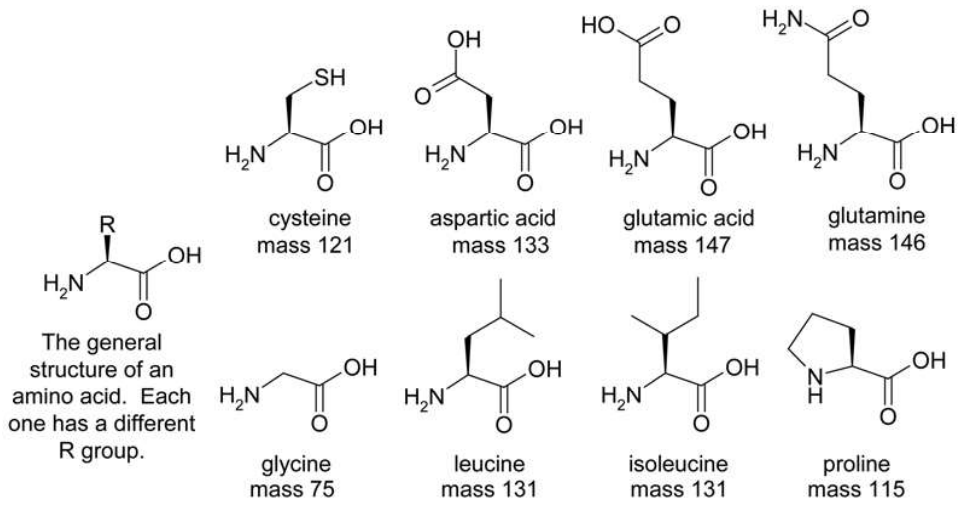
5. This question is about toxins from cone snails

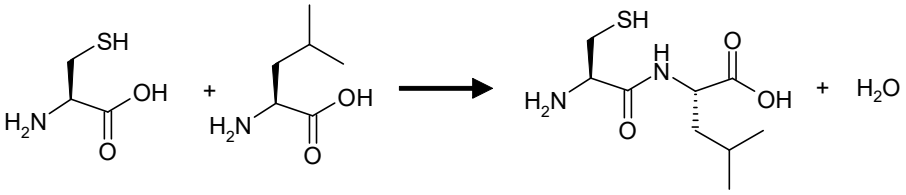
Cone snails are predators that use venom to capture prey. The toxic species in the venom are polypeptides. Cone snail toxins are of pharmaceutical interest as starting points for the development of new anaesthetics. A number of research groups are working towards identifying the amino acid sequences of new cone snail toxins.

The cone snail *Conus textile*

Polypeptides are polymers of amino acids; the structures and relative masses of some amino acids are shown below:

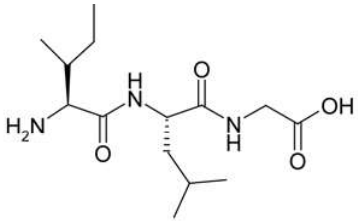


When amino acids form a polypeptide, an amide bond is made and water is lost:

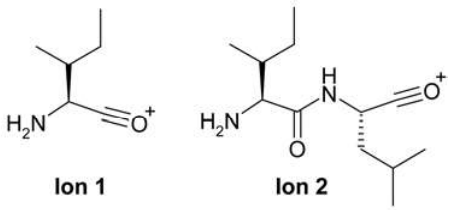


In biological systems the function of a polypeptide depends on the order of the amino acids in the sequence. By convention, a polypeptide is drawn starting with the amine group on the left, hence the sequence of the polypeptide shown above is cysteine-leucine NOT leucine-cysteine.

Polypeptides are often sequenced using mass spectrometry. In a mass spectrometer the polypeptide breaks into fragments with the amide bonds being the most likely to be broken. By comparing the masses of the different ions formed it is possible to work out the amino acid sequence. The major ions seen in the fragmentation of an isoleucine-leucine-glycine polypeptide are shown below:



isoleucine-leucine-glycine polypeptide



Fragment ions seen in the mass spectrum

In all parts of this question you should use the mass of the most common isotopes of each of the elements: 12 for C, 14 for N, 16 for O and 1 for H.

(a) **i)** What is the mass of the isoleucine-leucine-glycine polypeptide?

ii) What is the mass of **Ion 1**?

