CRISPR

The acronym “CRISPR” has often been invoked when considering recent revolutionary genetic engineering advances. This presentation will be devoted to the origin of the term, its biological significance and how further knowledge led to the expansion of the meaning of this term as well as its underlying processes in genetic engineering. Finally, the potential of CRISPR for deriving further molecular biological information and uses from experimentation and analysis will be discussed.

Dick Wagner

Lexington Computer and Technology Group

October 20, 2021

CRISPR

Clustered Regularly Interspaced
Short Palindromic Repeats

Chemical Bonds

Covalent -C-C-, -C-O-, -C=O, -C-N-C-

electrons are shared between atoms

Ionic NaCl

electrons partially transferred from one atom to another

Hydrogen -O--H-O-, -N--H-O-, -O--H-N-

electrostatic attraction between hydrogen and lone pair electrons of electronegative atom

Hydrophobic nonpolar molecules associate to exclude water

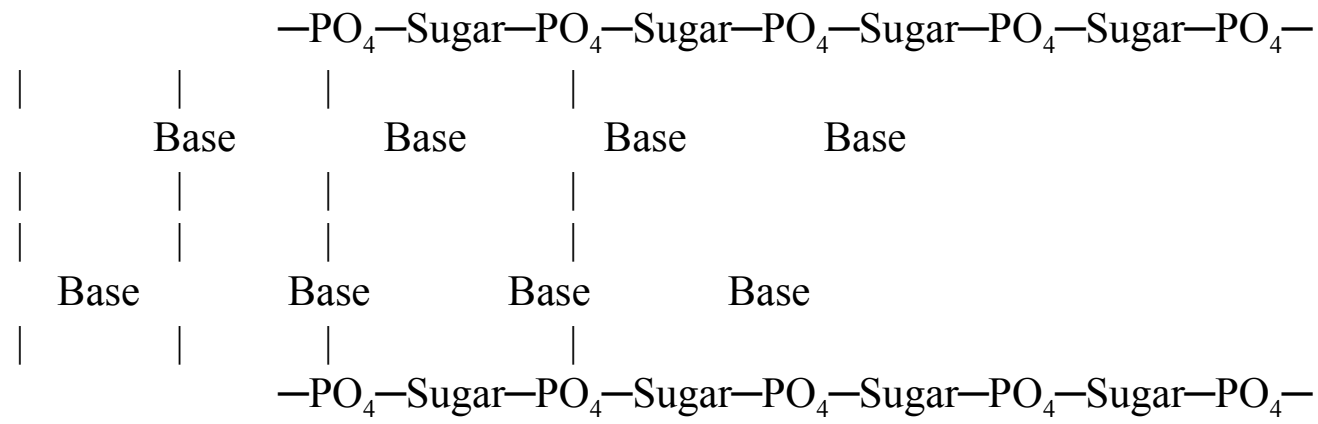
Central Dogma of Biology

DNA→RNA →Protein

DNA (DeoxyriboNucleic Acid)

Sugar (deoxyribose); Phosphate (acid portion); Base: Adenine (A), Thymine (T), Guanine (G), Cytosine (C)

Adenine always H-bonds to Thymine; Guanine always H-bonds to Cytosine



The phosphate-sugar-base configuration is referred to as a nucleotide. Complementary nucleotides in DNA are referred to as base pairs.

