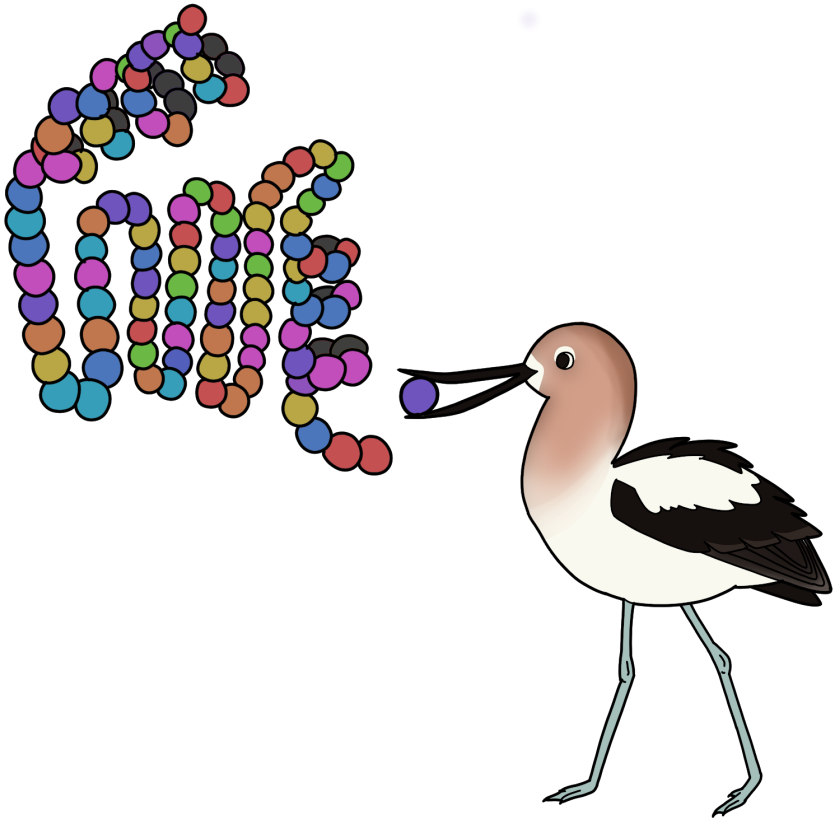


Protein Modeling TEST - Division C

BirdSO 2021

Written by Megan Luo and Eric Ma

Name:_____

Team Number: _____

School: _____

Score: /258

PART 2: Jmol Exploration

Open the PDB 1AML.
This is an NMR conformational ensemble. For now, consider only the first model, which we can see by typing "model 1." We define C = carbon, CA = alpha carbon, N = nitrogen.

1. List the first five amino acids in the N-terminus of the protein. (4 points)

2. The N-terminus of the protein resembles a special conserved protein motif that is often found in beta meanders, or other beta sheets. Name the motif. (4 points)

3. Give the amino acid identity and position numbers of the amino acids that indicate the existence of the motif in the previous question. (4 points)

4. Give the dihedral angle defined by atoms CA-12, C-12, N-13, and CA-13. What is the name of this angle, as a greek letter? (6 points)

5. Give the dihedral angle defined by atoms N-12, CA-12, C-12, and N-13. What is the name of this angle, as a greek letter? Same deal as with the previous question. (6 points)

6. Give the dihedral angle defined by atoms C-12, and N-13, CA-13, and C-13. What is the name of this angle, as a greek letter? Same deal as with the previous question. (6 points)

7. What is the distance between the alpha carbon of the first and last residues in the protein? (4 points)

8. What's the longest distance between any two atoms in the protein? Name or describe the two atoms and give their position in the sequence. (4 points)

9. The alpha helices in the PDB are joined by a linker. Describe the flexibility of this linker in terms of the different conformations in the PDB file and the phi-psi angles of the residues. (6 points)

Now, open model 14 in the same PDB file by typing "model 14."

