1) Determine the amino acid sequence an unknown polypeptide based on the following observations. Show the reasoning behind your conclusion. (7 pts)

i) CNBr (cyanogen bromide) cleaves the polypeptide into two fragments with compositions: (HSer, Ala) and (Ala, Gly, 2 x Lys, Pro).

ii) Trypsin cleaves the polypeptide into two fragments with compositions: (Ala, Lys, Gly, Met) and (Lys, Ala, Pro).

iii) Treatment of the fragment containing (Lys, Ala, Pro) with FDNB followed by hydrolysis with 6N HCl yields DNP-Ala as one of the products.

\[\text{CNBr cleaves after Met (converting Met to Hser)} \rightarrow \text{Ala-Met.}\]

\[\text{Trypsin cleaves after Lys or Arg} \rightarrow \text{Ala-Met-Gly-Lys}\]

\[\rightarrow \text{Ala-Lys-Pro or Ala-Pro-Lys}\]

Hence, the sequence is Ala-Met-Gly-Lys-Ala-Lys-Pro or Ala-Met-Gly-Lys-Ala-Pro-Lys.

2. a) Acid hydrolysis gave equimolar amounts of Lys, Glu, Phe, Ala, NH₃. FDNB followed by hydrolysis generated DNP-alanine and ε-DNP-lysine.

NH₃ implies Glu was originally Gln.

DNP-alanine implies Ala is at N-terminus.

ε-DNP-lysine implies internal or C-terminal Lys.

Ala – (?) – Lys – (?)

b) Digestion with trypsin yielded one peptide.

Ala – (?) – Lys

c) Digestion with chymotrypsin yielded two peptides and free lysine.

Implies Phe-Lys and two Phe residues.

Ala – (?) – Phe | – (?) – Phe | – Lys

d) Reaction of the chymotrypsin cleavage products with FDNB followed by hydrolysis generated one mole equivalent of DNP-alanine and two mole equivalents of α,ε-DNP-lysines.

Fragment with N-terminal Ala.

Two fragments with N-terminal Lys.

Ala – (?) – Phe | – Lys – (?) – Phe | – Lys
Result (a) says the amino acids are in equimolar amounts. This means there are two of each residue type in the polypeptide. But from (b), there is only one peptide after trypsin treatment, which will cleave after Lys. This implies that the sequence is repeated yielding identical fragments after trypsin digestion.

Ala-Gln-Phe-Lys-Ala-Gln-Phe-Lys

Let’s check this: Trypsin digestion gives Ala-Gln-Phe-Lys and Ala-Gln-Phe-Lys.
(b) is ok.
Chymotrypsin digestion gives Ala-Gln-Phe + Lys-Ala-Gln-Phe + Lys
(c) is ok.
FDNB of the above fragments gives DNP-ala + two $\alpha,\varepsilon$-DNP-lysines.
(d) is ok.

Tricky!

3) While on an expedition to the Amazon jungle, you isolate a highly potent polypeptide toxin from a new spider species. Use the following information to reconstruct the primary sequence.

a) Acid hydrolysis released (Arg, Asx, 2 x Cys, Glx, Ile, Leu, Lys, Met, Phe, Pro, Ser)

Presence of Cys $\rightarrow$ possibility of disulfide bond

b) A single cycle of Edman degradation released Leu and Ser

Two N-terminal amino acids. Two chains with disulfide bond between them.

c) Cyanogen bromide treatment released the free amino acid asparagine.

Met-Asn. Asn is at the C-terminus of one of the chains.

d) Dithiothreitol (DTT) plus iodoacetic acid followed by trypsin digestion released 4 peptide fragments

i) (Arg, Ser) implies Ser-Arg
ii) (Asx, Met)
iii) (Cys, Glx, Ile, Leu, Phe, Pro) implies Leu-(Cys, Glx, Ile, Phe, Pro)
iv) (Cys, Lys) implies Cys-Lys

e) (Treatment of sulfhydryl groups with 2-bromoethylamine converts the Cys to Lysine-like amino acids, i.e., recognized by trypsin.) Treatment with dithithreitol (DTT) followed by 2-bromoethylamine and trypsin digestion released 6 peptides:

i) (Arg, Ser) implies Ser-Arg. Trypsin cleaves after Arg.
ii) (Asx, Met) implies Met-Asn is preceded by a Lys, Arg or Cys.
iii) Cys
iv) (Cys, Glx, Leu) implies (Glx,Leu)-Cys implies Leu-Glx-Cys. Trypsin cleaves after modified Cys. Leu is N-terminal from (b).
v) (Ile, Phe, Pro)
vi) Lys (iii) and (vi) implies Cys-Lys (See also (iv) in (d).)
Ser-Arg and (Cys-Lys) are in the same chain because the chain starting with Leu already has a Cys. (iii) and (vi) implies preceding (Cys, Lys) is Arg. This implies Ser-Arg-(Cys,Lys). Now, from (ii) we see that Met-Asn must be at the end of the second chain. What we have so far is

\[
\text{Leu-Glx-Cys-(Ile, Phe, Pro)}
\]

\[
\text{Ser-Arg-Cys-Lys-Met-Asn}
\]

f) Treatment with chymotrypsin released (Ile, Pro).

The order of (Ile, Phe, Pro) must then be Phe-Ile-Pro.

g) Treatment with staphylococcal V8 protease released (Glx, Leu).

This means Glx is Glu. The final structure is then

\[
\text{Leu-Glu-Cys-Phe-Ile-Pro}
\]

\[
\text{Ser-Arg-Cys-Lys-Met-Asn}
\]

4) Determine the sequence of an unknown peptide based on the following clues. Note: at each step, the relative amounts of each amino acid were not determined. Each step refers to the original peptide. (13 points)

a) Hydrolysis with 6N HCl shows the presence of the following amino acids: Ala, Arg, Gly, Lys, Met, Phe, Pro.

b) Treatment with FDNB followed by hydrolysis with 6N HCl shows the presence of DNP-Gly as well as ε-DNP-Lys.

c) Treatment with trypsin yields two peptides. The first peptide contains Ala, Gly and Lys. The second one contains Ala, Arg, Gly, Lys, Met, Phe, and Pro.

d) Treatment with chymotrypsin yields two peptides. The first peptide contains Ala, Gly, Lys, Met, and Phe. The second one contains Ala, Arg, Lys and Pro.

e) Treatment with carboxypeptidase B releases Lys.

f) CNBr cleavage yields two peptides. The first peptide contains Ala, Gly, Homoserine, and Lys. The second contains Ala, Arg, Gly, Lys, Phe, and Pro.

(b) implies N-terminal Gly

(c) (Ala,Gly)-Lys and (Ala,Arg,Gly,Lys, Met,Phe,Pro)

(d) (Ala,Gly,Lys,Met)-Phe and (Ala,Arg,Lys,Pro)

(e) C-terminal Lys

(f) (Ala,Gly,Lys)-Met and (Ala,Arg,Gly,Lys,Phe,Pro)

(b), (c), (d) and (f) suggest one fragment is Gly-Ala-Lys-Met-Gly-Phe and the other is (Ala,Arg,Pro)-Lys.
(c) suggests an Arg-Pro sequence, otherwise trypsin would cleave after Arg and would yield three peptides instead of two.

We end up with two possibilities:

Gly-Ala-Lys-Met-Gly-Phe-Ala-Arg-Pro-Lys or