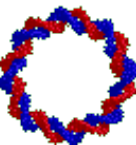


Informationsgehalt von DNA



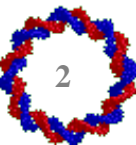
Welche Themen werden behandelt?

Gene

- Code, Genorganisation
- Signale in DNA
- Detektion von Genen

Genome

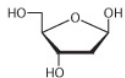
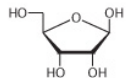
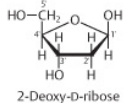
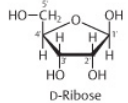
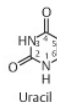
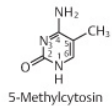
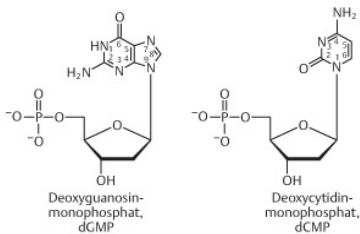
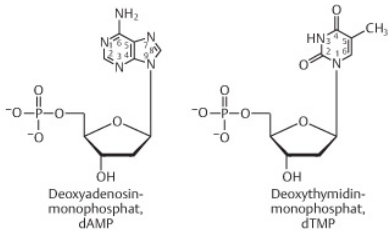
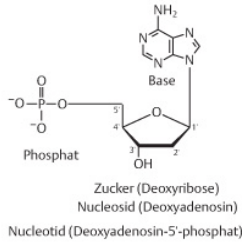
- Genomorganisation
- Nukleotidmuster
- Junk DNA



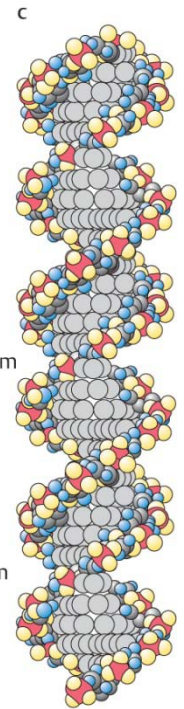
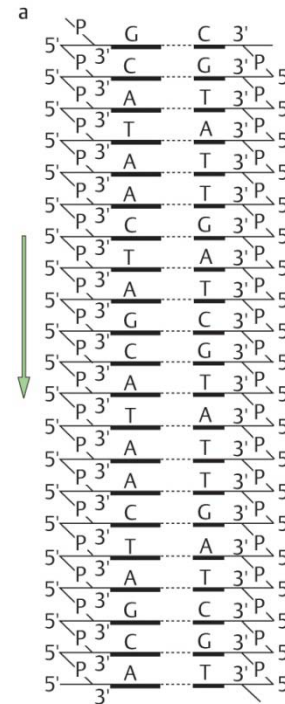
DNA als Informationsträger



