DNA repair and MSH2 gene

Types of DNA Damage Repair

Direct Repair of Damage

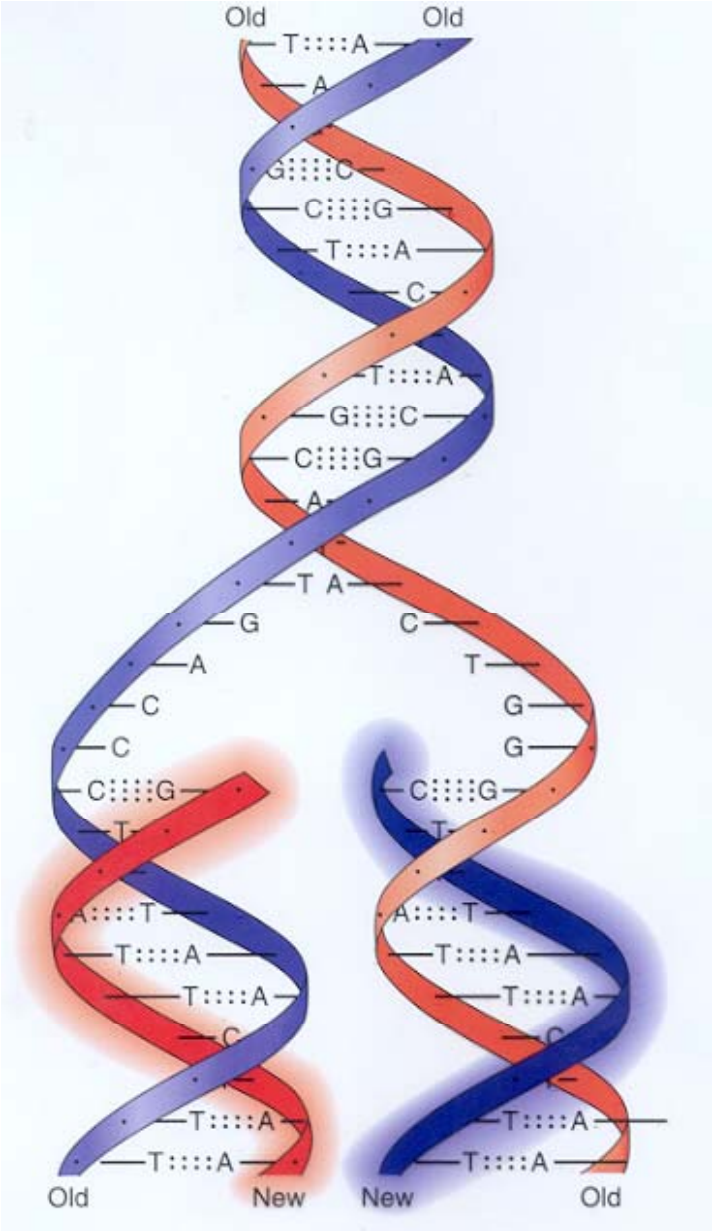
Mismatch Repair

Base Excision Repair

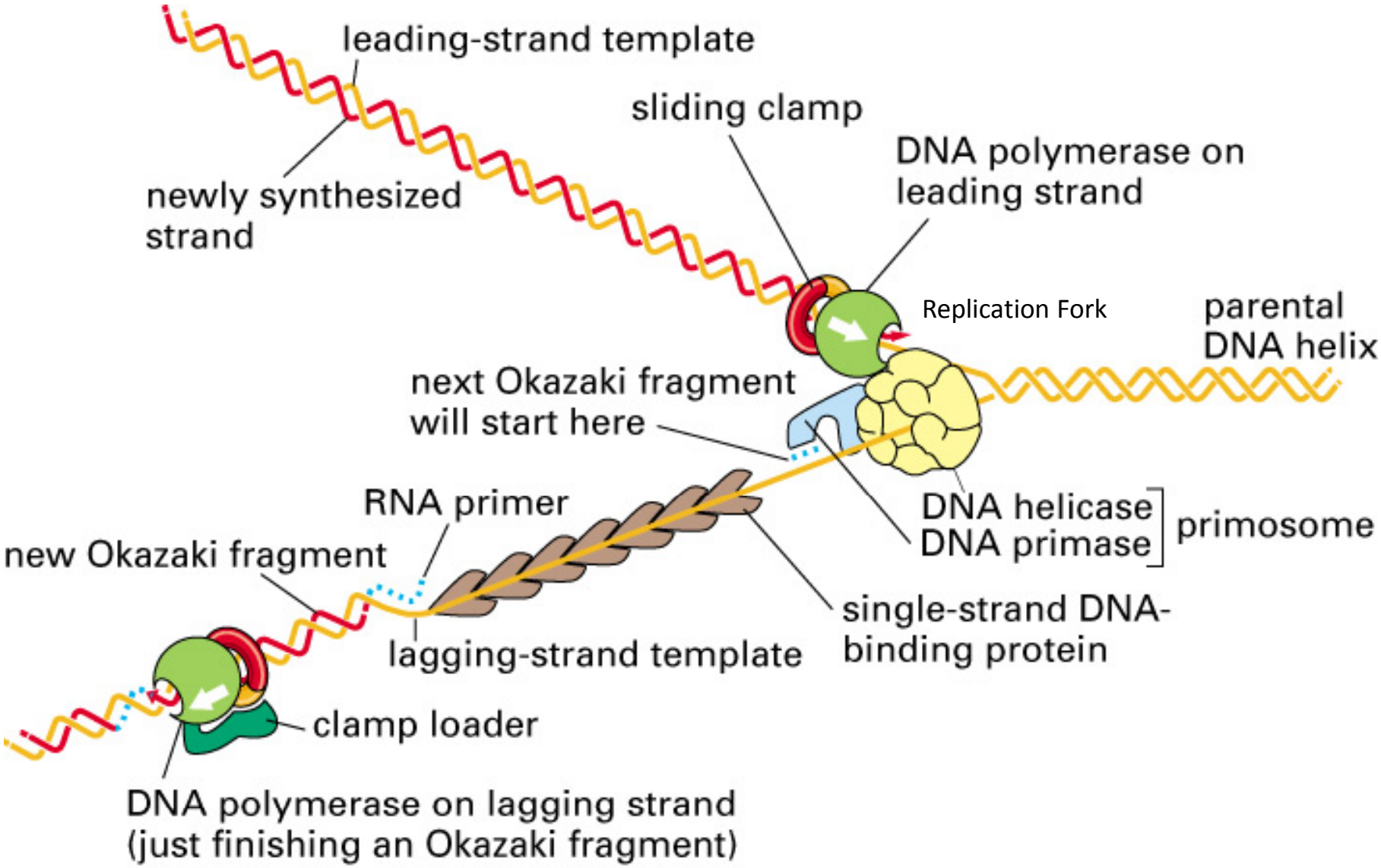
Nucleotide Excision Repair

Double-strand Break Repair

Damage Bypass



Review of Replication



From, Molecular Biology of the Cell

Mechanism of Mismatch repair

Recognition

Protein Binding

Incision

Degradation

Synthesis of new strand

Complement

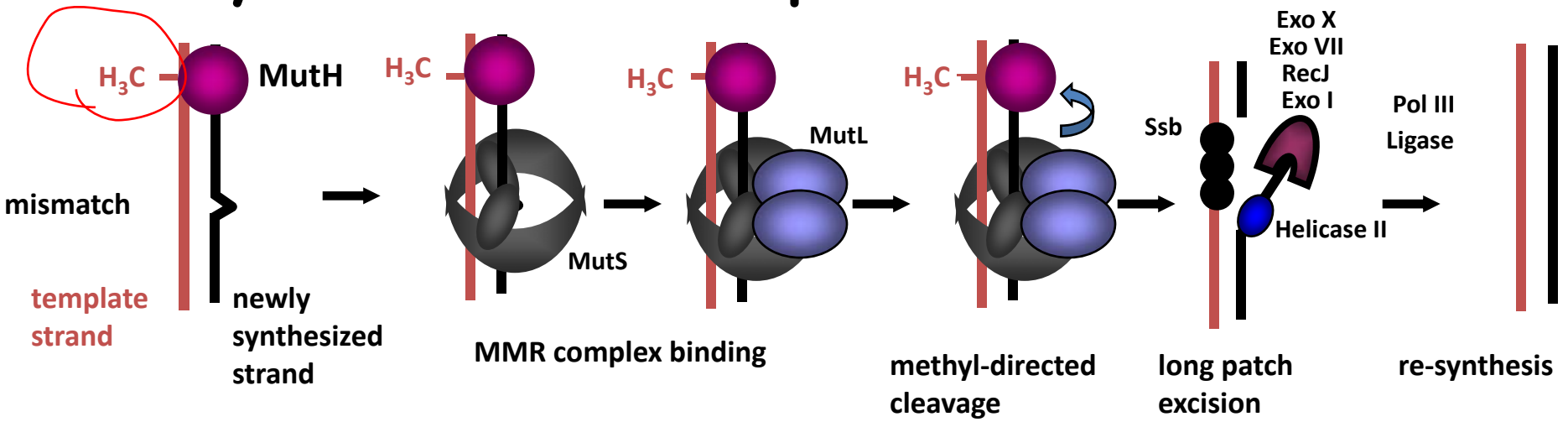
Closing gaps

Genes Encoding Enzymes of Mismatch Repair

