BM2023 Assignment 1

Find the FASTA sequence of the RAP2 protein in cartilage. Find the corresponding amino acids. Collect information about the different amino acids. What secondary structures can be predicted from it?

RAP 2 (*Homo sapiens*) PROTEIN

(RAP2 protein in cartilage)

FASTA Sequence:

MREYKVVVLGSGGVGKSALTVQFVTGTFIEKYDPTIEDFYRKEIEVDSSPSVLEILDTAG
TEQFASMRDLYIKNGQGFILVYSLVNQQSFQDIKPMRDQIIRVKRYEKVPVILVGNKVDL
ESEREVSSSEGRALAEEWGCPFMETSAKSKTMVDELFAEIVRQMNYAAQPDKDDPC
CSACNIQ

Reference: https://www.ncbi.nlm.nih.gov/protein/CAA00806.1?report=fasta

Corresponding Amino Acids:

Methionine – Arginine – Glutamic Acid – Tyrosine – Lysine – Valine – Valine – Valine – Leucine – Glycine – Serine – Glycine – Glycine – Valine – Glycine – Lysine – Serine – Alanine – Leucine – Threonine – Valine – Glutamine – Phenylalanine – Valine – Threonine – Glycine – Threonine – Phenylalanine – Isoleucine – Glutamic Acid – Lysine – Tyrosine – Aspartic Acid – Proline – Threonine – Isoleucine – Glutamic Acid – Aspartic Acid – Phenylalanine – Tyrosine – Arginine – Lysine – Glutamic Acid – Isoleucine – Glutamic Acid – Valine – Aspartic Acid – Serine – Serine – Proline – Serine – Valine – Leucine – Glutamic Acid – Isoleucine – Leucine – Aspartic Acid – Threonine – Alanine – Glycine – Threonine – Glutamic Acid – Glutamine – Phenylalanine – Alanine – Serine – Methionine – Arginine – Aspartic Acid – Leucine – Tyrosine –

Isoleucine – Lysine – Asparagine – Glycine – Glutamine – Glycine – Phenylalanine – Isoleucine – Leucine – Valine – Tyrosine – Serine – Leucine – Valine – Asparagine – Glutamine – Glutamine – Serine – Phenylalanine – Glutamine – Aspartic Acid – Isoleucine – Lysine – Proline – Methionine – Arginine – Aspartic Acid – Glutamine – Isoleucine – Isoleucine – Arginine – Valine – Lysine – Arginine – Tyrosine – Glutamic Acid – Lysine – Valine – Proline – Valine – Isoleucine – Leucine – Valine – Glycine – Asparagine – Lysine – Valine – Aspartic Acid – Leucine – Glutamic Acid – Serine – Glutamic Acid – Arginine – Glutamic Acid – Valine – Serine – Serine – Serine – Glutamic Acid – Glycine – Arginine – Alanine – Leucine – Alanine – Glutamic Acid – Glutamic Acid – Tryptophan – Glycine – Cysteine – Proline – Phenylalanine – Methionine – Glutamic Acid – Threonine – Serine – Alanine – Lysine – Serine – Lysine – Threonine – Methionine – Valine – Aspartic Acid – Glutamic Acid – Leucine – Phenylalanine – Alanine – Glutamic Acid – Isoleucine – Valine – Arginine – Glutamine – Methionine – Asparagine – Tyrosine – Alanine – Alanine – Glutamine – Proline – Aspartic Acid – Lysine – Aspartic Acid – Aspartic Acid – Proline – Cysteine – Cysteine – Serine – Alanine – Cysteine – Asparagine – Isoleucine – Glutamine

Reference: https://www.cup.uni-muenchen.de/ch/compchem/tink/as.html

Functional properties of the corresponding amino acids:

1. Methionine: Its chemical formula is $C_5H_{11}NO_2S$. It is non polar and hydrophobic. Methionine is one of nine essential amino acids in humans (provided by food), Methionine is required for growth and tissue repair. A sulphur-containing amino acid, methionine improves the tone and pliability of skin, hair, and strengthens nails. Involved in many detoxifying processes, sulphur provided by methionine protects cells from pollutants, slows cell aging, and is essential for absorption and bio-availability of selenium and zinc. Methionine chelates heavy metals, such as lead and mercury, aiding their excretion. It also acts as a lipotropic agent and prevents excess fat buildup in the liver.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Methionine

