Molecular Biology—Transcription and Translation

Today we return to take another look at the yeast cystathionine beta synthase (CBS) protein. Below is the sequence of the yeast CBS gene and surrounding DNA. We will use this sequence for parts A and B of today's section.

The sequences of both strands of the DNA duplex are shown: the top strand reads 5' to 3' left to right (1 to 2040); the bottom, complimentary, strand reads 5' to 3' right to left (2040 to 1).

```
5'-  CAACTTCACCCAAGTAAGGATAATCAGCTCTGTCGTGACTGATAAATGCTATATCCGGCA
   1 ---------+---------+---------+---------+---------+---------+
3'-  GTTGAAGTGGGTTCATTCCTATTAGTCGAGACAGCACTGACTATTTACGATATAGGCCGT

     TATGCAGTCCACACGGCATTACCGTTTCACTAATTTATTGCCATCTTCCTCCACAGTTTT
  61 ---------+---------+---------+---------+---------+---------+
     ATACGTCAGGTGTGCCGTAATGGCAAAGTGATTAAATAACGGTAGAAGGAGGTGTCAAAA

     GCACCGAAAGGAAAAAAAGAAACCAACACCGAAAATTTTTTTCTCCTAAAGGTTAAAGTA
 121 ---------+---------+---------+---------+---------+---------+
     CGTGGCTTTCCTTTTTTTCTTTGGTTGTGGCTTTTAAAAAAGAGGATTTCCAATTTCAT

     AACGCAAGGCACTTCACCAGGCTTGTATATATAAATGTCGTGATGCTTCTATGCCAAAGT
 181 ---------+---------+---------+---------+---------+---------+
     TTGCGTTCCGTGAAGTGGTCCGAACATATATATTTACAGCACTACGAACATACGGTTTCA

     AAAAGGCAACACTTGAAGATTTCGTTGTAGGCCACTTGCTCAAAGGACATCTAGATAAAT
 241 ---------+---------+---------+---------+---------+---------+
     TTTTCCGTTGTGAACTTCTAAAGCAACATCCGGTGAACGAGTTTCCTGTAGATCTATTTA

     ACGACGTAAGAATAAAAATGACTAAATCTGAGCAGCAAGCCGATTCAAGACATAACGTTA
 301 ---------+---------+--------c+--------b+---------+---a-----+
     TGCTGCATTCTTATTTTTACTGATTTAGACTCGTCGTTCGGCTAAGTTCTGTATTGCAAT

     TCGACTTAGTTGGTAACACCCCATTGATCGCACTGAAAAAATTGCCTAAGGCTTTGGGTA
 361 ---------+----e----+--d------+---------+---------+---------+
     AGCTCAATCAACCATTGTGGGGTAACTAGCGTGACTTTTTTAACGGATTCCGAAACCCAT

     TCAAACCACAAATTTATGCTAAGCTGGAACTATACAATCCAGGTGGTTCCATCAAAGACA
 421 ---------+---------+---------+---------+---------+---------+
     AGTTTGGTGTTTAAATACGATTCGACCTTGATATGTTAGGTCCACCAAGGTAGTTTCTGT

     GAATTGCCAAGTCTATGGTGGAAGAAGCTGAAGCTTCCGGTAGAATTCATCCTTCCAGAT
 481 ---------+---------+---------+---------+---------+---------+
     CTTAACGGTTCAGATACCACCTTCTTCGACTTCGAAGGCCATCTTAAGTAGGAAGGTCTA

     CTACTCTGATCGAACCTACTTCTGGTAACACCGGTATCGGTCTAGCTTTAATCGGCGCCA
 541 ---------+---------+---------+---------+---------+---------+
     GATGAGACTAGCTTGGATGAAGACCATTGTGGCCATAGCCAGATCGAAATTAGCCGCGGT