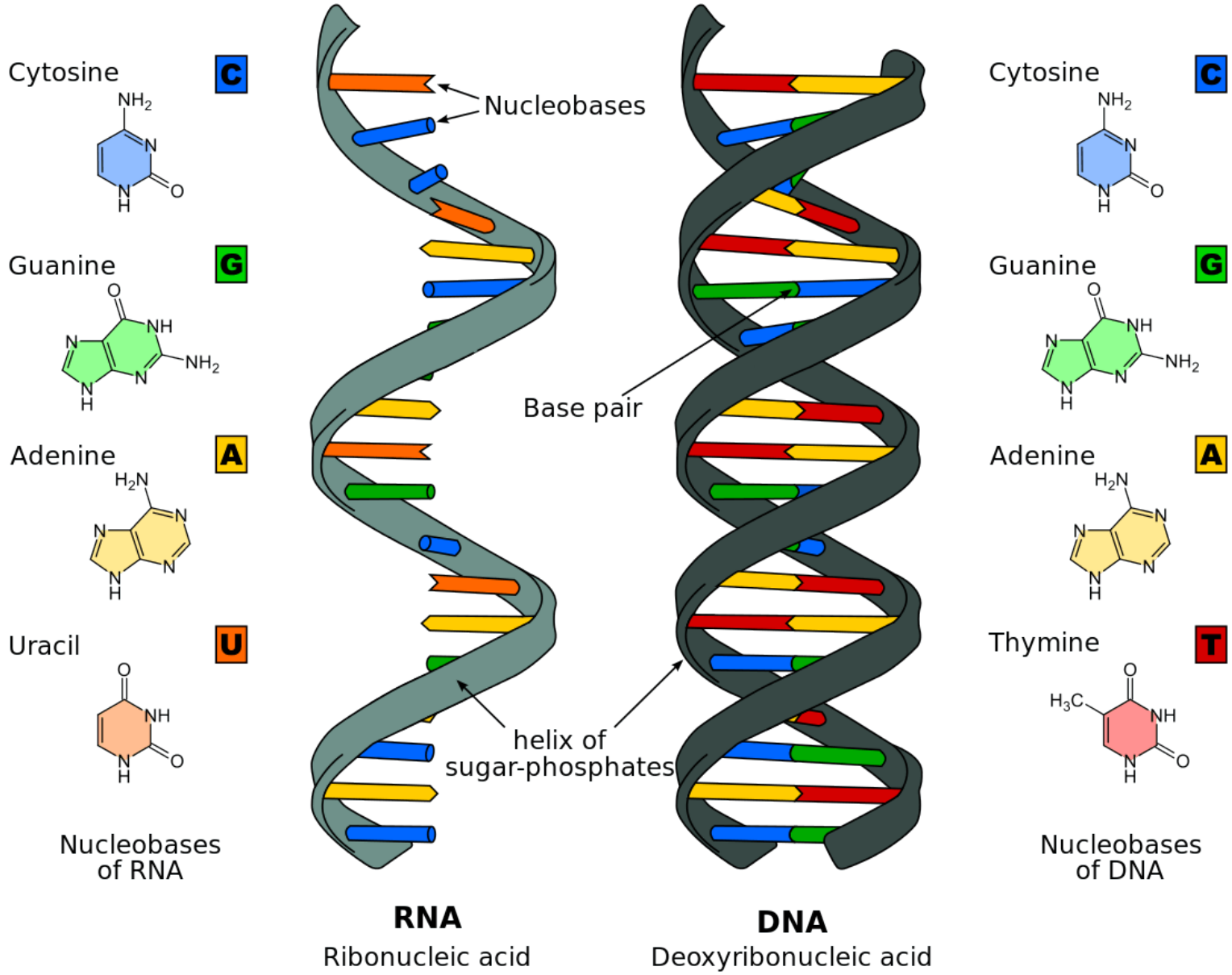
RNA (RiboNucleic Acid)