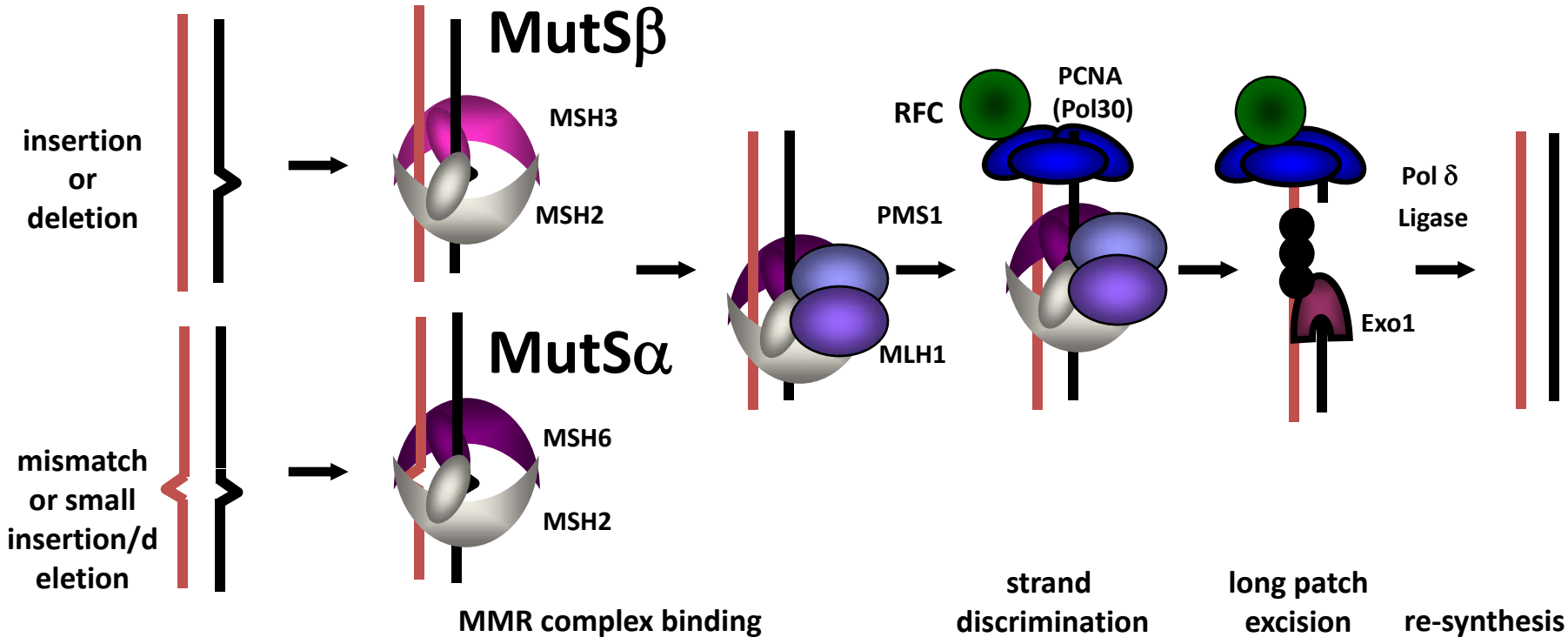
E. coli	*S. cerevisiae*	Human	Functions of Eukaryotic Proteins
MutS	MSH2	MSH2	MutSα (with MSH6; 80-90%); MutSβ (with MSH3)
"	MSH3	MSH3	MutSβ (with MSH2); repair of larger loops
"	MSH6	MSH6	MutSα (with MSH2); repair of mismatches and small loops
MutL	MLH1	MLH1	Forms heterodimers with the other three MutL homologs
"	PMS1	PMS2	MutLα (90%); Mismatch repair; endonuclease motif
"	MLH2	PMS1	MutLβ; Role unknown
"	MLH3	MLH3	MutLγ; Mismatch repair; endonuclease motif
MutH	?	?	?
uvrD	?	?	?
?	Exonuclease I	Exonuclease I	Excision (5' to 3' polarity)
?	RFC, PCNA, Polγ	RFC, PCNA, Polγ	Nick identification; gap filling

The eukaryotic genes are homologs of the corresponding *E. coli* genes both in terms of amino acid sequence and in terms of functional similarities. Whereas MutS and MutL function as homodimers, the eukaryotic proteins function as heterodimers. Heterodimers of MutS homologs are responsible for initial recognition of mismatches and small insertions/deletions, and heterodimers of MutL homologs interact with the resulting complex, as in *E. coli*.