```

```
      TCAAAGGTTACAGAACTATCATCACCTTGCCGGAAAAAATGTCTAACGAGAAAGTTTCTG
 601  ---------+---------+---------+---------+---------+---------+
      AGTTTCCAATGTCTTGATAGTAGTGGAACGGCCTTTTTTACAGATTGCTCTTTCAAAGAC

      TCCTAAAGGCTCTGGGTGCTGAAATCATCAGAACTCCAACTGCTGCTGCCTGGGATTCTC
 661  ---------+---------+---------+---------+---------+---------+
      AGGATTTCCGAGACCCACGACTTTAGTAGTCTTGAGGTTGACGACGACGGACCCTAAGAG

      CAGAATCACATATTGGTGTTGCTAAGAAGTTGGAAAAAGAGATTCCTGGTGCTGTTATAC
 721  ---------+---------+---------+---------+---------+---------+
      GTCTTAGTGTATAACCACAACGATTCTTCAACCTTTTTCTCTAAGGACCACGACAATATG

      TTGACCAATATAACAATATGATGAACCCAGAAGCTCATTACTTTGGTACTGGTCGCGAAA
 781  ---------+---------+---------+---------+---------+---------+
      AACTGGTTATATTGTTATACTACTTGGGTCTTCGAGTAATGAAACCATGACCAGCGCTTT

      TCCAAAGACAGCTAGAAGACTTGAATTTATTTGATAATCTACGCGCTGTTGTTGCTGGTG
 841  ---------+---------+---------+---------+---------+---------+
      AGGTTTCTGTCGATCTTCTGAAGTTAAATAAAGTATTAGATGCGCGACAACAACGACCAC

      CTGGTACTGGTGGGACTATTAGCGGTATTTCCAAGTACTTGAAAGAACAGAATGATAAGA
 901  ---------+---------+---------+---------+---------+---------+
      GACCATGACCACCCTGATAATCGCCATAAAGGTTCATGAACTTTCTTGTCTTACTATTCT

      TCCAAATC**GTT**GGTGCTGACCCATTCGGTTCAATTTTAGCCCAACCTGAAAACTTGAATA
 961  ---------**f**---------+---------+---------+---------+---------+
      AGGTTTAG**CAA**CCACGACTGGGTAAGCCAAGTTAAAATCGGGTTGGACTTTTGAACTTAT

      AGACTGATATCACTGACTACAAAGTTGAGGGTATT**GGT**TATGATTTTGTTCCTCAGGTTT
1021  ---------+---------+---------+------**g**--+---------+---------+
      TCTGACTATAGTGACTGATGTTTCAACTCCCATAA**CCA**ATACTAAAACAAGGAGTCCAAA

      TGGACAGAAAATTAATTGATGTTTGGTATAAGACAGACGACAAGCCTTCTTTCAAATACG
1081  ---------+---------+---------+---------+---------+---------+
      ACCTGTCTTTTAATTAACTACAAACCATATTCTGTCTGCTGTTCGGAAGAAAGTTTATGC

      CCAGACAATTGATTTCTAACGAAGGTGTCTTGGTGGGTGGTTCTTCCGGTTCTACCTTCA
1141  ---------+---------+---------+---------+---------+---------+
      GGTCTGTTAACTAAAGATTGCTTCCACAGAACCACCCACCAAGAAGGCCAAGATGGAAGT

      CTGCGGTTGTGAAATACTGTGAAGACCACCCTGAACTGACTGAAGATGATGTCATTGTTG
1201  ---------+---------+---------+---------+---------+---------+
      GACGCCAACACTTTATGACACTTCTGGTGGGACTTGACTGACTTCTACTACAGTAACAAC

      CCATATTCCCAGATTCCATCAGGTCGTACCTAACCAAATTCGTCGATGACGAATGGTTGA
1261  ---------+---------+---------+---------+---------+---------+
      GGTATAAGGGTCTAAGGTAGTCCAGCATGGATTGGTTTAAGCAGCTACTGCTTACCAACT