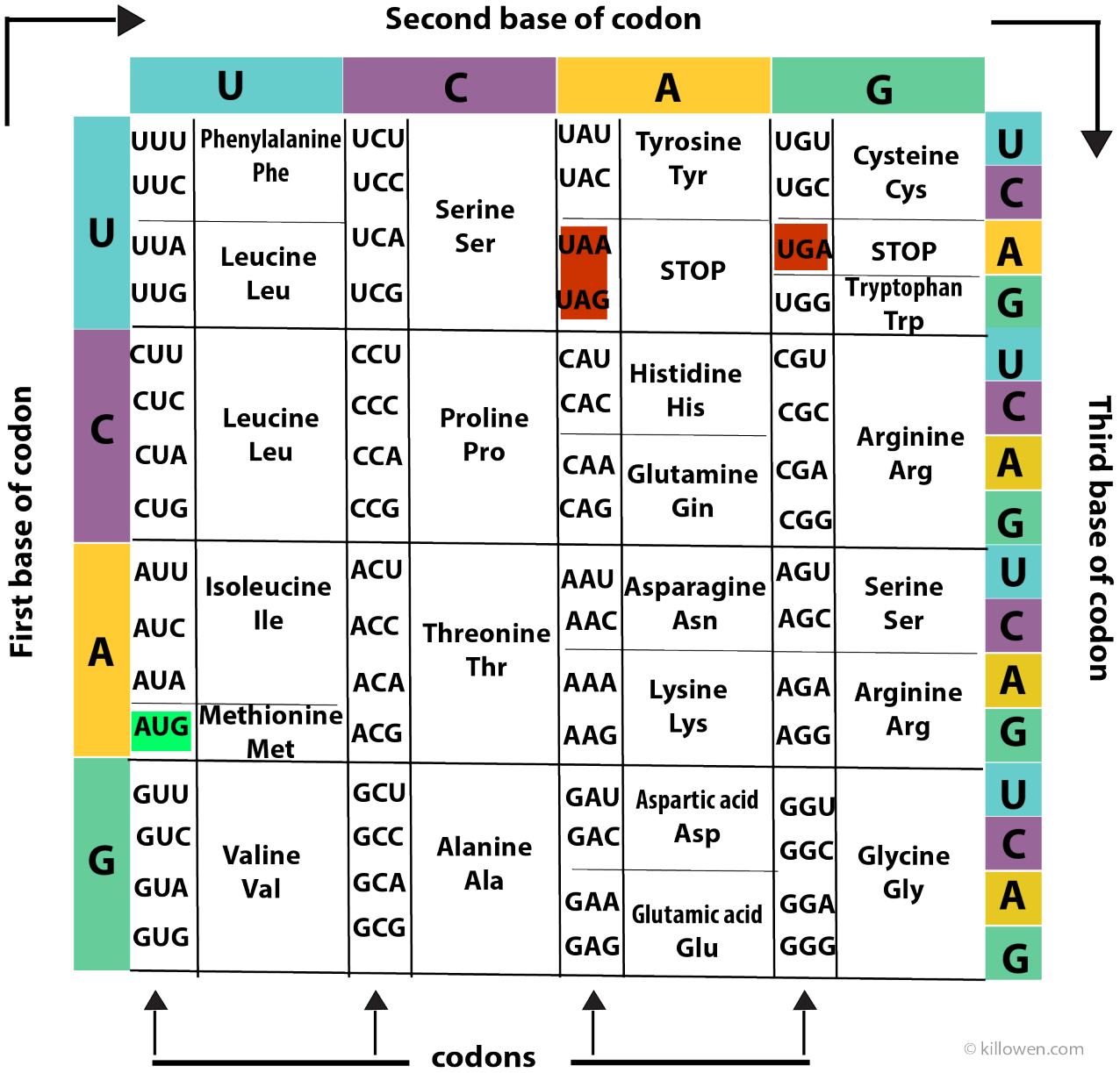
Sugar (ribose); Phosphate (acid portion); Base: Adenine (A), Uracil (U), Guanine (G), Cytosine (C)

Same Sugar, Phosphate, Base arrangement as for DNA , <u>BUT</u> no H-bond double strands when free in solution.

