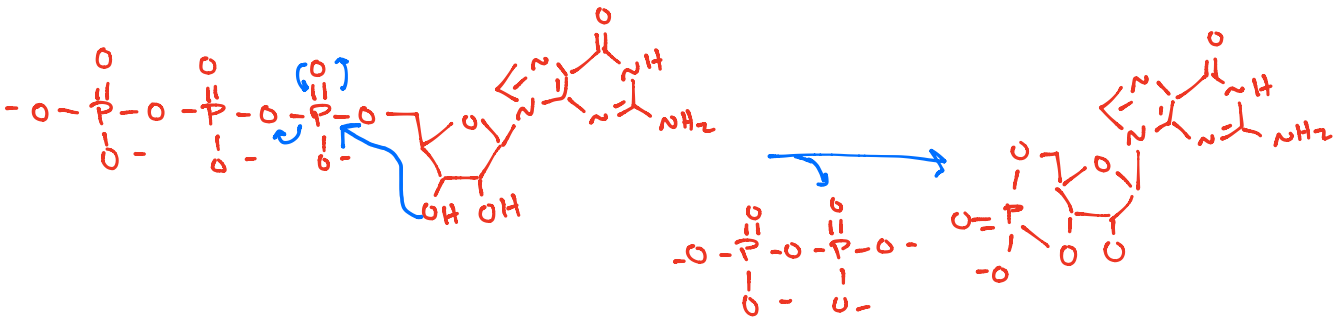
Problem Set 2 **(Due Sept 9th)**

1. Consider these two polynucleotides: AAGCGT GCACTG
   a. Draw each molecule in the 2'deoxy form (DNA). SEE BELOW
   b. What is the sequence of the complementary strand? Write this in the 5'→3' direction. 5'-ACGCTT-3' 5'-CAGTGC-3'
   c. Draw the first base pair in each case. You can use 'X' in place of the deoxyribose. SEE BELOW
   d. Which oligonucleotide will anneal more favorably (higher T_m) with its complement? Why? GCACTG → GC base pairs are more stable in dsDNA than AT base pairs.
   e. If Adenine were added to the 2nd strand by DNA Polymerase, where would it go? Why? 3' end because polymerization reactions need a 3' hydroxyl to be the nucleophile.

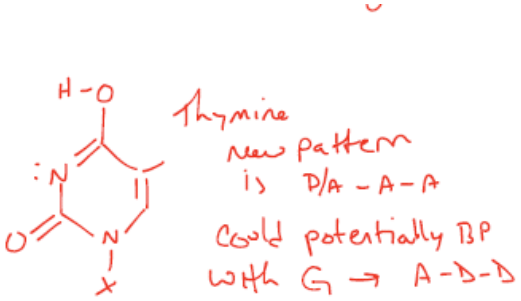
2. What is meant by cooperative folding/unfolding with respect to DNA structure? Folding and unfolding are not all or none phenomenon; instead, it is a stepwise process that becomes more favorable as previous events occur. If we consider unfolding, once the terminal base pairs separate, it's easier for the next base pairs to separate.

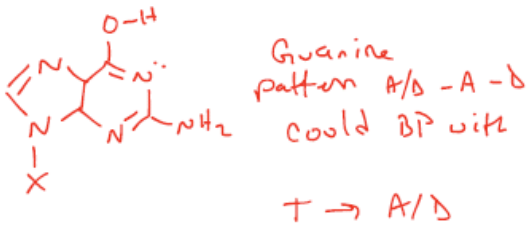
3. Describe all intermolecular forces that contribute to DNA structure. H-bonds are not very important because they reform with water upon unfolding. Ion-ion interactions are important to stabilize the anionic phosphordiester backbone. A combination of H-bond, dipole-dipole, and London Dispersion forces play critical roles in the base stacking interactions.