DNA Bausteine



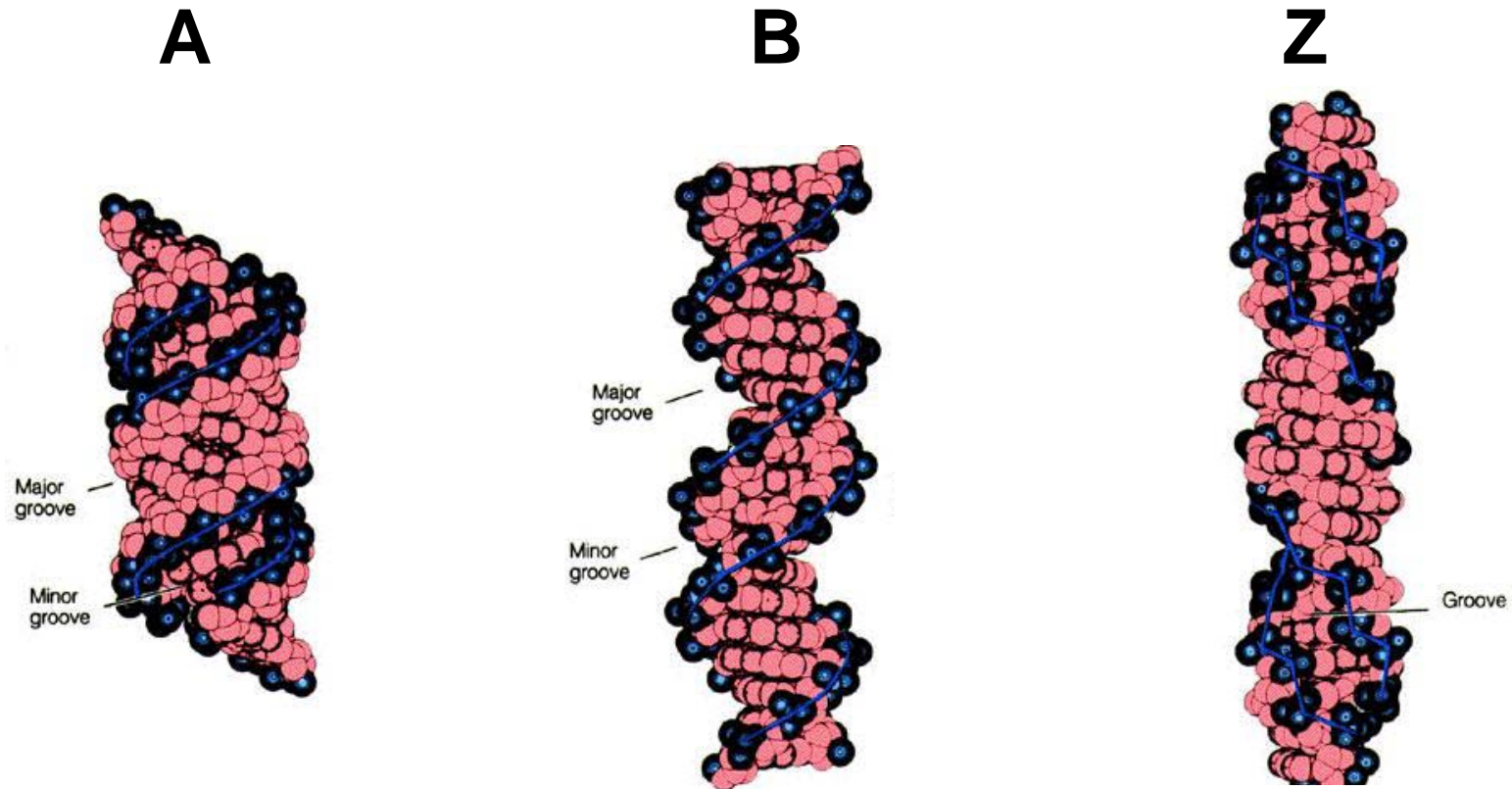
Desoxyribose
Phosphate
Base



● H ● O ● C in der Phosphodiester-Kette ● C bzw. N in den Basen ● P

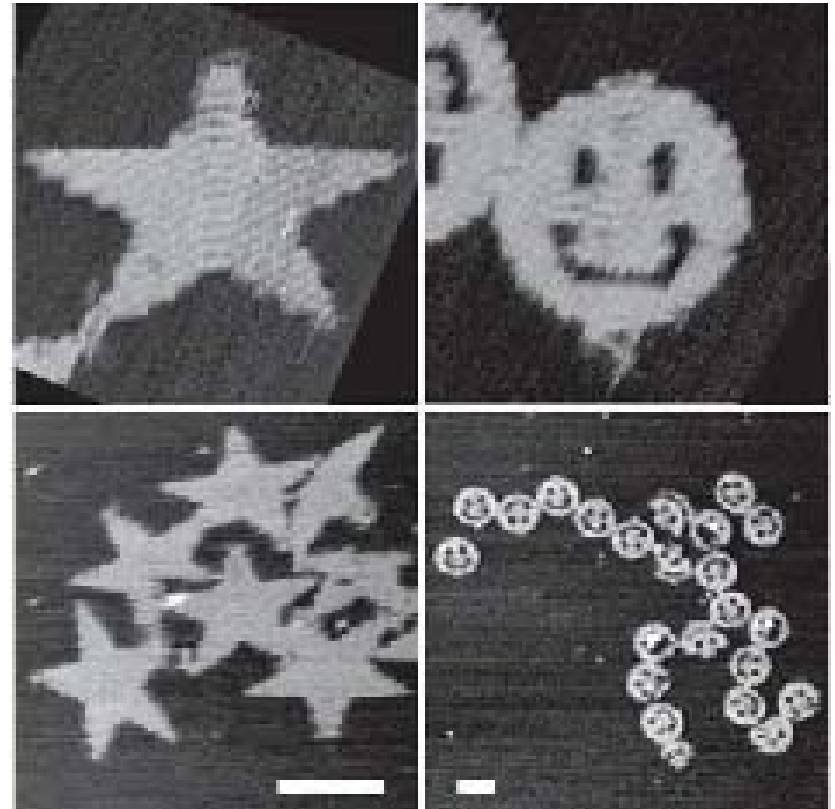
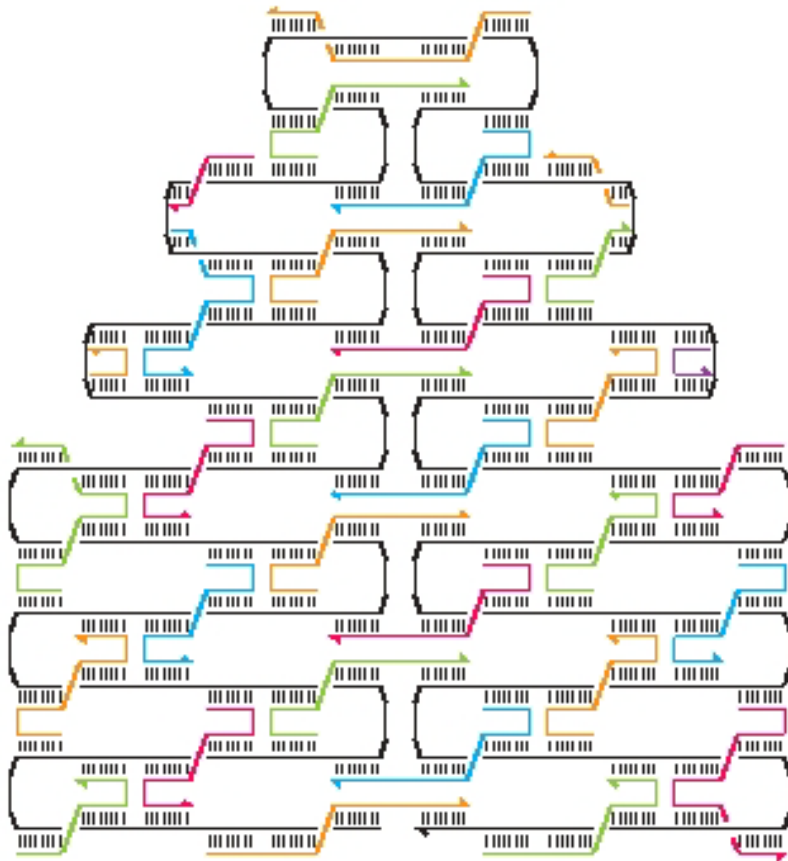
DNA Konformationen

Neben der Normalform (B) gibt es noch weitere Konformationen

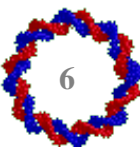


A und Z Formen könnten möglicherweise regulatorische Funktionen besitzen

DNA-Spielereien



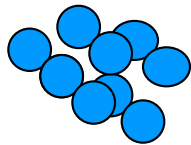
Mische M13 mit den richtigen Oligos! (aus Nature 440, 297-302, 2006)