10. Give the dihedral angle defined by atoms CA-12, C-12, N-13, and CA-13. (5 points)

11. Give the dihedral angle defined by atoms N-12, CA-12, C-12, and N-13. (5 points)

12. Give the dihedral angle defined by atoms C-12, and N-13, CA-13, and C-13. (5 points)

13. Color the protein based on the B-factor. Explain what you see. (4 points)

We often use Jmol to get an intuitive sense of protein structure and analyze interactions, but we take a lot of the behind-the-scenes work for granted. Thus, for the rest of Part II, you will answer questions based on how protein structures are found experimentally and displayed using software like Jmol.

14. The structure found in 1AML was solved using NMR. Describe when NMR can be used to solve structures, and its advantages and disadvantages compared to X-ray diffraction. (8 points)

15. Based on these ideas, are Cas proteins and base editors (such as Cas9 or the adenine base editor) structures usually solved with X-ray diffraction, NMR, or something else? What information do they lose out on by solving the structures with this method? (4 points)

16. Having structures of homologous proteins to your protein of interest can greatly improve the quality of the structure you are interested in. What computational technique does this describe? (2 points)

17. There are typically four steps in protein structure prediction. List them. (12 points)

18. One process for determining protein structure involves starting from a linear amino acid chain and simulating the forces in the protein until a folded state is reached. What is the name of this process? (4 points)

19. What formula/metric (that outputs a single number) is typically used to quantify the structural similarity between proteins? Explain the formula. (6 points)

20. Suppose we just solved the structures of two homologous proteins, and we want to use the formula discussed in the previous question to determine structural similarity. Explain why naively applying that formula would not be useful. (8 points)

21. What can we do before using the formula discussed in the previous two questions to make the calculation sensible? What algorithm can be used to achieve this? You don't have to explain the algorithm. (8 points)

22. According to the above formula, protein A and B have a similarity of X. Suppose all the bond lengths are tripled. What's the new similarity in terms of X? (4 points)

23. A representation of an object is referred to as invariant to a property if a change in that property does not change the representation. For example, if a rotation of a protein in some representation does not change a given property Y, that representation is invariant to rotation. A distance matrix is a matrix where the $i,j^\\text{th}$ entry is the distance between the $i^\\text{th}$ and $j^\\text{th}$ residues. Are distance matrices invariant to rotation and/or translation? What kind of symmetry to distance matrices have, if any? Explain. (6 points)

24. In the context of protein structure prediction, explain why it is useful to have rotation and translation invariant representations of proteins. (4 points)

PART 3: Written Exam

For questions 25-33, determine if the statement is true or false.

25. Peptide bonds are usually rigid and planar because of their partial double bond character. (1 point)

26. In the secondary structure, beta-strands are usually straight which is why beta-hairpins are very common. (1 point)

27. Whilst folding, a protein will form by folding into the final state through specific and non-specific interactions without any temporary intermediate forms. (1 point)

28. The major difference between beta barrels and alpha helices is found in secondary structure because alpha helices are rolled up creating a shorter and more hydrophobic bundle than bundled beta-sheets. (1 point)

29. The Ramachandran plot is a plot based on the phi and psi angles of a residue in a peptide. (1 point)

30. Most naturally occurring amino acids have the L-configuration on the main carbon. (1 point)

31. Cryo-electron microscopy is a common tool used for imaging of protein complexes at the molecular level. (1 point)

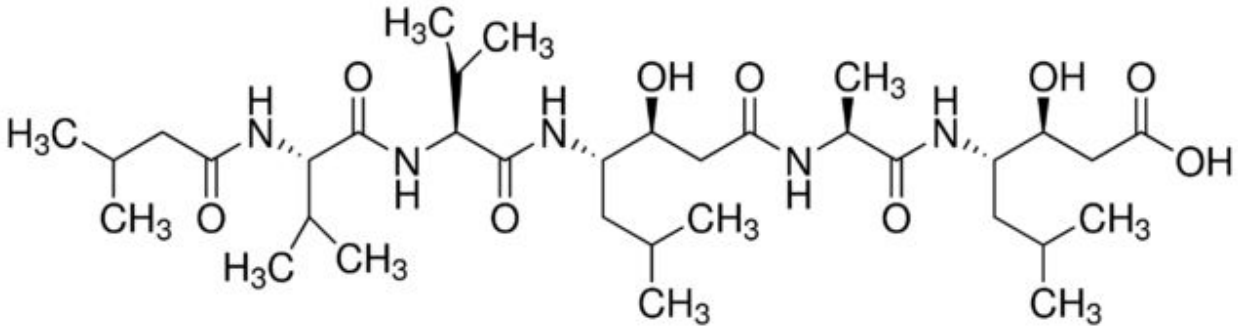
32. Disulfide bonds occur between all sulfur-containing amino acids, like methionine and cysteine. (1 point)