      AAAAGAACAATTTGTGGGATGATGACGTGTTGGCCCGTTTTGACTCTTCAAAGCTGGAGG
1321  ---------+---------+---------+---------+---------+---------+
      TTTTCTTGTTAAACACCCTACTACTGCACAACCGGGCAAAACTGAGAAGTTTCGACCTCC

      CTTCGACGACAAAATACGCTGATGTGTTTGGTAACGCTACTGTAAAGGATCTTCACTTGA
1381  ---------+---------+---------+---------+---------+---------+
      GAAGCTGCTGTTTTATGCGACTACACAAACCATTGCGATGACATTTCCTAGAAGTGAACT
```

```
      AACCGGTTGTTTCCGTTAAGGAAACCGCTAAGGTCACTGATGTTATCAAGATATTAAAAG
1441  ---------+---------+---------+---------+---------+---------+
      TTGGCCAACAAAGGCAATTCCTTTGGCGATTCCAGTGACAACAATAGTTCTATAATTTTC

      ACAATGGCTTTGACCAATTGCCTGTGTTGACTGAAGACGGCAAGTTGTCTGGTTTAGTTA
1501  ---------+---------+---------+---------+---------+---------+
      TGTTACCGAAACTGGTTAACGGACACAACTGACTTCTGCCGTTCAACAGACCAAATCAAT

      CTCTCTCTGAGCTTCTAAGAAAACTATCAATCAATAATTCAAACAACGACAACACTATAA
1561  ---------+---------+---------+---------+---------+---------+
      GAGAGAGACTCGAAGATTCTTTTGATAGTTAGTTATTAAGTTTGTTGCTGTTGTGATATT

      AGGGTAAATACTTGGACTTCAAGAAATTAAACAATTTCAATGATGTTTCCTCTTACAACG
1621  ---------+---------+---------+---------+---------+---------+
      TCCCATTTATGAACCTGAAGTTCTTTAATTTGTTAAAGTTACTACAAAGGAGAATGTTGC

      AAAATAAATCCGGTAAGAAGAAGTTTATTAAATTCGATGAAAACTCAAAGCTATCTGACT
1681  ---------+---------+---------+---------+---------+---------+
      TTTTATTTAGGCCATTCTTCTTCAAATAATTTAAGCTACTTTTGAGTTTCGATAGACTGA

      TGAATCGTTTCTTTGAAAAAAACTCATCTGCCGTTATCACTGATGGCTTGAAACCAATCC
1741  ---------+---------+---------+---------+---------+---------+
      ACTTAGCAAAGAAACTTTTTTTGAGTAGACGGCAATAGTGACTACCGAACTTTGGTTAGG

      ATATCGTTACTAAGATGGATTTACTGAGCTACTTAGCATAAATAAGAACCCACGCTTCAA
1801  ---------+---------+---------+---------+---------+---------+
      TATAGCAATGATTCTACCTAAATGACTCGATGAATCGTATTTATTCTTGGGTGCGAAGTT

      ATAAAAGCAAACATAGAAGCAAAATCCGTCATTCCTTTCCTATTCAATTGCACCGTTCTC
1861  ---------+---------+---------+---------+---------+---------+
      TATTTTCGTTTGTATCTTCGTTTTAGGCAGTAAGGAAAGGATAAGTTAACGTGGCAAGAG

      TTTATATAACTACTTAATTAAATAGCGCCTATACGAAGCAGCATTGTTCTATTATTTTTA
1921  ---------+---------+---------+---------+---------+---------+
      AAATATATTGATGAATTAATTTATCGCGGATATGCTTCGTCGTAACAAGATAATAAAAAT

      CAAATTCCTTATCATGCATGCATCACATCAGTGTTTGAATCTGTTAACTTTTCACTTTAT
1981  ---------+---------+---------+---------+---------+---------+
      GTTTAAGGAATAGTACGTACGTAGTGTAGTCACAAACTTAGACAATTGAAAAGTGAAATA
```