2. Cysteine: Its chemical formula is $C_3H_7NO_2S$. Cysteine (Cys) is an enigmatic amino acid residue. Although one of the least abundant, it often occurs in functional sites of proteins. Free Cys is a polar amino acid, Cys in proteins is often buried and its classification on the hydrophobicity scale is ambiguous. Cysteine is a non-essential sulphur-containing amino acid in humans, related to cystine, Cysteine is important for protein synthesis, detoxification, and diverse metabolic functions. Found in beta-keratin, the main protein in nails, skin, and hair, Cysteine is important in collagen production, as well as skin elasticity and texture. It is also required in the manufacture of amino acid taurine; Cysteine is a component of the antioxidant glutathione, and plays a role in the metabolism of essential biochemicals such as coenzyme A, heparin, and biotin.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Cysteine

2. Cysteine: Its chemical formula is $C_3H_7NO_2S$. Cysteine (Cys) is an enigmatic amino acid residue. Although one of the least abundant, it often occurs in functional sites of proteins. Free Cys is a polar amino acid, Cys in proteins is often buried and its classification on the hydrophobicity scale is ambiguous. Cysteine is a non-essential sulphur-containing amino acid in humans, related to cystine, Cysteine is important for protein synthesis, detoxification, and diverse metabolic functions. Found in beta-keratin, the main protein in nails, skin, and hair, Cysteine is important in collagen production, as well as skin elasticity and texture. It is also required in the manufacture of amino acid taurine; Cysteine is a component of the antioxidant glutathione, and plays a role in the metabolism of essential biochemicals such as coenzyme A, heparin, and biotin.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Cysteine

3. Isoleucine: Its chemical formula is $C_6H_{13}NO_2$. It is non-polar, uncharged (at physiological pH), hydrophobic, aliphatic amino acid. L-Isoleucine is one of the essential amino acids that cannot be made by the body and is known for its ability to improve endurance and assist in the repair and

rebuilding of muscle. This amino acid is important to body-builders as it helps boost energy and helps the body to recover from training. L-Isoleucine is also classified as a branched-chain amino acid (BCAA). It helps promote muscle recovery after exercise. Isoleucine is broken down for energy within the muscle tissue. It is important in hemoglobin synthesis and regulation of blood sugar and energy levels.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/l-Isoleucine

4. Leucine: Its chemical formula is $C_6H_{13}NO_2$.It is non-polar and hydrophobic. Leucine is one of nine essential amino acids in humans (provided by food), Leucine is important for protein synthesis and many metabolic functions. Leucine contributes to regulation of blood-sugar levels; growth and repair of muscle and bone tissue; growth hormone production; and wound healing. Leucine also prevents breakdown of muscle proteins after trauma or severe stress and may be beneficial for individuals with phenylketonuria. Leucine is available in many foods and deficiency is rare.

   Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Leucine

5. Lysine: Its chemical formula is C6H14N2O2.It is polar and hydrophilic. Lysine is one of nine essential amino acids in humans required for growth and tissue repair, Lysine is supplied by many foods, especially red meats, fish, and dairy products. Lysine seems to be active against herpes simplex viruses and present in many forms of diet supplements. The mechanism underlying this effect is based on the viral need for amino acid arginine; lysine competes with arginine for absorption and entry into cells. Lysine inhibits HSV growth by knocking out arginine.

   Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Lysine

6. Phenylalanine: Its chemical formula is $C_9H_{11}NO_2$. It is non-polar and hydrophobic. Phenylalanine is an essential aromatic amino acid in humans (provided by food), Phenylalanine plays a key role in the biosynthesis of

other amino acids and is important in the structure and function of many proteins and enzymes. Phenylalanine is converted to tyrosine, used in the biosynthesis of dopamine and norepinephrine neurotransmitters. The L-form of Phenylalanine is incorporated into proteins, while the D-form acts as a painkiller. Absorption of ultraviolet radiation by Phenylalanine is used to quantify protein amounts.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Phenylalanine

7. Threonine: Its chemical formula is $C_4H_9NO_3$. It is polar but neutral and hydrophilic. Threonine is an essential amino acid in humans (provided by food), Threonine is an important residue of many proteins, such as tooth enamel, collagen, and elastin. An important amino acid for the nervous system, threonine also plays an important role in porphyrin and fat metabolism and prevents fat build-up in the liver. Useful with intestinal disorders and indigestion, threonine has also been used to alleviate anxiety and mild depression.