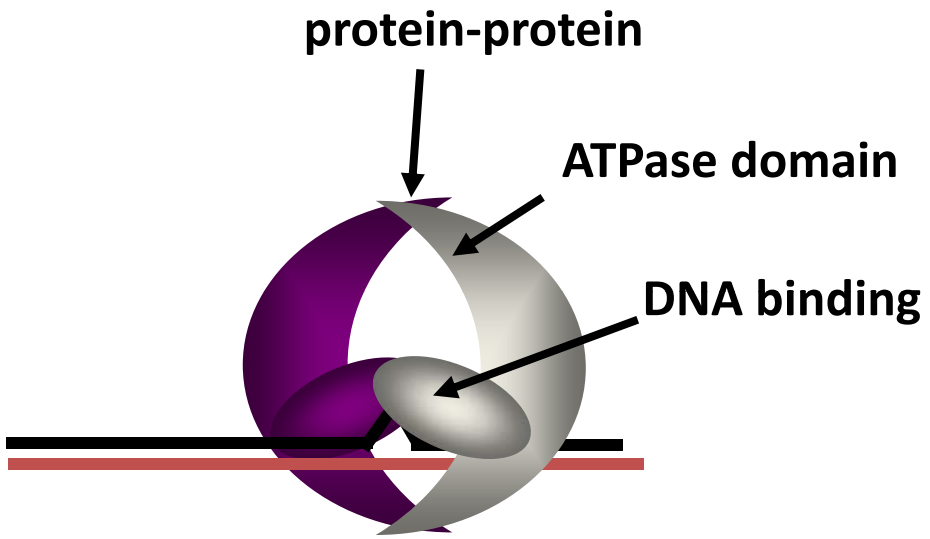
Prokaryotic Mismatch Repair



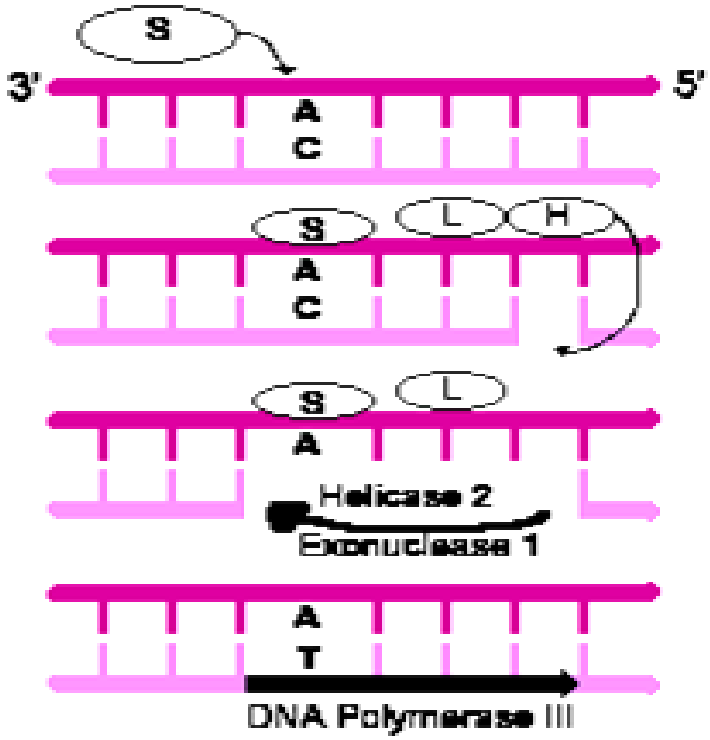
Eukaryotic Mismatch Repair

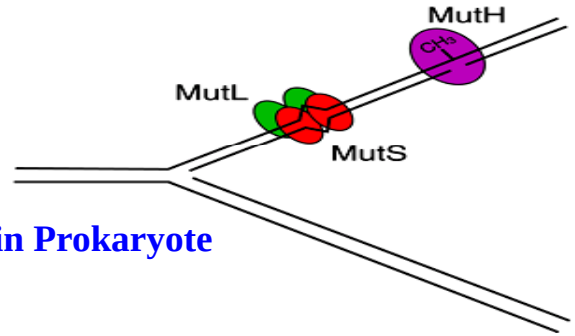


Msh2p Important Domains

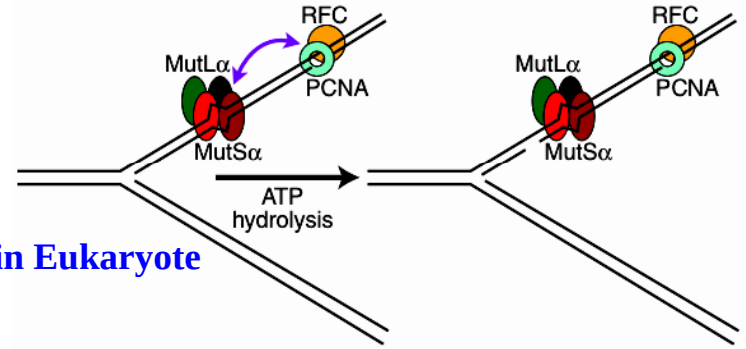


- Structural integrity residues (throughout)
- DNA binding region
- Protein-protein interacting regions
- ATPase domain