Protein

Amino Acids (21 common ones): aa_1─aa_2─aa_3─aa_4─ --- ─aa_n


Cytosine
C
Guanine
G
Adenine
A
Uracil
U
Nucleobases of RNA
Nucleobases
Base pair
helix of sugar-phosphates
RNA
Ribonucleic acid
DNA
Deoxyribonucleic acid
Cytosine
C
Guanine
G
Adenine
A
Thymine
T
Nucleobases of DNA



First base	U	C	A	G	Third base
U	UUU Phenylalanine Phe	UCU Serine Ser	UAU Tyrosine Tyr	UGU Cysteine Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leucine Leu	UCA	UAA STOP	UGA STOP	A
	UUG	UCG	UAG	UGG Tryptophan Trp	G
C	CUU Leucine Leu	CCU Proline Pro	CAU Histidine His	CGU Arginine Arg	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Glutamine Gin	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Isoleucine Ile	ACU Threonine Thr	AAU Asparagine Asn	AGU Serine Ser	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lysine Lys	AGA Arginine Arg	A
	AUG Methionine Met	ACG	AAG	AGG	G
G	GUU Valine Val	GCU Alanine Ala	GAU Aspartic acid Asp	GGU Glycine Gly	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glutamic acid Glu	GGA	A
	GUG	GCG	GAG	GGG	G

The DNA (genome):

---TGG-TTG-ACC-CAT---

is transcribed as RNA:

---UGG-UUG-ACC-CAU---

(transcription is from complementary DNA strand):

---ACC-AAC-UGG-GUA---

RNA translates to aa (protein):

---Trp-Leu-Thr-His---

Bacterial (& Archaeal) Immune Response to Phage

1. Phage attaches to cell membrane and inserts its genetic DNA into cell cytoplasm
2. Cell machinery uniquely inserts portions of this phage DNA into its cellular DNA

Subsequently, bacterial RNA contains (e.g.):

---spacer$_1$-**<u>CUUATG</u>**AAUCG**<u>CAUAAG</u>**-spacer$_2$-**<u>CUUATG</u>**AAUCG**<u>CAUAAG</u>**---

spacer$_n$-**<u>CUUATG</u>**AAUCG**<u>CAUAAG</u>**---

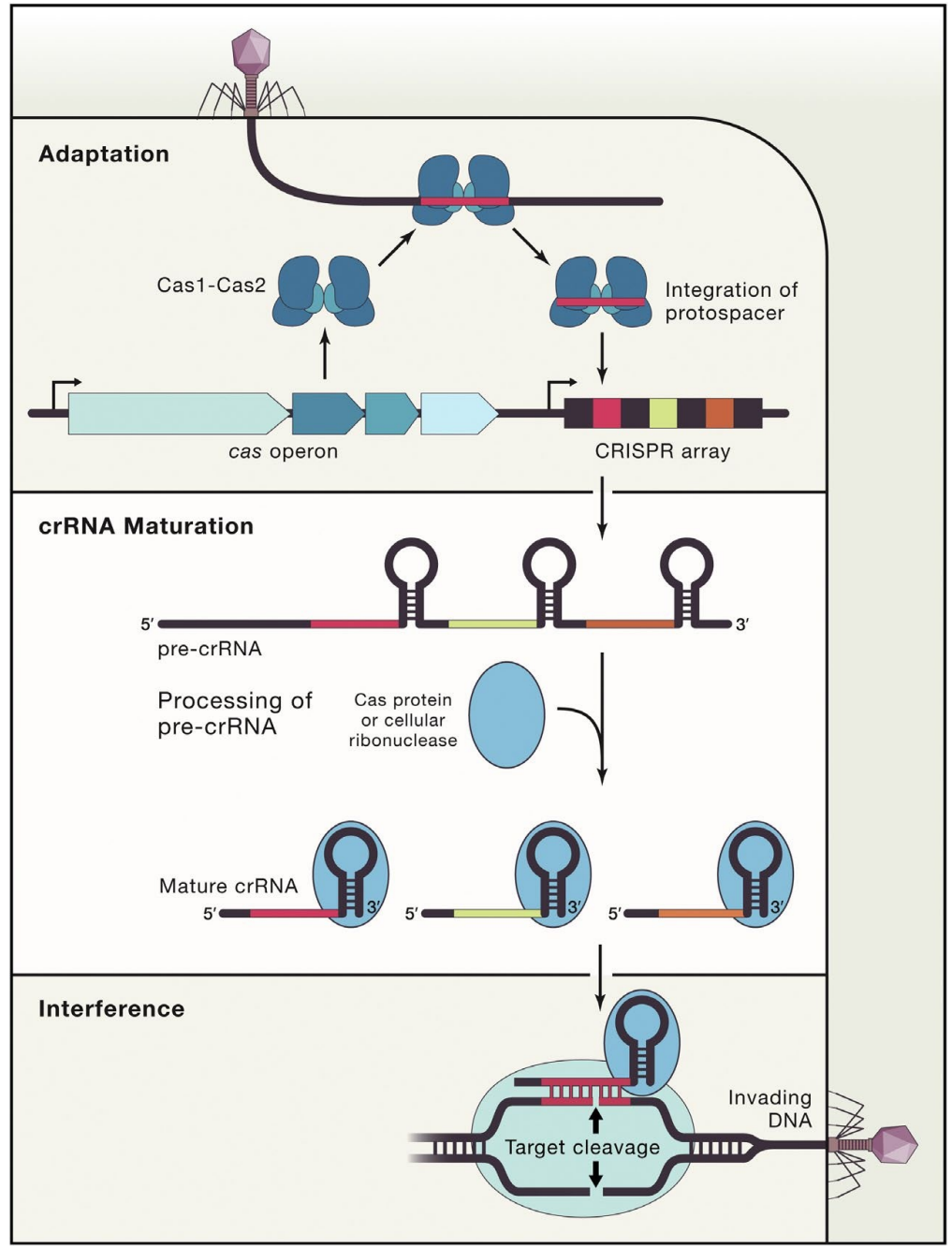
where the spacers are RNA portions from the phage DNA. The **<u>CUUATG</u>**AAUCG**<u>CAUAAG</u>** palindromic elements are often referred to as dyads or repeats of ~28-37 nt in length. The spacers are of constant length of ~25-35 nt.