   Reference: https://pubchem.ncbi.nlm.nih.gov/compound/L-threonine

8. Tryptophan: Its chemical formula is $C_{11}H_{12}N_2O_2$. It is non-polar and hydrophobic. Tryptophan is the least plentiful of all 22 amino acids and an essential amino acid in humans (provided by food), Tryptophan is found in most proteins and a precursor of serotonin. Tryptophan is converted to 5-hydroxy-tryptophan (5-HTP), converted in turn to serotonin, a neurotransmitter essential in regulating appetite, sleep, mood, and pain. Tryptophan is a natural sedative and present in dairy products, meats, brown rice, fish, and soybeans.

   Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Tryptophan

9. Valine: Its chemical formula is $C_5H_{11}NO_2$. Valine is an aliphatic, non-polar and extremely hydrophobic essential amino acid in humans related to leucine, Valine is found in many proteins, mostly in the interior of globular proteins helping to determine three-dimensional structure. A glycogenic

amino acid, valine maintains mental vigor, muscle coordination, and emotional calm. Valine is obtained from soy, cheese, fish, meats and vegetables. Valine supplements are used for muscle growth, tissue repair, and energy.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Valine

10. Alanine: Its chemical formula is $C_3H_7NO_2$. It is non-polar and hydrophobic. Alanine is a small non-essential amino acid in humans, Alanine is one of the most widely used for protein construction and is involved in the metabolism of tryptophan and vitamin pyridoxine. Alanine is an important source of energy for muscles and the central nervous system, strengthens the immune system, helps in the metabolism of sugars and organic acids, and displays a cholesterol-reducing effect in animals.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Alanine

11. Arginine: Its chemical formula is $C_6H_{14}N_4O_2$. It is polar and hydrophilic. Arginine is an essential amino acid in juvenile humans, Arginine is a complex amino acid, often found at active site in proteins and enzymes due to its amine-containing side chain. Arginine may prevent or treat heart and circulatory diseases, combat fatigue, and stimulate the immune system. It also boosts production of nitric oxide, relaxing blood vessels, and treating angina and other cardiovascular problems. Arginine is also an important intermediate in the urea cycle and in detoxification of nitrogenous wastes.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Arginine

12. Aspartic Acid: Its chemical formula is $C_4H_7NO_4$. It is polar and hydrophilic. Aspartic Acid is a non-essential amino acid in humans, Aspartic Acid has an overall negative charge and plays an important role in the synthesis of other amino acids and in the citric acid and urea cycles. Asparagine, arginine, lysine, methionine, isoleucine, and some nucleotides are synthesized from aspartic acid. Aspartic acid also serves as a neurotransmitter.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Aspartic%20Acid

13. Asparagine: Its chemical formula $C_4H_8N_2O_3$. It is polar and hydrophilic. Asparagine is a non-essential amino acid in humans, Asparagine is a beta-amido derivative of aspartic acid and plays an important role in the biosynthesis of glycoproteins and other proteins. A metabolic precursor to aspartate, Asparagine is a nontoxic carrier of residual ammonia to be eliminated from the body. Asparagine acts as diuretic.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Asparagine

14. Glutamic Acid: Its chemical formula $C_5H_9NO_4$. It is polar and hydrophilic. Glutamic acid (Glu), also referred to as glutamate (the anion), is one of the 20 proteinogenic amino acids. It is not among the essential amino acids. Glutamate is a key molecule in cellular metabolism. In humans, dietary proteins are broken down by digestion into amino acids, which serves as metabolic fuel or other functional roles in the body. Glutamate is the most abundant fast excitatory neurotransmitter in the mammalian nervous system.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Glutamic%20Acid

15. Glutamine: Its chemical formula $C_5H_{1o}N_2O_3$. It is polar and hydrophilic. Glutamine is a nonessential amino acid. Glutamine can donate the ammonia on its side chain to the formation of urea (for eventual excretion by the kidneys) and to purines (necessary for the synthesis of nucleic acids). Glutamic acid-to-glutamine conversion, in which an ammonia group is added to glutamic acid (catalyzed by glutamine synthase), is of central importance in the regulation of toxic levels of ammonia in the body. This agent is a substrate for the production of both excitatory and inhibitory neurotransmitters (glutamate and GABA) and is also an important source of energy for the nervous system. Glutamine may become a conditionally essential amino acid during certain catabolic states.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Glutamine