33. Coiled coils usually contain both hydrophobic and charged residues. (1 point)

34. P53 is a tumor suppressor protein, it prevents cancer formation. Listed below are some of the structures found on the protein and their general roles. Mutation or damage of which region would most likely lead to the deactivation p53. Explain. (4 points)

a. N-terminus: activates transcription factors;
b. Central DNA-binding core domain: binding the p53 co-repressor;
c. C-terminal: downregulation of DNA binding;
d. Activation domain 2: apoptotic activity

35. Pepstatin A is an oligopeptide and an excellent inhibitor of aspartyl proteases because it contains the amino acid statine. However, it is a very selective inhibitor, e.g. it only inhibits aspartic proteases not others such as thiol proteases. Please explain why this might be the case. An image of the structure is pictured below for your reference. (3 points)



36. Below is a polypeptide strand. Which amino acids would typically be found on the surface of the protein when folded? Explain. (4 points)

RNDCQKMFPSTWYVA

37. Which pair of amino acids can convert from one to another through biosynthesis? (3 points)

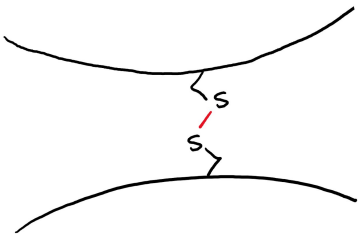
    a. Glutamate to Proline
    b. Glycine to Proline
    c. Glutamate to Glutamine
    d. Alanine to Serine
    e. None of the above.

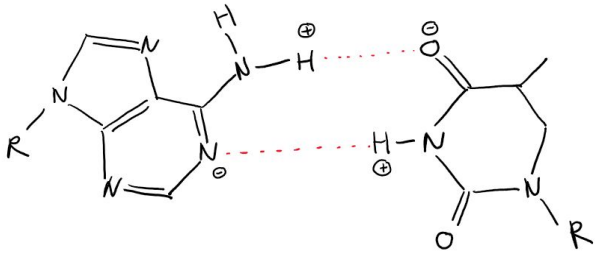
38. Which of the following reasons cause proline to affect a protein's secondary structure? (3 points)

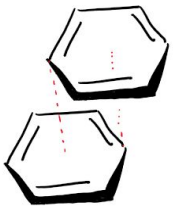
    a. The R-Group connects to both the alpha and beta Carbon.
    b. It often exists as a zwitterion, therefore the minimal charges cause structural issues when folding into a beta sheet.
    c. Proline is aromatic so its unsaturated ring has a positive charge.
    d. The *cis/trans* forms of proline are almost isoenergetic allowing for easy structural errors.
    e. None of the above.

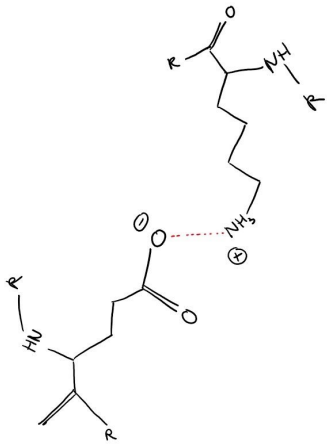
Match an interaction to each picture. Each choice can be used once, more than once, or not at all. Use this for questions 39-42. (2 points each)

A. Salt bridge
B. London Dispersion
C. Disulfide bonds
D. Hydrophobic/hydrophilic
E. Pi stacking
F. Hydrogen bonding

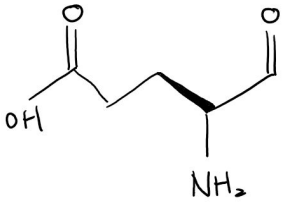
39. 