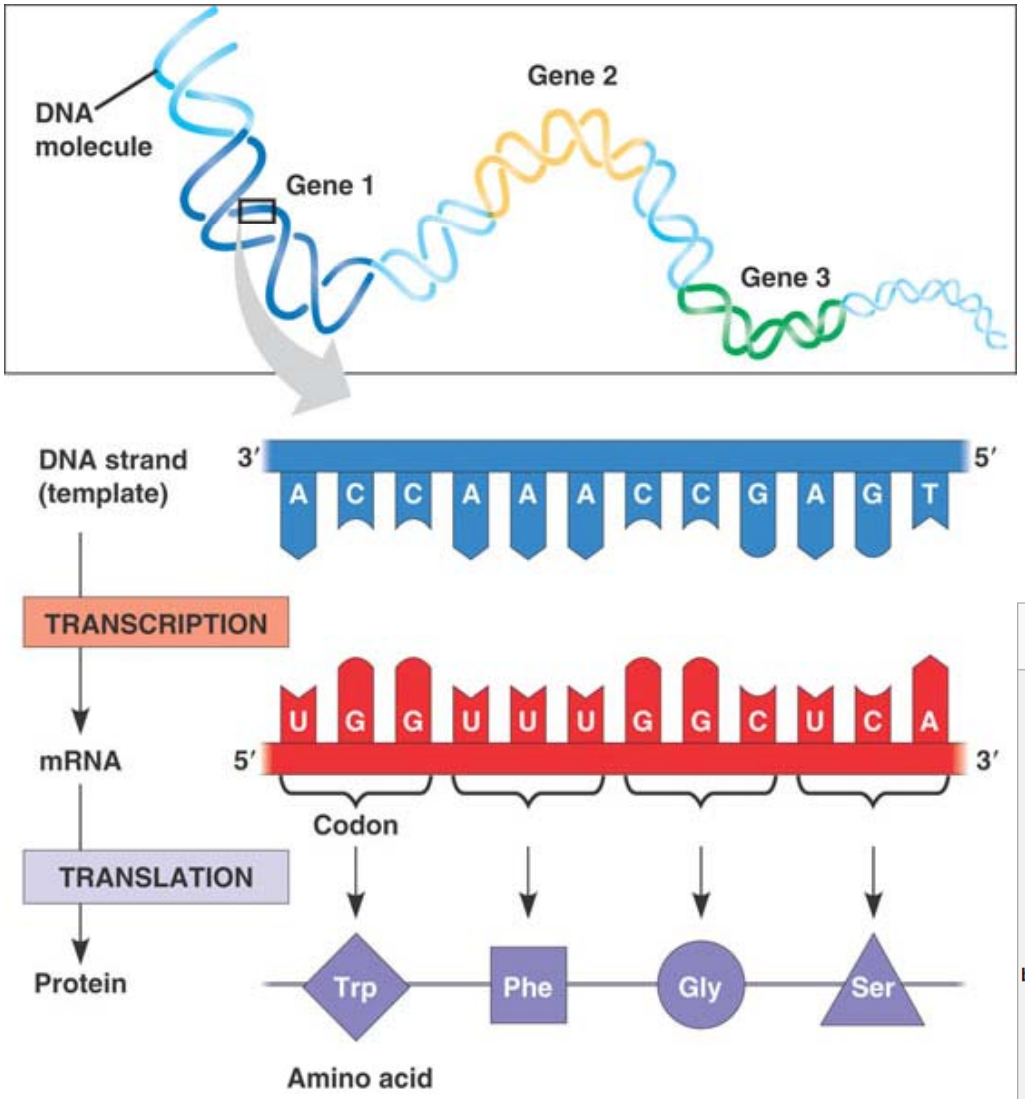


MMR in Prokaryote

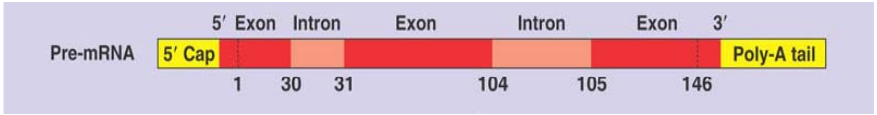


MMR in Eukaryote

Review of Gene Expression



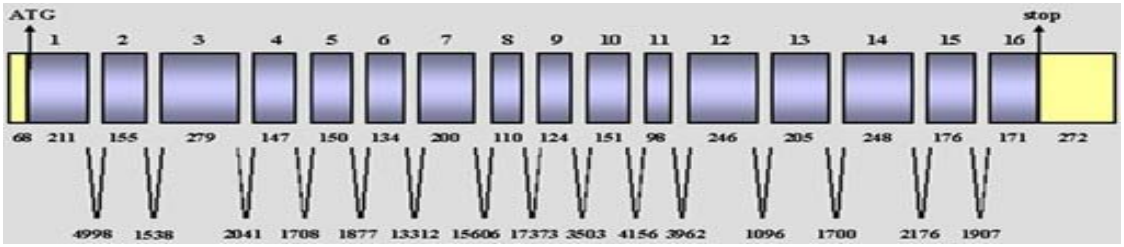
ORF Open Reading Frame



		2nd base			
		U	C	A	G
1st base	U	UUU (Phe/F) Phenylalanine	UCU (Ser/S) Serine	UAU (Tyr/Y) Tyrosine	UGU (Cys/C) Cysteine
		UUC (Phe/F) Phenylalanine	UCC (Ser/S) Serine	UAC (Tyr/Y) Tyrosine	UGC (Cys/C) Cysteine
		UUA (Leu/L) Leucine	UCA (Ser/S) Serine	UAA Ochre (*Stop*)	UGA Opal (*Stop*)
		UUG (Leu/L) Leucine	UCG (Ser/S) Serine	UAG Amber (*Stop*)	UGG (Trp/W) Tryptophan
	C	CUU (Leu/L) Leucine	CCU (Pro/P) Proline	CAU (His/H) Histidine	CGU (Arg/R) Arginine
		CUC (Leu/L) Leucine	CCC (Pro/P) Proline	CAC (His/H) Histidine	CGC (Arg/R) Arginine
		CUA (Leu/L) Leucine	CCA (Pro/P) Proline	CAA (Gln/Q) Glutamine	CGA (Arg/R) Arginine
		CUG (Leu/L) Leucine	CCG (Pro/P) Proline	CAG (Gln/Q) Glutamine	CGG (Arg/R) Arginine
	A	AUU (Ile/I) Isoleucine	ACU (Thr/T) Threonine	AAU (Asn/N) Asparagine	AGU (Ser/S) Serine
		AUC (Ile/I) Isoleucine	ACC (Thr/T) Threonine	AAC (Asn/N) Asparagine	AGC (Ser/S) Serine
		AUA (Ile/I) Isoleucine	ACA (Thr/T) Threonine	AAA (Lys/K) Lysine	AGA (Arg/R) Arginine
		AUG (Met/M) Methionine, *Start* [A]	ACG (Thr/T) Threonine	AAG (Lys/K) Lysine	AGG (Arg/R) Arginine
	G	GUU (Val/V) Valine	GCU (Ala/A) Alanine	GAU (Asp/D) Aspartic acid	GGU (Gly/G) Glycine
		GUC (Val/V) Valine	GCC (Ala/A) Alanine	GAC (Asp/D) Aspartic acid	GGC (Gly/G) Glycine
		GUA (Val/V) Valine	GCA (Ala/A) Alanine	GAA (Glu/E) Glutamic acid	GGA (Gly/G) Glycine
		GUG (Val/V) Valine	GCG (Ala/A) Alanine	GAG (Glu/E) Glutamic acid	GGG (Gly/G) Glycine