This RNA is cleaved by a bacterial enzyme (an RNase or a Cas protein) to form:

-spacer$_x$-CUUATGAAUCGCAUAAG-

where the palindromic sequence repeats hybridize thereby forming loops.

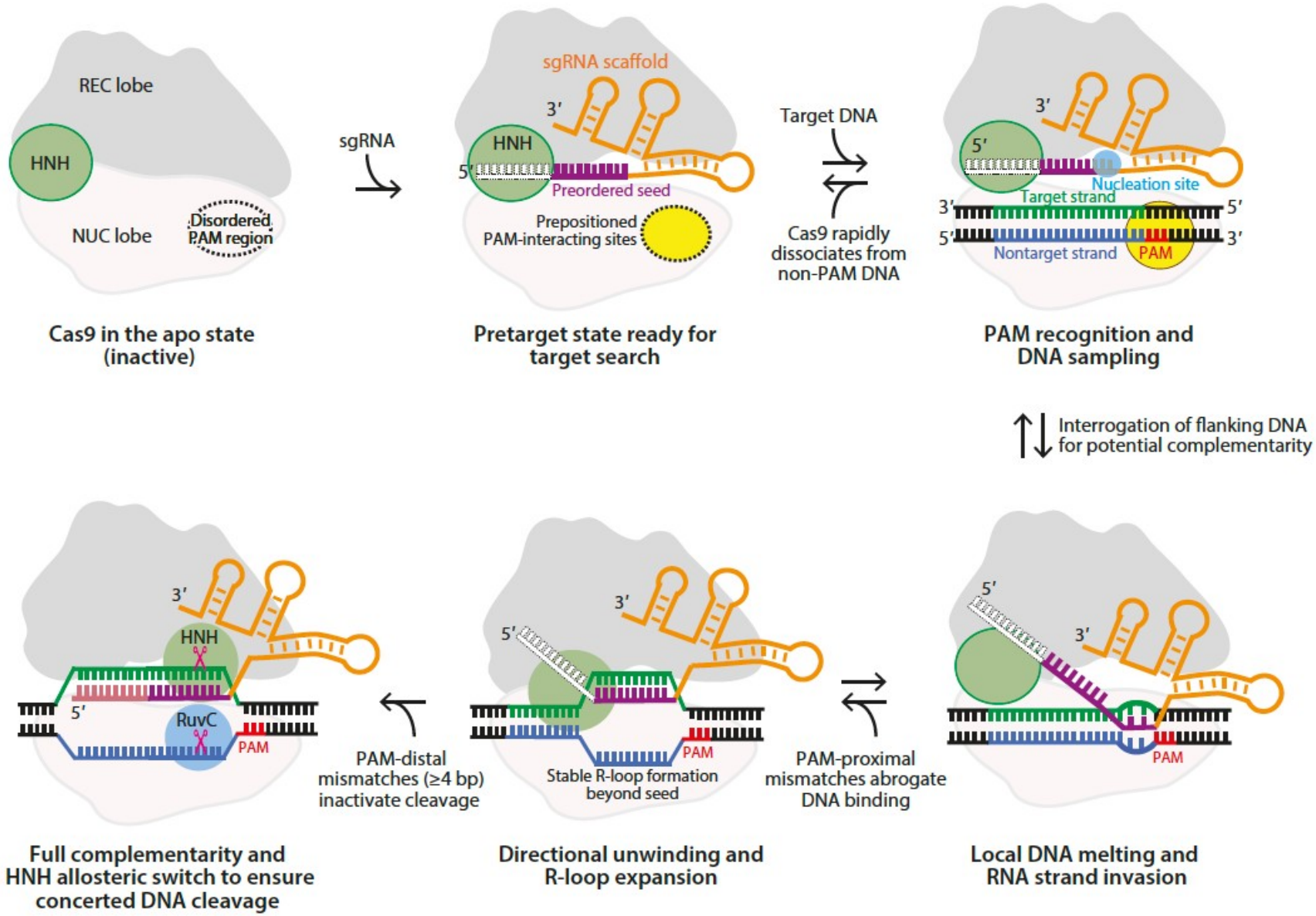
These spacer/palindromic nucleic acid repeats bind to the DNA of subsequent invasions by these phage via a Cas (CRISPR associated) group of proteins. The Cas proteins use the spacers to bind the phage DNA and then cleave the resultant RNA/DNA complexes, within the spacer regions, into pieces for eventual riddance by the bacteria.


Adaptation
Cas1-Cas2
Integration of protospacer
cas operon
CRISPR array
crRNA Maturation
5′
3′
pre-crRNA
Processing of pre-crRNA
Cas protein or cellular ribonuclease
Mature crRNA
5′
3′
5′
3′
5′
3′
Interference
Invading DNA
Target cleavage

Further Knowledge

---*tracrRNA*---Cas_1-Cas_2--Cas_m---$spacer_1$-repeat-$spacer_2$-repeat--$spacer_n$-repeat---

1. The Cas genes are located upstream and closely adjacent to the CRISPR loci.
2. *tracr*RNA (trans-activating crRNA), located upstream from the Cas genes, is formed from the opposite (complementary) DNA strand; this *tracr*RNA can hybridize with individual palindromic repeats. The *tracr*RNA is necessary and acts in conjunction with the CRISPR RNA (crRNA) to allow the Cas proteins to cleave the target DNA (or RNA). The *tracr*RNA and crRNA can be covalently combined and still retain their activity.
3. PAMs (protospacer adjacent motifs), a 2-6 nt sequence (e.g., NGG) located a couple nt downstream from the sequences complementary to the protospacer or target sequences. This short sequence is on the complementary non-target strand and is recognized by the Cas proteins in an initial reaction of the Cas system with the target DNA.
4. “Seed region”, a ~10 nt sequence located at the downstream end of the spacer sequence and found in *tracr*RNA. Necessary for target recognition by Cas-*tracr*RNA-crRNA complex.
5. Cleavage of DNA occurs in the region the spacer recognizes. The cleavages are 3 nt away from the PAM site.