4. cGMP(3'-5' cyclic guanosine monophosphate) is generated from GTP. Propose a mechanism by which this could happen. What is the product? dGMP and pyrophosphate Do you expect this to be energetically favorable? Why? Yes because a high energy phosphate bond is hydrolyzed. Additionally, the PPi can be hydrolyzed to inorganic phosphate (Pi) to provide more energy.



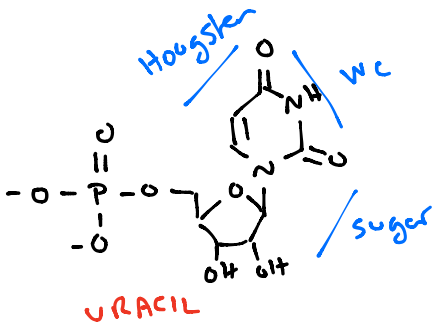
5. We have seen that base pairing occurs between keto tautomers of the nucletides. However, base pairing could potentially occur between enol-enol or enol-keto forms of the nucleotides as well. Determine if enol tautomers of the nucleotides have base pairing potential. Assume that at least two hydrogen bonds must be present to create a stable base pair.

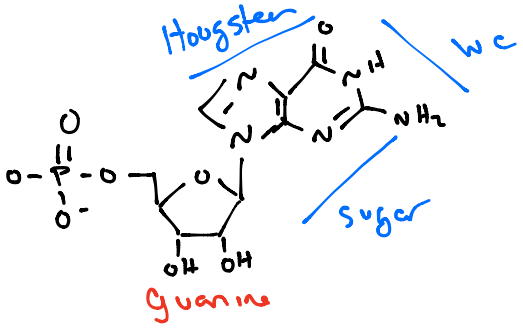


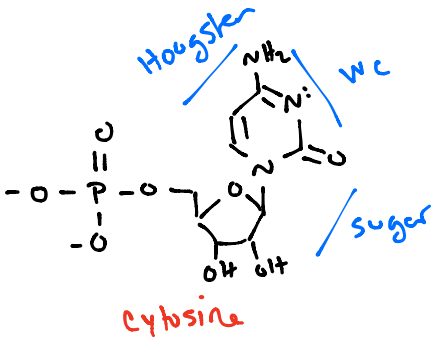


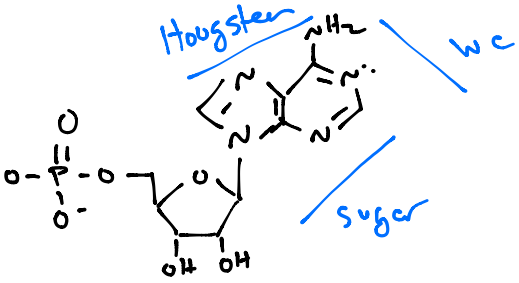
6. Describe the distinct characteristics of A-DNA, B-DNA and Z-DNA. A-DNA is a right handed helix with a wider and more compressed helical structure. It has a shallow major groove but a deep minor groove. B-DNA is the standard structure with a right handed helix, a deep and wide major groove and a narrow and wide minor groove. Z-DNA is a long skinny left handed helix that almost never forms.

7. Draw each of the nucleotide monophosphates. Identify the Watson-Crick, Hoogsteen, and sugar faces. Explain why the H-bonding pattern of each of these faces is important.