A. Transcription and Translation—Practice

RNA Polymerase complex binds to the two TATA-type elements underlined and bolded above. Once bound, RNA polymerase starts making mRNA somewhere between 40 and 120 base pairs downstream of the last TATA element.

1. If transcription starts 40 base pairs downstream of the last TATA element, at base pair #253, write the sequence of the first 10 nucleotides of the resulting mRNA. Label 5' and 3' ends.
   5'-UUGAAGAUUU...-3'

2. If transcription starts 90 base pairs downstream of the last TATA element, at base pair #303, write the sequence of the first 10 nucleotides of the resulting mRNA. Label 5' and 3' ends.
   5'-GACGUAAGAA...-3'

3. In each case, what are the first eight amino acids of the resulting protein? In this case, does the transcription start site influence the sequence of the resulting protein?
N-Met-Thr-Lys-Ser-Glu-Gln-Gln-Ala-…C
Transcription start site does not influence the sequence of protein, because the first ATG occurs after either of the start sites.

4. Does translation terminate at the TAA at the underlined position 353 (**a**)? Why or why not?
No, this TAA is not in frame of translation.

5. How would your answer to 3 change if the C/G base pair at position 339 (**b**, bold) was deleted? What effect, if any, do you expect this mutation to have on the resulting CBS protein?
Amino acid 8, Ala would be changed to Pro.

6. How would your answer to 3 change if an A/T base pair was added between 328 & 329 (**c**, bold)? What effect, if any, do you expect this mutation to have on the resulting CBS protein?
Early stop would be created, and the sequence would read
N-Met-Thr-Lys-Ser-C

7. How would your answer to 3 change if the A/T base pair at position 383 (**d**, bold) were changed to a G/C? What effect, if any, do you expect this mutation to have on the resulting CBS protein?
No change for answer in 3. And no change for the protein—silent mutation.

8. Give a single base change (substitution, deletion, or addition of a single base and it's partner on the other strand) that would cause termination of the polypeptide chain at TAA codon 374 (**e**, underlined).
Insert a base pair anywhere between the ATG and the TAA in question.

9. Give an example of a nonsense mutation (codon --> stop codon).
TGT→TGA (cys→stop)

10. Give an example of missense mutation (codon --> codon for another amino acid).
AAT→AAA (ans→lys)

11. Give an example of silent mutation (codon ---> codon for the same amino acid).
AGA→CGA (arg→arg)

12. A mutation in what codon(s) always results in a change in the primary sequence of the resulting protein?
Trp, Met—only one codon per each of these amino acids.

13. The actual stop codon for CBS is at bases 1839-1841 (underlined). What would the effect of the A1841C mutation be on the CBS protein?
The protein would acquire a "tail"—there is another TAA stop codon 1935-1937.

B. Transcription and Translation—Functional effects

Recall that yeast that lack CBS protein can not synthesize the amino acid cysteine. Mutant versions of the human CBS protein can lead to a very serious disease called homocystinuria. Some symptoms include mental retardation, early strokes and heart disease.

Recall also that in Section 3 we encountered a crystallographic model of a form of CBS protein. Consider the model again. This model is of a truncated form of the human CBS protein. Yeast and human proteins are very similar (72%), and some sections are identical (38%).

I. Mutation 1-- Treatable

1. Take a look at bases 969-971 (**f**, bold).
    a. What amino acid is encoded?
       Val
    b. What kind of amino acid is it?
       hydrophobic
    c. Where in a protein might you expect to find it?
       Inside the protein, with other hydrophobics
In the human CBS protein, this amino acid corresponds to Ile278.

2. What type of amino acid is Ile? Would you expect to find it in the same environment as the amino acid in 1 or not? Explain.
    Also hydrophobic—same environment.
Choose "Isoleucine 278" on the crystallographic model exercise.

3. What type of secondary structure element is it located in?
    β-sheet
4. What type of amino acids would you expect to surround it?
    Other hydrophobic residues
Choose "Isoleucine 278 and Neighbors" on the crystallographic model exercise.

5. What amino acids is Ile278 in contact with?
    Ala, Ile, Leu

6. What type of amino acids are those?
    All hydrophobic

A human mutation leading to a serious but somewhat treatable case of homocystinurea is caused by the equivalent of G969A and T970C mutations in yeast.

7. What amino acid would be created in yeast with those substitutions?
    Thr
8. Is a mutation causing this amino acid substitution more likely in humans or in yeast? Why?
    More likely in humans—only one change required.