Flow of Genetic Information: Gene to Protein

Human MSH2 gene has 16 exons, coding for one protein with three functional domains:



hMSH2 gene

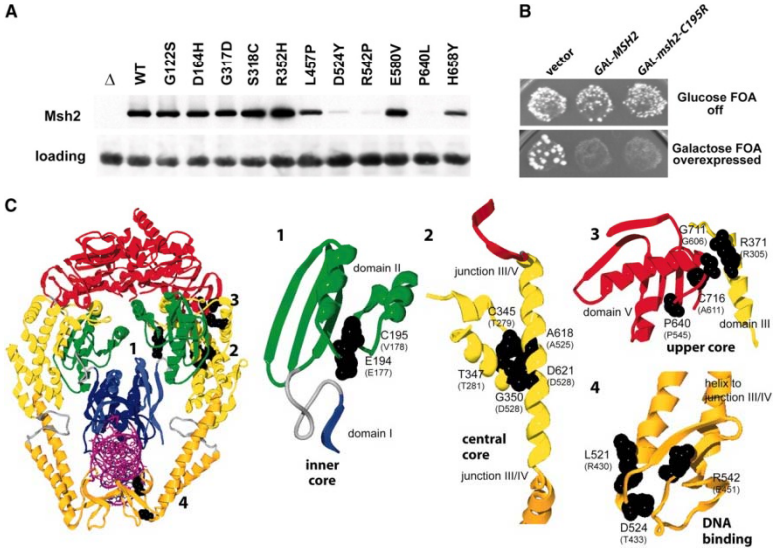
hMSH2 gene product (protein)