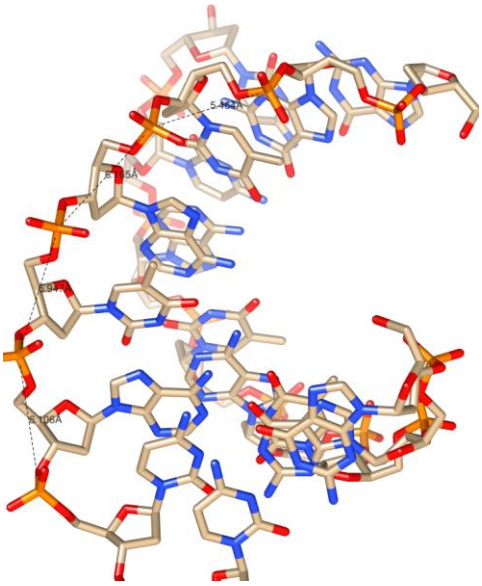


8. Summarize the factors that dictate DNA stability. Include stabilizing forces as well as structural arguments (torsion angles, etc.). Torsion angles are the angles in the sugar phosphate backbone. The stable conformations are the 3' endo and the 2' endo. The bond between the ribose and the base, called chi, can be cis or trans. The trans conformation, which orients that Watson and Crick face of the base away from the sugar, is highly favored. H-bonds form on the W-C face across DNA strands and permit selection of the 'correct' base pair. VDW forces allow the pi system of the bases to stack up and form a VERY stable helical core. Mg^{2+} ions interact with the phosphate backbone to stabilize the negative charge.

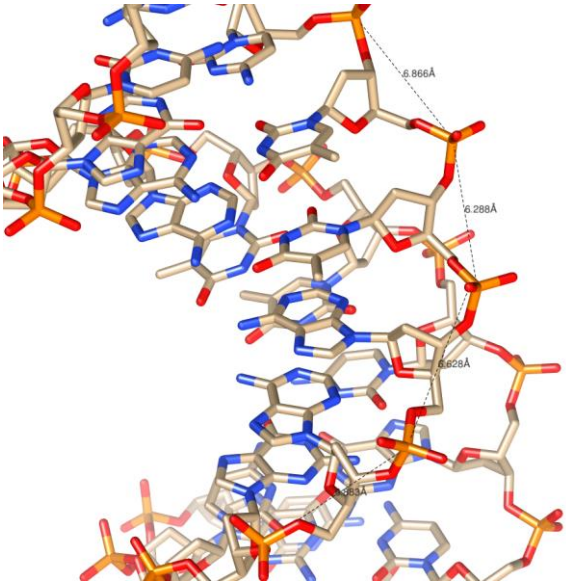
9. Hydrogen bonding patterns are important for DNA to obey Chargaff's Rule but do not play a very important role when considering the stability of dsDNA. Justify both of these statements. Chargaff's Rule necessitates that G-C and A-T base pairs are formed in typical B-form DNA. However, since the base pairing interactions occur through hydrogen bonds, they don't lead to a net stabilization of the folded helix relative to the single stranded molecules. This is because the H-bonds that form in base pairs are replaced by H-bonds with water in the ssDNA structures. The BIG reason that G-C base pairs stabilize DNA more than AT base pairs is because they form more favorable base stacking interactions.

10. Why is supercoiling necessary? What contributes to stress in a supercoiled DNA molecule and how is that stress relieved (so how is the Linking Number decreased)? It's necessary so that chromosomal DNA will fit into cells. It allows the otherwise long chains to DNA to form more compact structures. Increasing the number of twists results in more superhelical stress; this is relieved through writhing.

11. Use Chimera to view the 3D structures of A-form (pdbID 1VJ4) and B-form (pdbID 1BNA) DNA. I recommend turning off nucleic acid objects (Actions → Atoms/Bonds → Nucleic Acid Objects → Off) and hiding the ribbon form. **For this question, please create a word document that contains images supporting the answers to each question**. Please do not waste paper and print this unless you really want to; email it instead.
    a. In both structures, determine the distance between adjacent phosphorus atoms. You can use the Distance tool under Structure Analysis. Select the two atoms and then 'create'. For each structure, do this for P-P distances and average them. A Form → average = 5.9 angstroms B-Form → Average = 6.667 angstroms