40. 

41. 

42. 

43. Which amino acid is this and is the stereochemistry S or R? Explain. (2 points)



44. Which of the following point mutations on the mRNA would change serine to glycine? (1 point)

    a. G → A
    b. U → A
    c. A → G
    d. G → C
    e. None of the above

45. Which kinds of substitution mutation occurred in the previous question 44? (1 point)

    a. Nonsense
    b. Silent
    c. Missense
    d. Insertion
    e. None of the above

46. Which of the following will commonly be found in the middle of beta-sheets? Choose all that apply. (1 point?)

    a. Tyrosine
    b. Valine
    c. Threonine
    d. Tryptophan
    e. None of the above

47. What is the HNH domain responsible for? (1 point)
    a. Binding histidine and asparagine residues to the DNA sequence to prepare it for cleaving.
    b. Cleaves the complementary DNA strand in the CRISPR-Cas9 system.
    c. Help the CRISPR-Cas9 system recognize the PAM sequence.
    d. Matures the crRNA into sgRNA.
    e. None of the above.

48. What is cytidine deaminase responsible for? (1 point)

49. Which of the following begins the cleavage of the non-complementary DNA strand when using a CRISPR-Cas9 system? (2 points)

    a. sgRNA
    b. Cas9
    c. PAM
    d. RuvC
    e. HNH
    f. None of the above

50. Which of the following statements about ionization are true? Choose all that may apply. (4 points)

    a. All amino acids have one ionizable amine group but no ionizable carboxyl group.
    b. All amino acids have one ionizable amine group and one ionizable carboxyl group.
    c. Not all amino acids have ionizable R-groups.
    d. Amino acids are able to ionize when bound by a peptide bond.
    e. None of the above

51. Which of the statements are true about the synthesis process of cysteine from methionine? Choose all that apply. (4 points)

    a. Both methionine and cysteine are neutrally charged and equally nonpolar.
    b. Converting from methionine to homocysteine (an analog of cysteine) requires the removal of one methyl group to form a thiol group.
    c. Homocysteine and serine combine to produce an intermediate to cysteine.
    d. Methionine and cysteine both are sulfur-containing amino acids.
    e. None of the above are true.

52. "Threonine and Valine are roughly the same shape and volume because there is only one functional group that differentiates them. As a result, they are chemically similar and can be used as a substitute for each other." Is this statement true or false? Explain using your understanding of amino acids structures and chemical characteristics. (3 points)

Acetylation is a very common modification in eukaryotes, especially histone acetylation and deacetylation. While the evolutionary purposes of acetylation are not pinpointed like phosphorylation, many new studies show that acetylation helps with gene regulation/expression, protein activity regulation, and more. Questions 53-57 will test your understanding of acetylation as a post-translational modification and its effects on a protein's different structures.

53. Chemically explain how acetylation modifies a substance, such as histones, and its electrostatic effects. (2 points)

54. Which residue would non-terminal acetylation normally occur on? (2 points)

    a. Proline
    b. Orthinione
    c. Arginine
    d. Lysine

55. Nt-acetylation is also a common modification; however, unlike other protein acetylations, it is irreversible. How could this influence protein properties? Explain. (4 points)

56. Nt-acetylation requires a specific enzyme complex and substrate (amino acid pairing). Complex NatB is known to be bind on the terminus where the first amino acid is MET and the following amino acid is highly hydrophilic? List 2 amino acids that could be second in the sequence. (2 point)

57. Despite its benefits, overexpression of Nt-acetylation and NATs (the enzymes that assist in the process) have been seen in diseases such as cancers. One specific enzyme hNaa10p has shown high levels of expression in tumors/prolific cell growth. How does the process of expression Nt-acetylation and NATs correlate with tumor growth? (5 points)

58. In a new protein, you find that there are lots of tetrahedral chemical bonding on the alpha-C atom and the following peptide backbone angles: $\varphi = -139°$ $\psi = 133°$. What structural motif might be present? (2 points)