Mutations

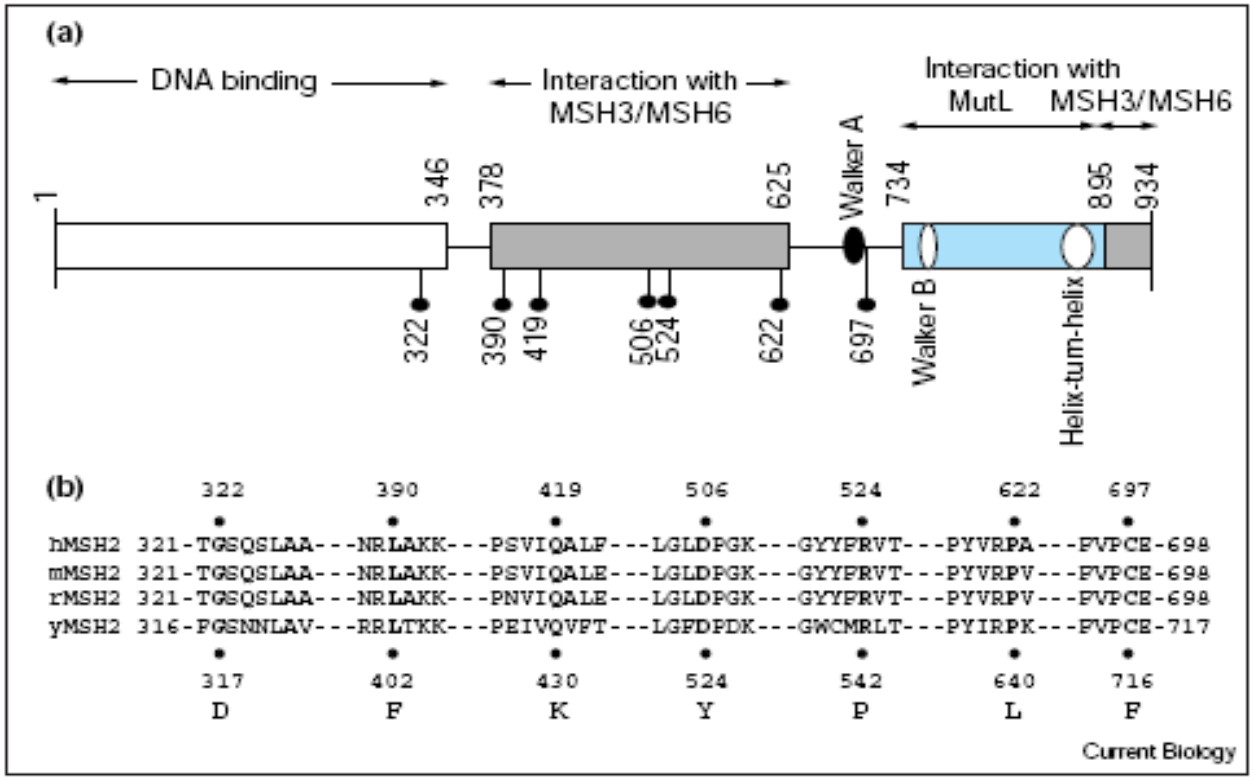
Missense

Nonsense

No protein



AMINO ACID			AMINO ACID		
Aspartic acid	Asp	D	Alanine	Ala	A
Glutamic acid	Glu	E	Glycine	Gly	G
Arginine	Arg	R	Valine	Val	V
Lysine	Lys	K	Leucine	Leu	L
Histidine	His	H	Isoleucine	Ile	I
Asparagine	Asn	N	Proline	Pro	P
Glutamine	Gln	Q	Phenylalanine	Phe	F
Serine	Ser	S	Methionine	Met	M
Threonine	Thr	T	Tryptophan	Trp	W
Tyrosine	Tyr	Y	Cysteine	Cys	C



Representations of MSH2. (a) Putative functional regions of hMSH2. An amino-terminal DNA-binding region, a region that interacts with the mismatch-repair protein MutL (blue), and a carboxy-terminal MutS dimerization region are suggested by studies of deletion mutants of *E. coli* MutS [20]. Amino acids 827–846 comprise a helix-turn-helix motif suggested to interact with yMSH6 [21]. Studies of hMSH2 have implicated two regions in heterodimerization of hMSH2 with MSH3 and MSH6 (grey) [22]. Also indicated are the Walker A and B motifs required for ATP binding and hydrolysis [23]. The numbers above the boxes indicate amino acids delineating the functional regions; those below correspond to the missense mutations. (b) Alignment of human, mouse, rat and yeast MSH2. The numbers above the alignment correspond to hMSH2 amino-acid positions; those below correspond to yMSH2 residues. The missense mutations are indicated below the alignment.

***MSH2* missense mutations examined in this study.**

Mutation in hMSH2	Homologous mutation in yMSH2	Allele frequency	Role in HNPCC*	Role in sporadic or early onset colon cancer*
R524P	R542P	ND	+†	–
P622L	P640L	ND	+ [7]	–
C697F	C716F	ND	+ [8]	–
D506Y	D524Y	ND	–	+ [9]
G322D	G317D	1–6% [8,10–14]	+ [17]	+ [15,16]
Q419K	Q430K	1% [18]	–	–
L390F	L402F	2% [18]	+ [19]	–

The mutations are indicated in the single-letter amino-acid code. *These columns indicate whether a particular polymorphism has (+) or has not (–) been implicated in the development of HNPCC, or of sporadic or early onset colon cancer. †Although described as HNPCC [4], this patient did not fulfill the Amsterdam criteria and had ovarian cancer. ND, not determined.

Problem Set

Creating mutagenic primers for site-directed mutagenesis in yeast VMA6

Here's the nucleotide sequence of the yeast VMA6 gene:

```
   1 ATGGAAGGCG TGTATTTCAA TATTGACAAT GGGTTTATTG AAGGTGTAGT GAGAGGCTAC
  61 AGAAATGGGT TGTTATCTAA TAACCAATAC ATCAACTTAA CACAATGTGA CACGTTGGAA
 121 GATCTAAAAT TACAATTATC ATCAACTGAT TATGGTAATT TTCTTTCCTC TGTTTCCTCA
 181 GAGTCTTTGA CCACGTCATT GATTCAAGAA TATGCTTCTA GCAAGTTGTA CCACGAATTC
 241 AACTACATAA GAGACCAATC CAGTGGATCC ACGAGAAAGT TCATGGACTA TATCACTTAT
 301 GGTTACATGA TCGACAATGT AGCATTGATG ATTACAGGTA CTATTCATGA TCGTGATAAG
 361 GGTGAAATTT TACAACGTTG TCATCCGCTA GGTTGGTTTG ATACTTTGCC TACGTTGAGT
 421 GTTGCTACTG ATCTTGAATC CCTATACGAA ACCGTATTGG TGGATACCCC ACTGGCACCT
 481 TACTTCAAAA ACTGTTTTGA CACGGCAGAG GAGCTAGACG ATATGAACAT TGAAATTATT
 541 AGAAATAAGC TGTACAAGGC TTATTTAGAA GACTTTTACA ATTTTGTCAC TGAAGAAATT
 601 CCGGAACCTG CTAAAGAATG TATGCAAACA TTACTAGGGT TTGAAGCTGA CAGAAGAAGT
 661 ATCAATATTG CACTCAACTC TTTGCAAAGT TCAGATATTG ACCCAGATTT GAAAAGTGAC
 721 TTGTTACCTA ACATAGGTAA GTTGTACCCT CTTGCAACGT TTCACTTGGC GCAAGCCCAA
 781 GATTTCGAAG GAGTTAGAGC TGCTTTAGCT AACGTCTATG AGTATAGGGG ATTTTTGGAG
 841 ACTGGTAACT TAGAAGATCA CTTTTACCAA TTGGAAATGG AACTATGTAG AGATGCTTTC
 901 ACGCAACAAT TTGCCATCAG CACTGTTTGG GCCTGGATGA AATCCAAGGA ACAAGAAGTT
 961 AGGAATATTA CCTGGATTGC AGAATGTATC GCACAAAACC AAAGAGAAAG AATCAACAAT
1021 TATATTTCCG TTTATTGA
```