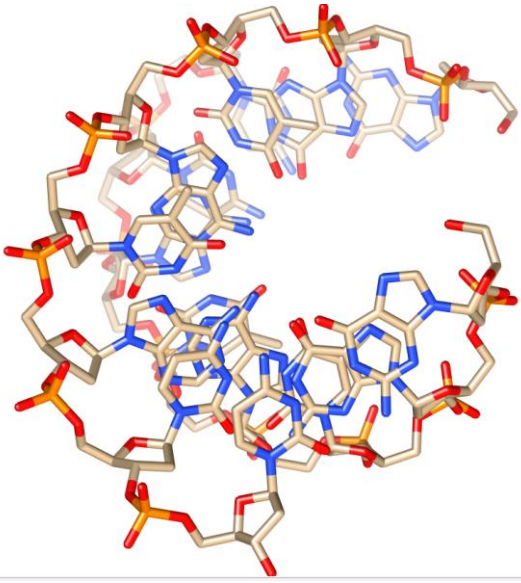


A-DNA

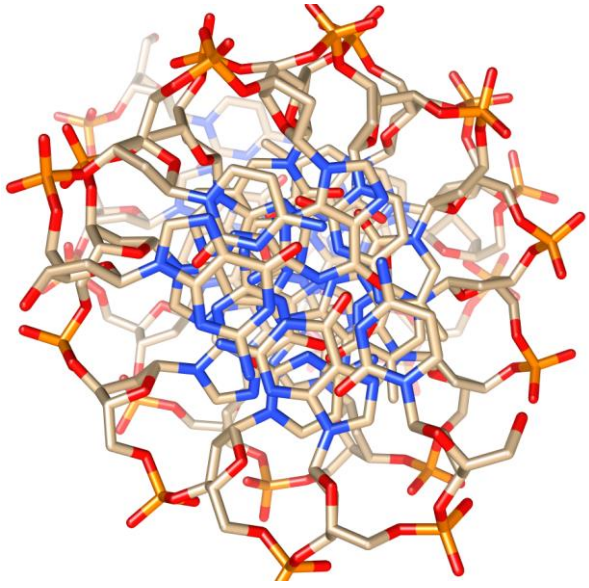


B-DNA

    b. Turn each molecule so that you're looking down the helical axis. What difference do you observe? A-DNA has a hole down the middle and the base pair are set off to one side, not centered like they are in B-DNA

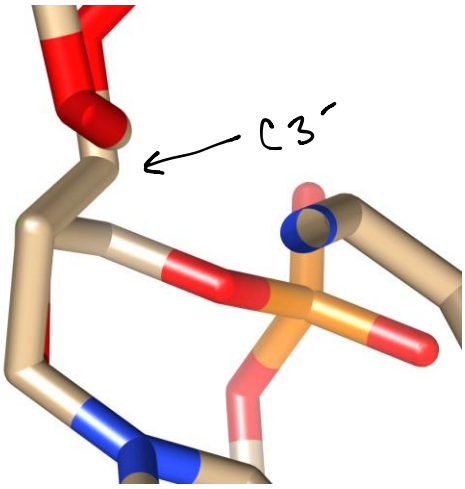


A-DNA

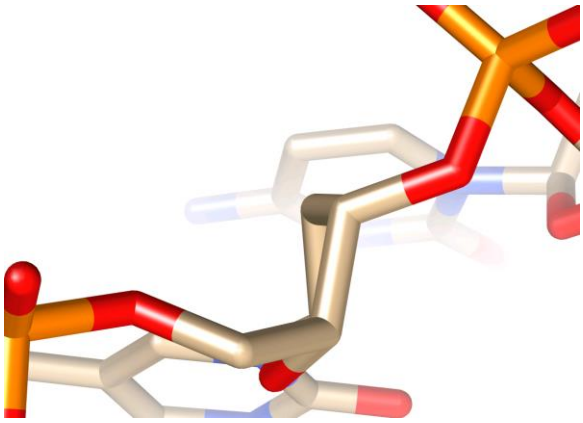


B-DNA

c. Zoom in on one of the deoxyribose units (if you select an atom, you can set it as your focus Actions→focus). Orient the sugar so that 4 of the atoms are roughly in the same plane. Which atom appears to be puckered out of the plane of the ring?