9. What would be the effect of this mutation be on the tertiary structure of CBS? Why?
    Structure less strongly held together—hydrophilic amino acid inserted into a hydrophobic cluster would cause the cluster to not associate as strongly as before..

The only known treatment for homocystinurea is to supplement patient's diet with large quantities of vitamin B_6. B_6 is the co-factor for human CBS.

10. In general, why would adding more co-factor to a deficient enzyme help restore some of the function?
    The co-factor can saturate the enzyme, such that the bottle neck step is the catalysis, and not complex formation.
11. How do you suppose does the treatment overcome the deficiency in the mutant CBS?
    It is possible that the bound co-factor holds this complex tighter together, overcoming some of the problem introduced by the mutation.

II. Mutation 2-- Untreatable

12. Take a look at bases 1056-1058 (**g**, bold).
    a. What amino acid is encoded?
       Gly
    b. What kind of amino acid is it? Does it have any special properties?
       This is a smallest amino acid.
    c. Where in a protein might you expect to find it?
       We would expect to see it in tight spaces in the protein.

In the human CBS protein, this amino acid corresponds to Gly307. Choose "Glycine 307" on the crystallographic model exercise.

13. Does it belong to any secondary structure element?
    No secondary structure
14. What type of amino acids would you expect to surround it?
    Hydrophobics—inside the protein.

Choose "Glycine 307 and Neighbors" on the crystallographic model exercise.

15. What amino acids is Gly307 in contact with?
    Tyr, Ile

16. What type of amino acids are those?
    Hydrophobic and large

A human mutation leading to a severe and untreatable case of homocystinurea is caused by the equivalent of G1056A mutation in yeast.

17. What amino acid would be created in yeast with that substitution?
    Ser
18. Is a mutation causing this amino acid substitution more likely in humans or in yeast? Why?
    Same likelihood—one mutation required in each case.

19. What would be the effect of this mutation be on the tertiary structure of CBS? Why?
    It would be less tightly held together because Ser might disrupt the tight packing and also disrupt the hydrophobic pocket.

This mutation in humans does not respond to the vitamin B$_6$ treatment.

20. What are the possible reasons that one mutation would respond to treatment and another would not?
    Responsiveness to treatment is likely influenced by the location of the affected amino acid in the protein. Location in the 3D structure of the protein influences the role of a particular amino acid in the function of the protein.

C. Transcription and Translation—Conclusions

1. Where is all the information needed to properly execute transcription and translation recorded?
   DNA

2. Is all this information accessible to the same machinery?
   No—different machinery reads each.

3. Is the stop codon in position 1839 where transcription of the CBS gene stops? Why or why not?
   No—stop codon is read by the ribosome, and transcription termination is enacted by the RNA Pol complex.

The Genetic Code

	U	C	A	G	
U	UUU phe	UCU ser	UAU tyr	UGU cys	U
	UUC phe	UCC ser	UAC tyr	UGC cys	C
	UUA leu	UCA ser	UAA STOP	UGA STOP	A
	UUG leu	UCG ser	UAG STOP	UGG trp	G
C	CUU leu	CCU pro	CAU his	CGU arg	U
	CUC leu	CCC pro	CAC his	CGC arg	C
	CUA leu	CCA pro	CAA gln	CGA arg	A
	CUG leu	CCG pro	CAG gln	CGG arg	G
A	AUU ile	ACU thr	AAU asn	AGU ser	U
	AUC ile	ACC thr	AAC asn	AGC ser	C
	AUA ile	ACA thr	AAA lys	AGA arg	A
	AUG met	ACG thr	AAG lys	AGG arg	G
G	GUU val	GCU ala	GAU asp	GGU gly	U
	GUC val	GCC ala	GAC asp	GGC gly	C
	GUA val	GCA ala	GAA glu	GGA gly	A
	GUG val	GCG ala	GAG glu	GGG gly	G