Using SIXFRAME, the correct open reading frame (ORF 1) can be identified:

```
M  E  G  V  Y  F  N  I  D  N  G  F  I  E  G  V  V  R  G  Y
1    atggaaggcgtgtatttcaatattgacaatgggtttattgaaggtgtagtgagaggctac  60
     R  N  G  L  L  S  N  N  Q  Y  I  N  L  T  Q  C  D  T  L  E
61   agaaatgggttgttatctaataaccaatacatcaacttaacacaatgtgacacgttggaa  120
     D  L  K  L  Q  L  S  S  T  D  Y  G  N  F  L  S  S  V  S  S
121  gatctaaaattacaattatcatcaactgattatggtaattttctttcctctgtttcctca  180
     E  S  L  T  T  S  L  I  Q  E  Y  A  S  S  K  L  Y  H  E  F
181  gagtctttgaccacgtcattgattcaagaatatgcttctagcaagttgtaccacgaattc  240
     N  Y  I  R  D  Q  S  S  G  S  T  R  K  F  M  D  Y  I  T  Y
241  aactacataagagaccaatccagtggatccacgagaaagttcatggactatatcacttat  300
     G  Y  M  I  D  N  V  A  L  M  I  T  G  T  I  H  D  R  D  K
301  ggttacatgatcgacaatgtagcattgatgattacaggtactattcatgatcgtgataag  360
     G  E  I  L  Q  R  C  H  P  L  G  W  F  D  T  L  P  T  L  S
361  ggtgaaattttacaacgttgtcatccgctaggttggtttgatactttgcctacgttgagt  420
     V  A  T  D  L  E  S  L  Y  E  T  V  L  V  D  T  P  L  A  P
421  gttgctactgatcttgaatccctatacgaaaccgtattggtggataccccactggcacct  480
     Y  F  K  N  C  F  D  T  A  E  E  L  D  D  M  N  I  E  I  I
481  tacttcaaaaactgttttgacacggcagaggagctagacgatatgaacattgaaattatt  540
     R  N  K  L  Y  K  A  Y  L  E  D  F  Y  N  F  V  T  E  E  I
541  agaaataagctgtacaaggcttatttagaagacttttacaattttgtcactgaagaaatt  600
     P  E  P  A  K  E  C  M  Q  T  L  L  G  F  E  A  D  R  R  S
601  ccggaacctgctaaagaatgtatgcaaacattactagggtttgaagctgacagaagaagt  660
     I  N  I  A  L  N  S  L  Q  S  S  D  I  D  P  D  L  K  S  D
661  atcaatattgcactcaactctttgcaaagttcagatattgacccagatttgaaaagtgac  720
     L  L  P  N  I  G  K  L  Y  P  L  A  T  F  H  L  A  Q  A  Q
721  ttgttacctaacataggtaagttgtaccctcttgcaacgtttcacttggcgcaagcccaa  780
     D  F  E  G  V  R  A  A  L  A  N  V  Y  E  Y  R  G  F  L  E
781  gatttcgaaggagttagagctgctttagctaacgtctatgagtataggggatttttggag  840
     T  G  N  L  E  D  H  F  Y  Q  L  E  M  E  L  C  R  D  A  F
841  actggtaacttagaagatcacttttaccaattggaaatggaactatgtagagatgctttc  900
     T  Q  Q  F  A  I  S  T  V  W  A  W  M  K  S  K  E  Q  E  V
901  acgcaacaatttgccatcagcactgtttgggcctggatgaaatccaaggaacaagaagtt  960
     R  N  I  T  W  I  A  E  C  I  A  Q  N  Q  R  E  R  I  N  N
961  aggaatattacctggattgcagaatgtatcgcacaaaaccaaagagaaagaatcaacaat  1020
     Y  I  S  V  Y  *
1021 tatatttccgtttattga  1038
```

To simplify the analysis, a shorter gene fragment can be generated using the nucleic editing tool

```
     L  L  P  N  I  G  K  L  Y  P  L  A  T  F  H  L  A  Q  A  Q
1    ttgttacctaacataggtaagttgtaccctcttgcaacgtttcacttggcgcaagcccaa  60
     D  F  E  G  V  R  A  A  L  A  N  V  Y  E  Y  R  G  F  L  E
61   gatttcgaaggagttagagctgctttagctaacgtctatgagtataggggatttttggag  120
     T  G  N  L  E  D  H  F  Y  Q  L  E  M  E  L  C  R  D  A  F
121  actggtaacttagaagatcactttaccaattggaaatggaactatgtagagatgctttc  180
     T  Q  Q  F  A  I  S  T  V  W  A  W  M  K  S  K  E  Q  E  V
181  acgcaacaatttgccatcagcactgtttgggcctggatgaaatccaaggaacaagaagtt  240
     R  N  I  T  W  I  A  E  C  I  A  Q  N  Q  R  E  R  I  N  N
241  aggaatattacctggattgcagaatgtatcgcacaaaaccaaagagaaagaatcaacaat  300
     Y  I  S  V  Y  *
301  tatatttccgtttattga  318
```