A-Form → C3' is puckered out



B-Form → C2' is puckered out – or depending on how you orient it, it might be the C1'

12. Using the tools you learned in Biochemistry Lab this week:
    a. Determine the nucleotide and amino acid sequence for glucokinase from Escherichia coli (strain K-12 sub-strain W3110). I recommend searching NCBI using the 'gene' option instead of 'nucleotide'. Make sure to list the nucleotide sequence from 5'→3' with the start codon at the 5' end. Note that there is a start codon (ATG) on the 5' end and a stop codon (TAA) on the 3' end.

```
ATGACAAAGTATGCATTAGTCGGTGATGTGGGCGGCACCAACGCACGTCTTGCTCTGTGTGATATTGCCAGTGGTGAAATCTCGCAGGC
TAAGACCTATTCAGGGCTTGATTACCCCAGCCTCGAAGCGGTCATTCGCGTTTATCTTGAAGAACATAAGGTCGAGGTGAAAGACGGCT
GTATTGCCATCGCTTGCCCAATTACCGGTGACTGGGTGGCGATGACCAACCATACCTGGGCGTTCTCAATTGCCGAAATGAAAAAGAAT
CTCGGTTTTAGCCATCTGGAAATTATTAACGATTTTACCGCTGTATCGATGGCGATCCCGATGCTGAAAAAAGAGCATCTGATTCAGTT
TGGTGGCGCAGAACCGGTCGAAGGTAAGCCTATTGCGGTTTACGGTGCCGGAACGGGGCTTGGGGTTGCGCATCTGGTCCATGTCGATA
AGCGTTGGGTAAGCTTGCCAGGCGAAGGCGGTCACGTTGATTTTGCGCCGAATAGTGAAGAAGAGGCCATTATCCTCGAAATATTGCGT
GCGGAAATTGGTCATGTTTCGGCGGAGCGCGTGCTTTCTGGCCCTGGGCTGGTGAATTTGTATCGCGCAATTGTGAAAGCTGACAACCG
CCTGCCAGAAAATCTCAAGCCAAAAGATATTACCGAACGCGCGCTGGCTGACAGCTGCACCGATTGCCGCCGCGCATTGTCGCTGTTTT
GCGTCATTATGGGCCGTTTTGGCGGCAATCTGGCGCTCAATCTCGGGACATTTGGCGGCGTGTTTATTGCGGGCGGTATCGTGCCGCGC
TTCCTTGAGTTCTTCAAAGCCTCCGGTTTCCGTGCCGCATTTGAAGATAAAGGGCGCTTTAAAGAATATGTCCATGATATTCCGGTGTA
TCTCATCGTCCATGACAATCCGGGCCTTCTCGGTTCCGGTGCACATTTACGCCAGACCTTAGGTCACATTCTGTAA
```

    b. From the gene page, you can quickly determine what genes are genetic neighbors (upstream and downstream of glucokinase in the genomic context window). ypdH is downstream and yfeo is upstream
    c. From the gene page, follow the KEGG link under "Links to other resources". This is on the right side of your screen. What metabolic Pathways are glucokinase involved in? List the top two. Glycolysis/gluconeogenesis and Galactose metabolism
    d. What amino acids are involved in nucleotide binding? List these in this format: Tyr4, meaning the tyrosine at the 4th position. Asp9, Val10, Gly11, Gly12, Asn14, Arg16, Asp100
    e. What conserved domains are present in this protein? NBD_sugar Kinase_SSP70_actin

f. Run a Blast search on the amino acid sequence. Are there any homologues in Yersinia pestis (the cause of the Bubonic Plague)? Yes

i. If yes, determine the % Identity and % Similarity (positives) between these two proteins. 79% Identical and 88% Similar (positive)

glucokinase [Yersinia pestis]