59. Proline is not usually found inside alpha-helices; however, they are often the first residue of an helix. Using your understanding of amino acids, please explain this phenomenon by specifically referencing components of proline's structure. (3 points)

60. Name another helix structure. Out of the helix structures, why is the alpha helix most common? (2 points)

61. Which of the following is false about endonuclease? Choose all that apply. (2 points)

    a. The endonuclease cleaves phosphodiester bonds.
    b. Type I, II, and III restriction endonuclease use ATP when cleaving a sequence.
    c. Endonucleases can cleave RNA, dsDNA, ssDNA.
    d. Type I endonuclease has a recognition site within the cleavage pattern.

62. What is required in Type II CRISPR-Cas systems to mature crRNA? (2 points)

    a. tracrRNA
    b. sgRNA
    c. mRNA
    d. Helicase
    e. None of the above.

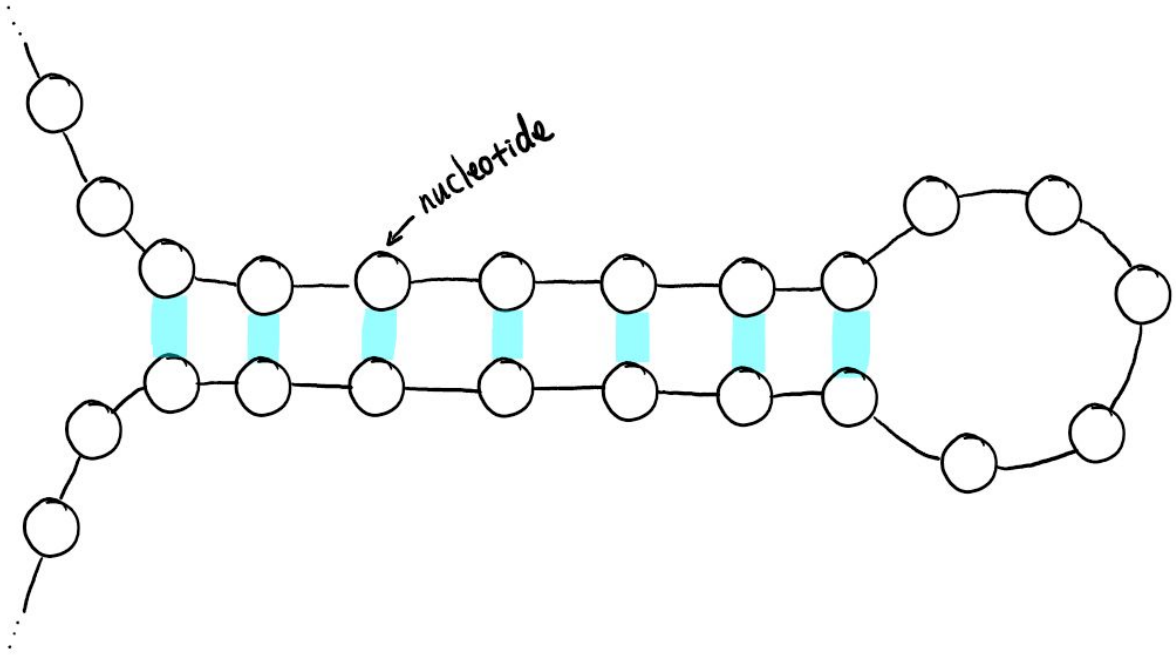
63. Which of the following is true about type I and type II CRISPR-Cas systems but not Type III? (2 points)

    a. The system will use crRNAs to identify and cleave complementary target systems.
    b. The system cleaves at a single point of the DNA.
    c. The system is made up of subunits that work cohesively with a Cas protein.
    d. The system recognizes and cleaves RNA.
    e. None of the above.

64. Which of the following is a similar structure to the R-loop formed in Type I systems? (2 points)

    a. Beta hairpin
    b. D-loop
    c. C-loop
    d. A-loop
    e. Beta barrel
    f. None of the above.