STRUCTURES OF AMINO ACIDS
at pH 7.0
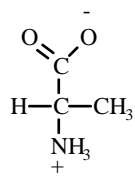
ALANINE
(ala)
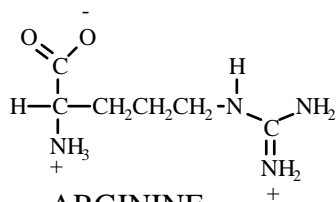
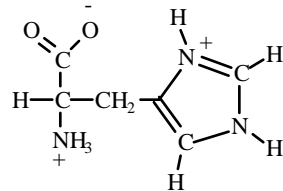ARGININE
(arg)
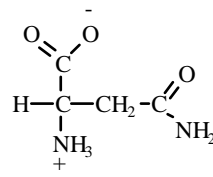
ASPARAGINE
(asN)
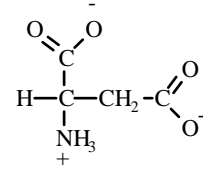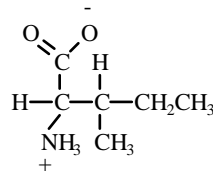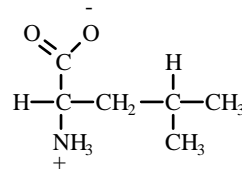
ASPARTIC ACID
(asp)
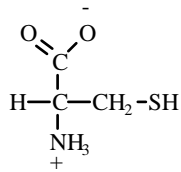
CYSTEINE
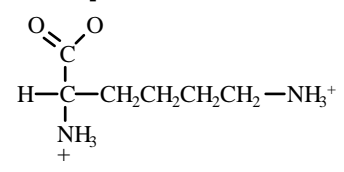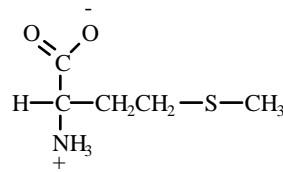(cys)
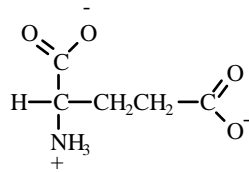
GLUTAMIC ACID
(glu)
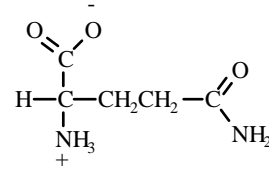
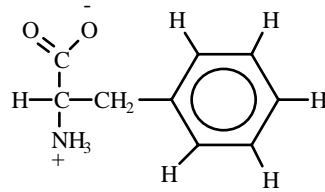GLUTAMINE
(glN)
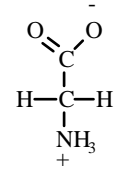
GLYCINE
(gly)
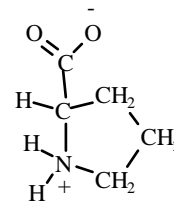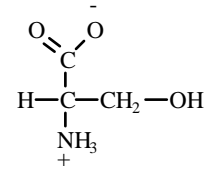
HISTIDINE
(his)

ISOLEUCINE
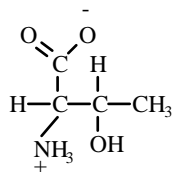(ile)

LEUCINE
(leu)

LYSINE
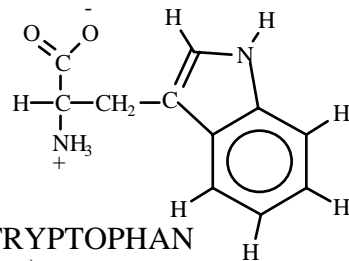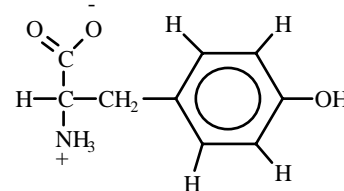(lys)

METHIONINE
(met)
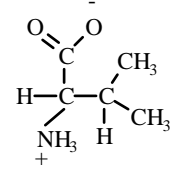
PHENYLALANINE
(phe)

PROLINE
(pro)

SERINE
(ser)

THREONINE
(thr)

TRYPTOPHAN
(trp)

TYROSINE
(tyr)

VALINE
(val)