Sequence ID: ref|WP_011055412.1| Length: 324 Number of Matches: 1

▸ See 2 more title(s)

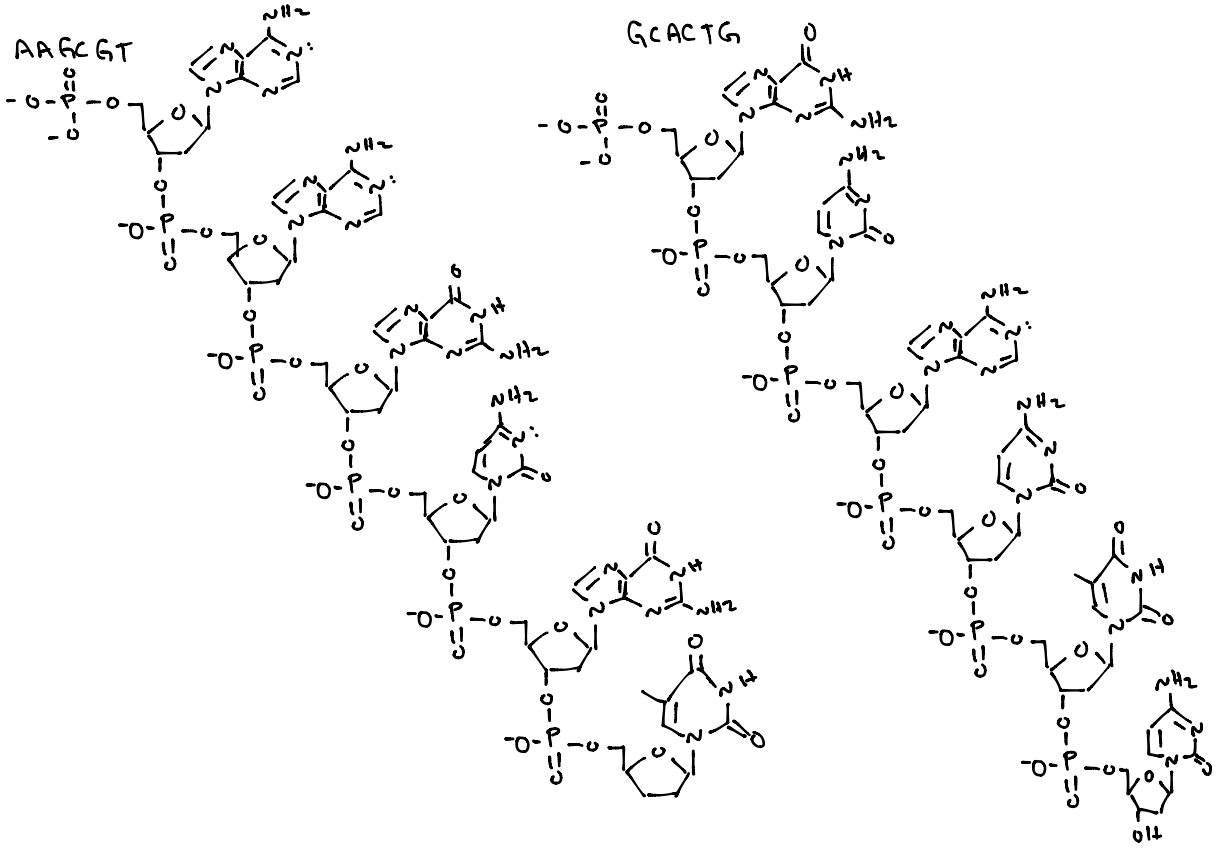
Range 1: 2 to 322 GenPept Graphics ▼ Next Match ▲ Previous Match