REC lobe
HNH
NUC lobe
Disordered PAM region
sgRNA
sgRNA scaffold
3′
HNH
5′
Preordered seed
Prepositioned PAM-interacting sites
Target DNA
Cas9 rapidly dissociates from non-PAM DNA
3′
5′
Nucleation site
Target strand
3′
5′
5′
3′
Nontarget strand
PAM
Interrogation of flanking DNA for potential complementarity
3′
HNH
5′
RuvC
PAM
PAM-distal mismatches (≥4 bp) inactivate cleavage
5′
3′
PAM
Stable R-loop formation beyond seed
PAM-proximal mismatches abrogate DNA binding
5′
3′
PAM
Cas9 in the apo state (inactive)
Pretarget state ready for target search
PAM recognition and DNA sampling
Full complementarity and HNH allosteric switch to ensure concerted DNA cleavage
Directional unwinding and R-loop expansion
Local DNA melting and RNA strand invasion

Further Knowledge Regarding Genetic Engineering

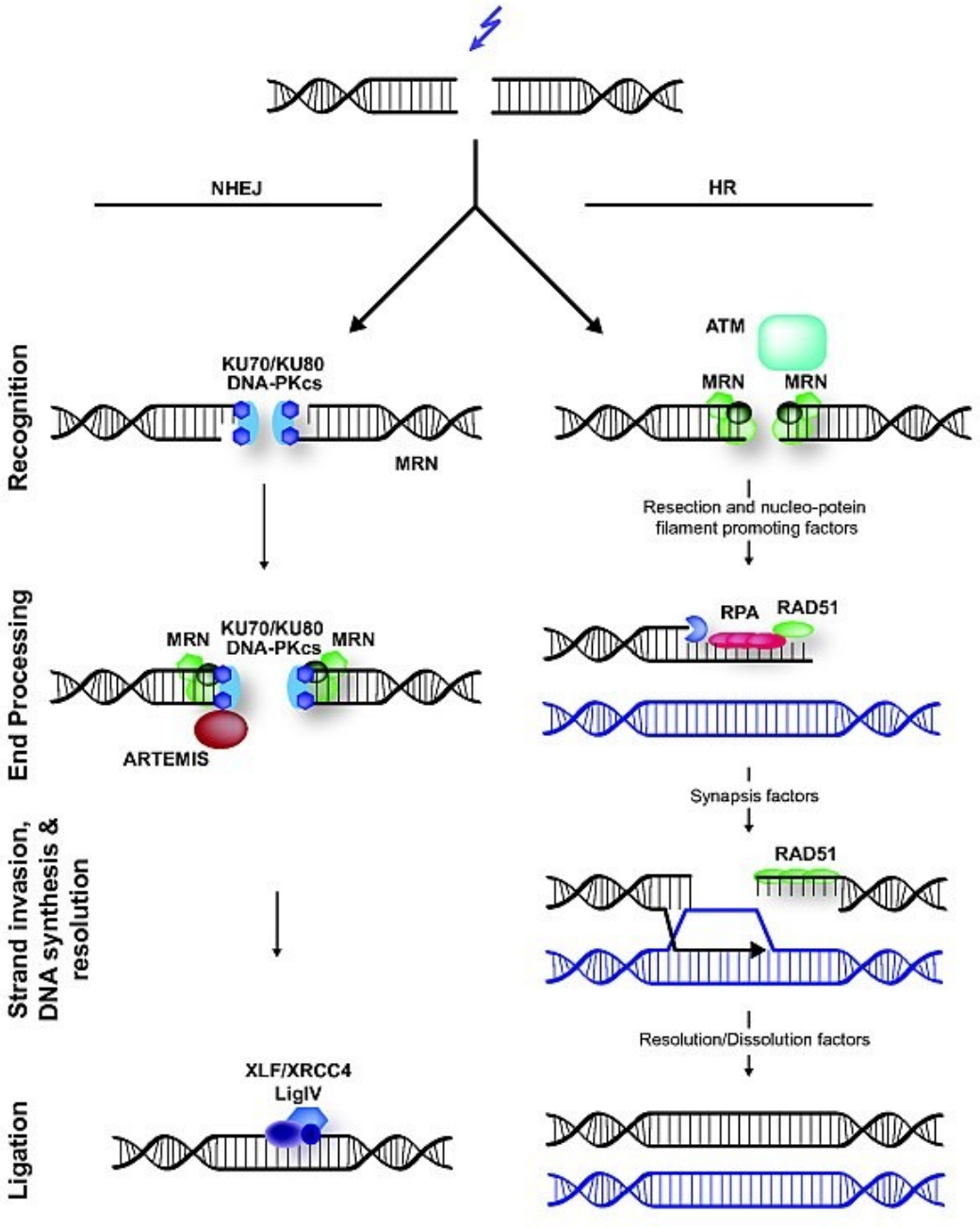
CRISPR Requirements:

A. One (or more) RNA(s):
   1) RNA containing CRISPR-like sequences (*tracr*RNA & crRNA or variation thereof)
   2) RNA containing 'seed' and remaining sequence that H-bonds to target DNA sequence

B. Cas system of proteins:
   1) Capable of incorporating 'signal' (sg) or 'guide' (g) RNA including:
      a. CRISPR-like sequences (*tracr*RNA & crRNA or variation thereof)
      b. 'seed' and further sequence complementary to target DNA sequence
   2) Recognizes PAM sequence in DNA strand complementary to target sequence
   3) Facility to uncoil target DNA when PAM sequence is recognized
   4) Allows 'seed' RNA sequence to interrogate and H-bond to target DNA sequence
   5) Allows the remaining RNA sequence to H-bond to target DNA sequence
   6} Causes cleavage of a covalent bond at the appropriate location in the target DNA
   7) Causes cleavage of a covalent bond at the appropriate location in the complementary DNA sequence

In addition to strand cleavage (breaking, scission), gene editing usually requires strand joining (repair or recombination). This can be done by homologous or non-homologous techniques (next slide).

Previous gene editing tools for localized strand scission were zinc finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN). For a variety of reasons, CRISPR appears to be better than these previous tools for gene editing because it uses PAM and 'seed', etc. sequence for its targeting, and contains its own nucleases.



Non-Homologous Repair (no homologous template needed):

1) DNA cleavage may leave overhangs on individual strand ends.
2) Enzymes are recruited to trim off damaged or mismatched ends.
3) Enzymes (ligases) join the ends. Strand gaps that may exist are also filled in.

Homologous Repair (Recombination):

1) Requires a homologous DNA sequence.
2) Strand ends are processed to remove unnecessary material and to form single strand overhangs.
3) Complementary DNA piece invades broken DNA to provide a template for a DNA polymerase.
4) This polymerase synthesizes missing portion of broken strand. The template/polymerase process can occur for either strand.
5) Repaired DNA and homologous DNA uncouple.

Further CRISPR Refinements

Cleavage of only one of the DNA strands (a nick).
Cleavage at different sites on the complementary DNA strands.
Cleavage of RNA.

Uses of the CRISPR Gene Editing Tool

Identify mutant genes in organisms (animals, plants, microbes) and edit them; i.e., remove or modify (e.g., "correct") them.
Identify mutant or investigational genes in humans and insert them at target locations in experimental systems; e.g., clone them.
Insert nucleotide sequences at specified locations in genomes for subsequent targeting; of themselves or of their transcripts.
Probe specific genome locations for their biological relevance; e.g., a gene, regulatory function site, genome ID site, etc.
Overall, use CRISPR technology for basic molecular biology research, to produce new biotech products, and to diagnose and/or treat diseases.