CRISPR-Cas9 typically only has a Cas9 protein and a sgRNA. The sgRNA has a core hairpin structure (the first stem loop of the sgRNA). It is a commonly secondary structure of RNA where complementary nucleotide sequences form Watson-Crick base pairs that leaves an unpaired loop at the end. See the image below for reference. Use this information for the questions 65-66.

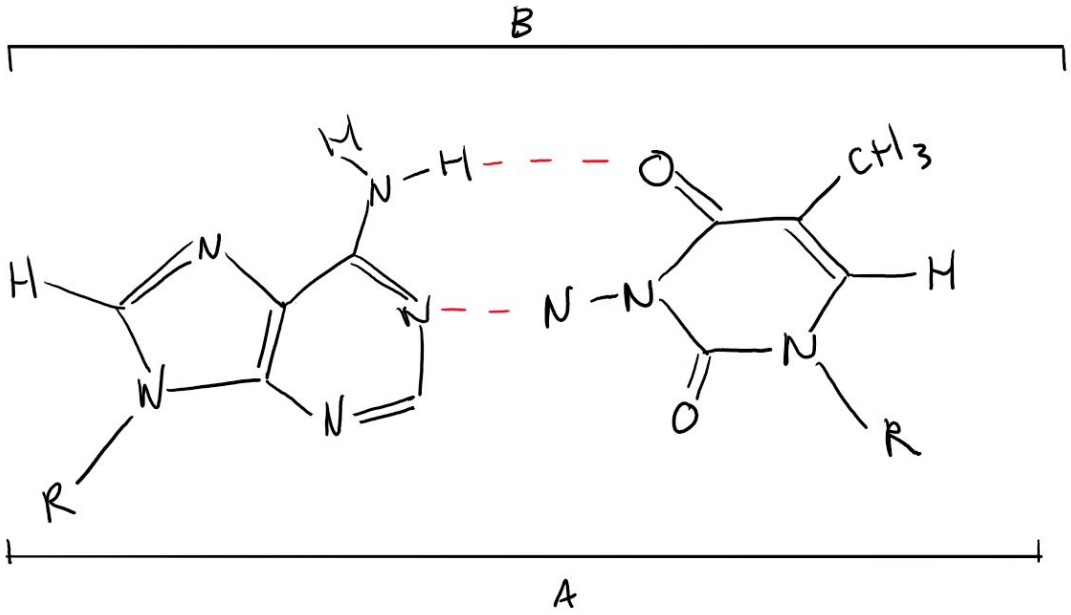


65. Which of the following conditions would be true for forming stable hairpin loops? Choose all that apply. (4 points)

    a. The loop cannot be longer than 5 base pairs or paired base pairs will pop apart.
    b. The RNA sequence can fold back onto itself to form a paired double helix.
    c. Loops made up of base stacking nucleotides.
    d. Base stacking interactions that align pi bonds of bases.

66. Why might this hairpin be necessary for the CRISPR-Cas9 system? (2 points)

Use the following diagram of two bounded base pairs to answer questions 67-71.



67. Which base pairs are shown? (1 point)

68. Which kind of bonds are represented by the dashed red lines? (1 point)

69. Which side is the minor groove side? Explain. (2 points)

70. Which secondary structure motif is commonly used for major groove specific DNA recognition? (2 points)

    a. Coiled coil
    b. Alpha helix
    c. Hairpin loop
    d. Beta-meander
    e. None of the above.

71. Major and minor grooves recognition is used for PAM recognition when using the CRISPR-Cas9 system. Explain how groove recognition is applied when reading for the PAM motif. (3 points)

Use the following scenario to answer questions 72-75.
This DNA was found in *Streptococcus pyogenes* by some friends last weekend. They want to edit the strand using CRISPR-Cas9 by adding in a small fragment of DNA.

72. Which of the following could be a possible PAM site? (2 points)

    a. CCA
    b. GCG
    c. AGG
    d. AAT
    e. None of the above.