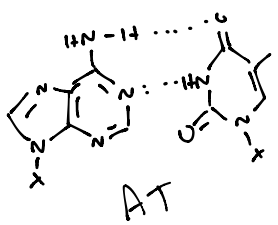
Score	Expect	Method	Identities	Positives	Gaps
541 bits(1394)	0.0	Compositional matrix adjust.	255/321(79%)	285/321(88%)	0/321(0%)

```
Query  1    MTKYALVGDVGGTNARLALCDIASGEISQAKTYSGLDYPSLEAVIRVYLEEHKVEVKDGC  60
            MT YALVGDVGGTNARLALC +A+GEI QAKTYSGL+Y SLE VI+ YL EH+ +V D C
Sbjct  2    MTTYALVGDVGGTNARLALCAVATGEILQAKTYSGLEYESLEDVIKQYLSEHQAKVTDAC  61

Query  61   IAIACPITGDWVAMTNHTWAFSIAEMKKNLGFSHLEIINDFTAVSMAIPMLKKEHLIQFG  120
            IAIACPITGDWVAMTNHTWAFSIA M++NLG  HLE+INDFTAVSMAIP+L  + ++QFG
Sbjct  62   IAIACPITGDWVAMTNHTWAFSIAAMQQNLGLDHLEVINDFTAVSMAIPVLPAQDVLQFG  121

Query  121  GAEPVEGKPIAVYGAGTGLGVAHLVHVDKRWVSLPGEGGHVDFAPNSEEEAIILEILRAE  180
            G +P  GKP+AVYGAGTGLGVAHLV+VD+RW+SL GEGGHVDFAPNSEEE  IL +LR E
Sbjct  122  GTQPQPGKPVAVYGAGTGLGVAHLVNVDRRWISLAGEGGHVDFAPNSEEEDQILAVLRQE  181

Query  181  IGHVSAERVLSGPGLVNLYRAIVKADNRLPENLKPKDITERALADSCTDCRRALSLFCVI  240
            +GHVSAERVLSGPGLVNLYRAIV +D RLPE L PKDIT RALADSCTDCRRALSLFCVI
Sbjct  182  LGHVSAERVLSGPGLVNLYRAIVISDARLPEKLAPKDITARALADSCTDCRRALSLFCVI  241

Query  241  MGRFGGNLALNLGTFGGVFIAGGIVPRFLEFFKASGFRAAFEDKGRFKEYVHDIPVYLIV  300
            MGRFGGNLALNL TFGGV+IAGGIVPRF+EFFKASGFRAAFEDKGRFK+++ DIPVY+I
Sbjct  242  MGRFGGNLALNLSTFGGVYIAGGIVPRFMEFFKASGFRAAFEDKGRFKDFLQDIPVYMIT  301

Query  301  HDNPGLLGSGAHLRQTLGHIL  321
            H  PGLLG+GA+LRQ LG+ L
Sbjct  302  HPQPGLLGAGAYLRQKLGYEL  322
```

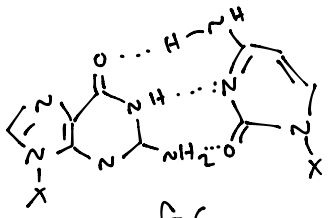
ii. List the amino acid sequence of this protein.

MMTTYALVGDVGGTNARLALCAVATGEILQAKTYSGLEYESLEDVIKQYLSEHQAKVTDACIAIACPITGDWVAMTNHTWAFSIAAMQQ
NLGLDHLEVINDFTAVSMAIPVLPAQDVLQFGGTQPQPGKPVAVYGAGTGLGVAHLVNVDRRWISLAGEGGHVDFAPNSEEEDQILAVL
RQELGHVSAERVLSGPGLVNLYRAIVISDARLPEKLAPKDITARALADSCTDCRRALSLFCVIMGRFGGNLALNLSTFGGVYIAGGIVP
RFMEFFKASGFRAAFEDKGRFKDFLQDIPVYMITHPQPGLLGAGAYLRQKLGYELSS





AT



GC