16. Glycine: Its chemical formula $C_2H_5NO_2$.Glycine is a non-essential,

non-polar, hydrophobic, non-optical, glucogenic amino acid. Glycine, an inhibitory neurotransmitter in the CNS, triggers chloride ion influx via ionotropic receptors, thereby creating an inhibitory post-synaptic potential. In contrast, this agent also acts as a co-agonist, along with glutamate, facilitating an excitatory potential at the glutaminergic N-methyl-D-aspartic acid (NMDA) receptors. Glycine is an important component and precursor for many macromolecules in the cells.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Glycine

17. Proline: Its chemical formula $C_5H_9NO_2$.It is nonpolar, hydrophobic, soluble in water and a moderately acidic compound. Proline is a cyclic, nonessential amino acid (actually, an imino acid) in humans (synthesized from glutamic acid and other amino acids), Proline is a constituent of many proteins. Found in high concentrations in collagen, proline constitutes almost a third of the residues. Collagen is the main supportive protein of skin, tendons, bones, and connective tissue and promotes their health and healing.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Proline

18. Serine: Its chemical formula $C_3H_7NO_3$. It is polar and hydrophilic. Serine is a non-essential amino acid in humans (synthesized by the body), Serine is present and functionally important in many proteins. With an alcohol group, serine is needed for the metabolism of fats, fatty acids, and cell membranes; muscle growth; and a healthy immune system. It also plays a major role in pyrimidine, purine, creatine, and porphyrin biosynthetic pathways. Serine is also found at the active site of the serine protease enzyme class that includes trypsin and chymotrypsin.

    Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Serine

19. Tyrosine: Its chemical formula $C_9H_{11}NO_3$. It is polar and hydrophobic. L-Tyrosine is the laevorotatory isomer of the aromatic amino acid tyrosine. L-tyrosine is a naturally occurring tyrosine and is synthesized in vivo from L-phenylalanine. It is considered a non-essential amino acid; however, in patients with phenylketonuria who lack phenylalanine hydroxylase and

cannot convert phenylalanine into tyrosine, it is considered an essential nutrient. In vivo, tyrosine plays a role in protein synthesis and serves as a precursor for the synthesis of catecholamines, thyroxine, and melanin.

Reference: https://pubchem.ncbi.nlm.nih.gov/compound/Tyrosine

Secondary structure predicted for RAP2 protein:

Predicted Complete Secondary Structure:

```
A '*' character above the overall prediction indicates the homolog's structure was used at this residue.
  1  ************************************************************ 60
     MREYKVVVLGSGGVGKSALTVQFVTGTFIEKYDPTIEDFYRKEIEVDSSPSVLEILDTAG
     CEEEEEEECCCCCCHHHHHHHHHCCCCCCCCCCEEEEEEEEECCCEEEEEEEECC
     9699999986788646788888755788888777757899999996897799999999976

 61  ************************************************************ 120
     TEQFASMRDLYIKNGQGFILVYSLVNQQSFQDIKPMRDQIIRVKRYEKVPVILVGNKVDL
     CCCCHHHHHHHHHCEEEEEEECCCHHHHHHHHHHHHHHHHCCCCEEEEEEECCCC
     665578999988598699999985886579999999999985788888699999858998

121  ************************************************************ 180
     ESEREVSSSEGRALAEEWGCPFMETSAKSKTMVDELFAEIVRQMNYAAQPDKDDPCCSAC
     CCCCCCCHHHHHHHHHHCEEEEEECCHHHHHHHHHHHHHHHHCCCCCCCCCCCCCC
     887776479999999985878999995888758999999999999857877777676777

181  *   183
     NIQ
     CCC
     778
```

Legend:

```
H = Helix
E = Beta Strand
C = Coil
T = Membrane helix
B = Membrane strand
S = Signal peptide
c = Cleavage site
Line 1 = sequence (single letter IUPAC code, 60 characters per line)
Line 2 = secondary structure (H, E or C)
Line 3 = confidence score (0-9, 0 = low, 9 = high)
```

Reference:

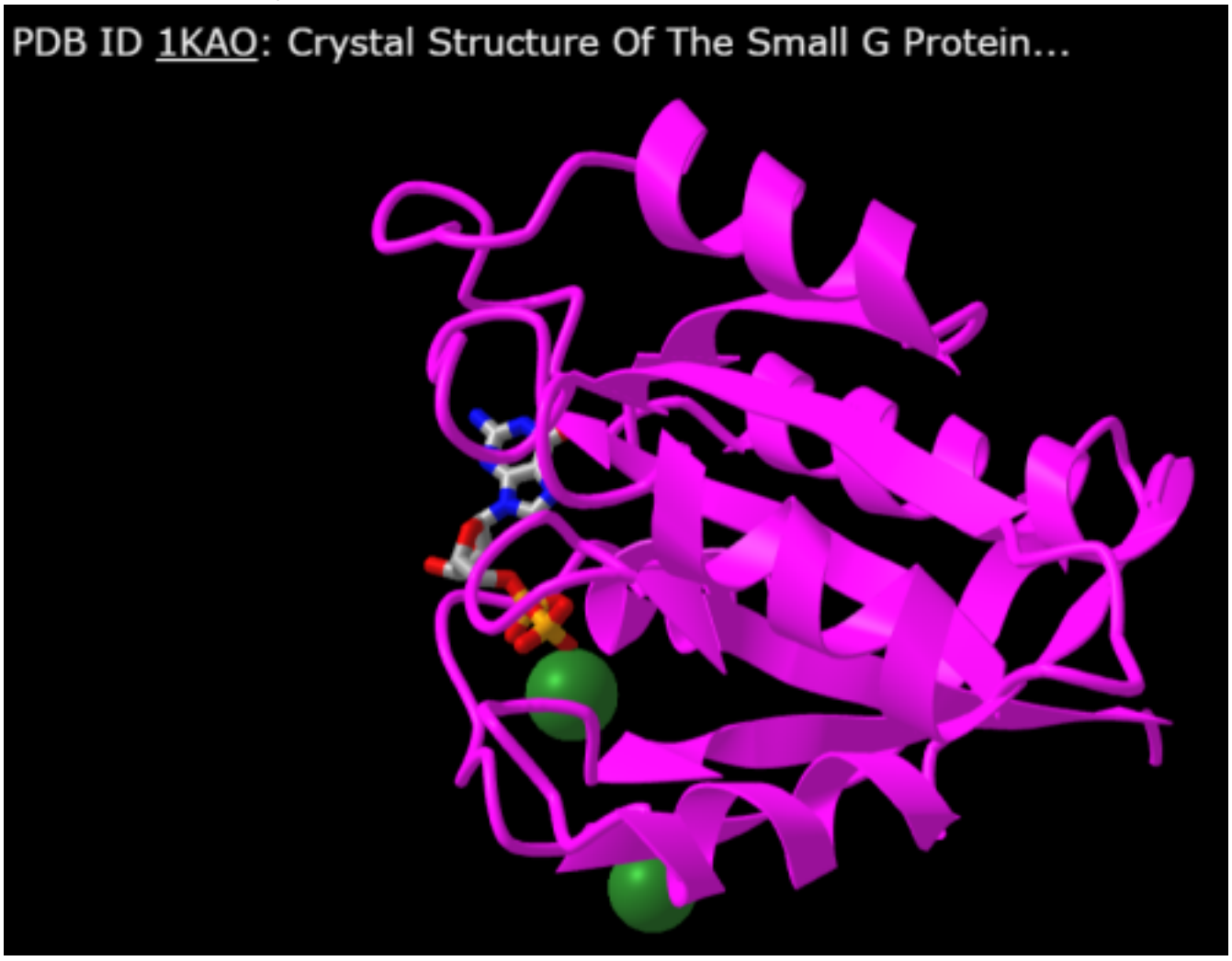
Predicted using Proteus Structure Prediction Server 2.0:

http://www.proteus2.ca/proteus2/Result.do?ID=3236852;jsessionid=646027EE3419317EE29BB1C3682B0C94

3D visualization of RAP2 protein (RAP2A):

(PDB ID: 1KAO; 1KAO_A):

NOTE: RAP2 family GTPase consists of RAP2A, RAP2B, and RAP2C.



Reference:

Obtained using iCn3D (NCBI):

https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?&mmdbid=56535&bu=1&showanno=1&source=full-feature