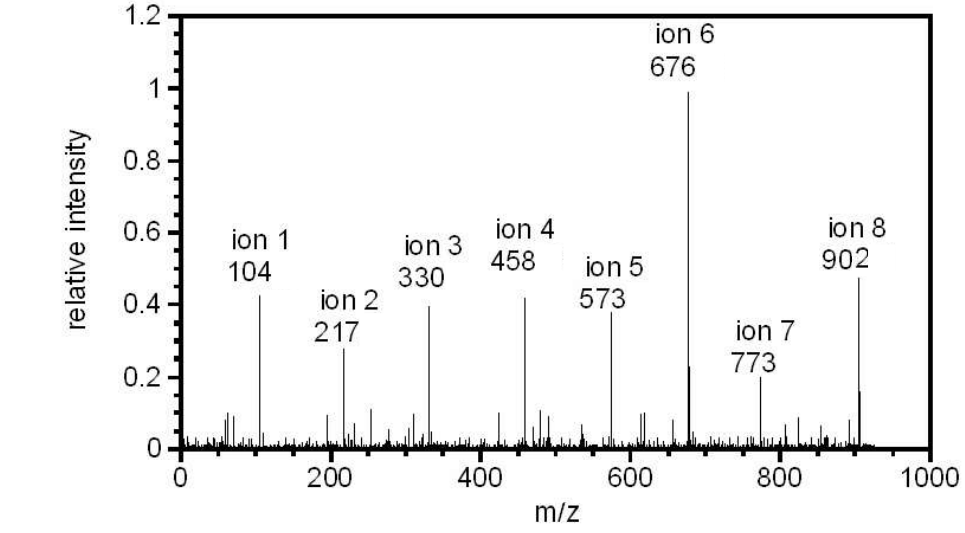
iii) What is the mass of **Ion 2**?

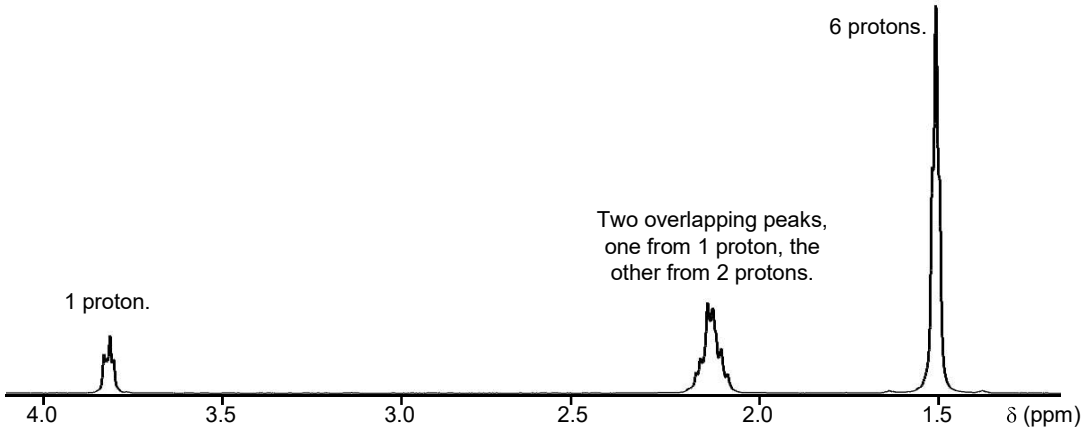
A polypeptide, **X**, of mass 976 was isolated from a cone snail. It was found by chemical analysis to have the following amino acid composition:

2 x cysteine, 1 x aspartic acid, 1 x glutamic acid, 1 x glutamine, 1 x glycine, 1 x isoleucine, 1 x leucine and 1 x proline.

(b) How many unique polypeptide sequences can be formed using all these amino acids?

The fragmentation mass spectrum of **X** and the ^{1}H NMR spectrum of the third amino acid in the sequence are shown below. Under the conditions used for the NMR spectrum no peaks are seen for the NH_2 and COOH protons.





(c) The last two amino acids in the sequence are glutamic acid-glycine. What is the sequence of the first 7 amino acids? [The structures and masses of the amino acids are shown at the top of the previous page.]