73. The friends want to design a guide RNA to help them add the small DNA fragment into the DNA sequence. Which of the things should they consider when choosing a sgRNA from a database? Choose all that apply. (4 points)

    a. Maximize off-target activity score to ensure that the sgRNA will bind to the correct target.
    b. Finding out whether out the experiment is to activate or inhibit a gene.
    c. The impact of in-frame mutations on functionality of a protein.
    d. Balancing the efficiency and specificity of a sgRNA**.**

74. In a followup experiment, the friends find out they can use CRISPR-Cas9 to knock out genes to test genetic function. This usually requires making a double strand break and then an insertion or deletion to cause a frameshift. Which kind of mechanism could they use alongside CRISPR-Cas9 to perform gene knockouts? Explain. (4 points)

75. In their gene knockout experiment, the friends are editing the DNA strand, a portion of the sequence is shown below. Which of the following edits can they make to inhibit the expression of the gene coding from base 12 to 20? Base 3 is this first base in the first codon. (Bases are numbered starting at 1 from the 5' end) Select all that apply. (4 points)

5' TTATAGCTTGTCTGTATAAT 3'

a. Edit base 11 from T → A.
b. Deletion of base 15.
c. Insertion of "AA" after base 3.
d. Deletion of base 1.

Type II and I CRISPR-Cas systems target double stranded DNA, Type III CRISPR-Cas systems will bind and cleave to single-stranded RNA sequences. Despite this difference, the effector complex of Type III systems (extracted from *Thermus thermophilus* aka Cmr) are structurally similar to the Type I CRISPR-Cascade complex with Cas3. The Type III CRISPR-Cas system has a 12-subunit assembly composed of six Cmr subunits (Cmr1–6) and a crRNA with a stoichiometry of $Cmr1_1 2_1 3_1 4_4 5_3 6_1$:$crRNA_1$.

Below is an excerpt from a research paper on type III CRISPR-Cas systems. Please use it to answer questions 76-79.

"The Cmr effector complex cleaves target ssRNAs at five sites in vitro, despite containing only four Cmr4 subunits. Reanalyzing our structures, we noticed a thumblike extension, nearly identical to those observed in individual Cmr4 subunits, originating in Cmr6. In the context of the target-bound Cmr structure, this thumb places the target strand in a position for cleavage of the 5′-most site on the target RNA, Similarly, a thumblike domain in Cmr3 stretches into the palm of the bottom Cmr4 subunit, which stacks on top of the 5′-handle and scaffolds the 3′-most discontinuous segment of crRNA:target.
These...type III complexes...use thumb-mediated local disruption of duplex geometry in their interactions with substrate sequences, leading to a lack of continuous double-helix formation between guide RNA and target strands. That RecA employs similar discontiguous DNA-DNA interactions for homology searches (18) hints at a common mode of substrate recognition among genome surveillance complexes...In the related type III CRISPR-Csm complex, discontinuous helix formation might occur during association with topologically constrained R loops formed during transcription."

76. The Cmr effector is able to cleave at 5 sites with four Cmr4 subunits and the thumblike extension from Cmr6 subunit. A similar thumblike domain is seen in Cmr3. Which of the following structures might be structurally similar to the thumblike structure? (3 points)

    a. Alpha helix
    b. Beta sheets
    c. Beta hairpins
    d. Beta barrel
    e. None of the above.

77. The following are descriptions of the Cmr complex from Type III systems. Which of the following sentences would suggest architectural similarities with Type I CRISPR-Cascade Systems? (4 points)

    a. There are long beta-strand extensions from the Cmr complex (specifically extends from the middle of a Cmr4 subunit) and interacts with an adjacent subunit.
    b. The effector complex is made up of six Cmr subunits and a crRNA.
    c. The crRNA 5'handle is attached to the Cmr2-Cmr3 heterodimer.
    d. The alpha-helical bundle in Cmr2 attaches to the bottom of Cmr5.

78. RecA homology searches suggest a common substrate recognition method. Name and briefly explain two substrate recognition methods that are used by CRISPR-Cas systems. (4 points)

79. The Type I-E CRISPR Cas3 system from *E. coli* uses sgRNAs to silence foreign DNA. Explain how this system differs from the CRISPR Cas9 in both structure and function? (3 points)