DNA ist das genetische Material

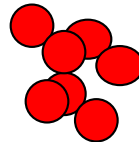
1920-er Jahre

harmlose
Bakterien

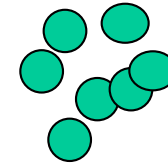


+

tote infektiöse
Bakterien

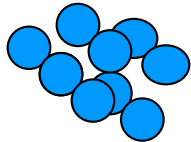


lebende infektiöse
Bakterien



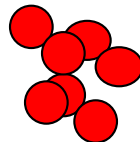
1944 (Avery/MacLeod/McCarthy):

harmlose
Bakterien

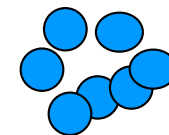


+

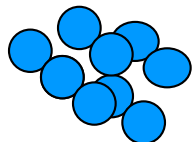
tote infektiöse
Bakterien + **DNAse**



harmlose
Bakterien

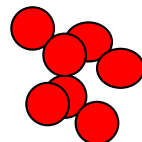


harmlose
Bakterien

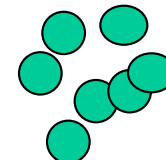


+

tote infektiöse
Bakterien + **Proteinase**



lebende infektiöse
Bakterien



DNA als Informationsquelle

- Information liegt in **linearer** Form vor
- Information ist **veschlüsselt**
- Information muss **übersetzt** werden
- Information kann **kopiert** werden
- Information kann **verändert** werden



Codierung

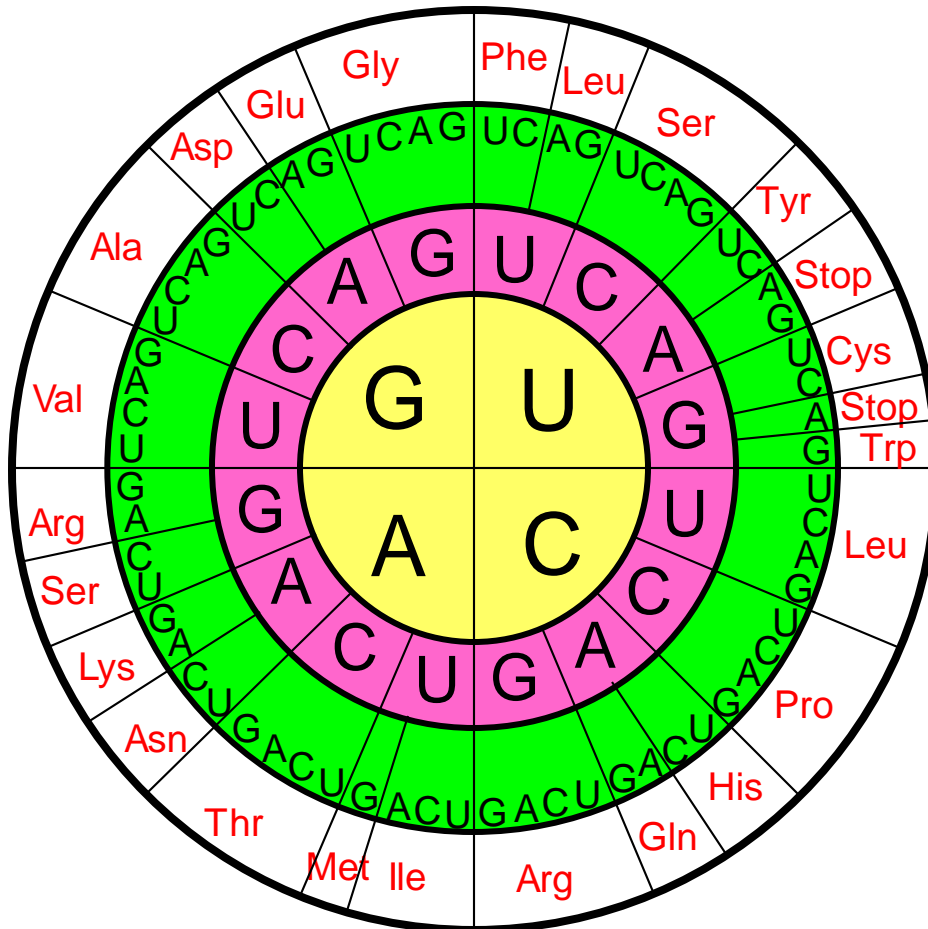
Vier Merkmale

- **Triplett-Code**
- **kommafrei**
- **degeneriert (wobble code)**
- **universell**



Wobble Code

Redundanz auf DNA-Ebene



1 Startcodon
3 Stopcodons
Bis zu 6 Codons/Aminosäure

Bakterien:

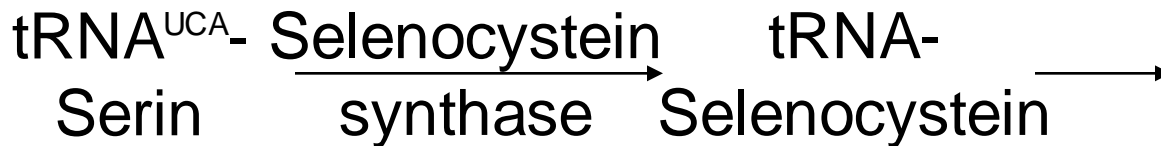
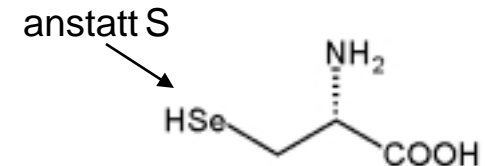
nicht alle Anticodons sind
im Genom vertreten

Zusätzliche Aminosäuren

Redundanz auf Protein-Ebene

UGA kann für **Selenocystein** kodieren

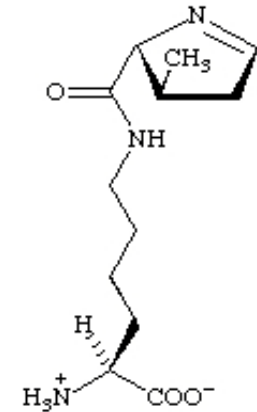
nur in 20 % (Bacteria) und 14 % (Archaea)
aller vollständig bekannten Genome



zum Ribosom
über „selB translation
elongation factor“

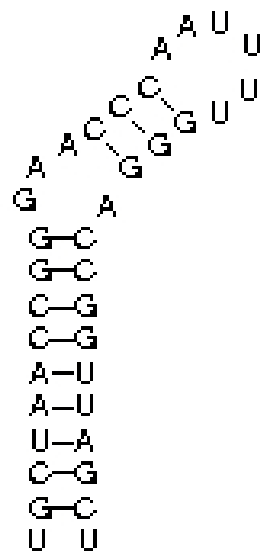
UAG kann für **Pyrrolysin** kodieren

von Lysin abgeleitet,
benutzt im Methan Metabolismus (Archaea)

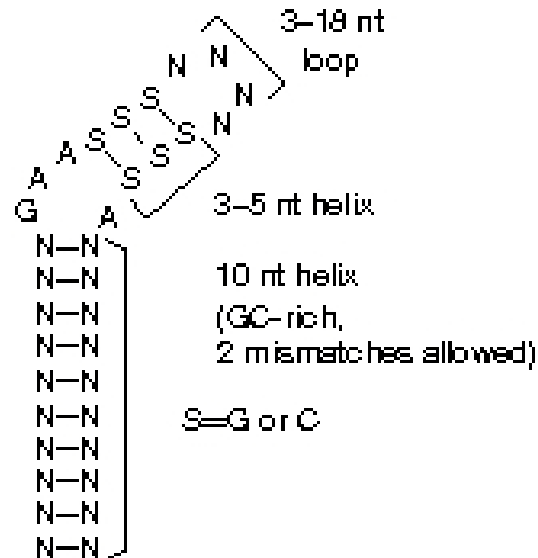


Strukturelle Voraussetzungen

Zusätzliche Erkennungsmerkmale sind nötig!



Methanococcus jannaschii
HeS-like SECIS element



Archaeal SECIS elements
consensus

From: EMBO Reports 5 (5):538-543 (2004)

SECIS Elemente direkt am Stop Codon (Bacteria)
oder in der 3' untranslatierten Region (Eukarya und Archaea)

Gen - Definition

Ein Gen ist:

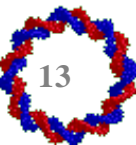
Eine lokalisierbare Region einer genomischen Sequenz, die einer vererbaren Einheit entspricht.

Es besteht aus mehreren Teilen:

regulatorische Regionen

transkribierte Regionen

Protein-codierende Gene sind eine Untergruppe!



Was ist ein Protein-codierendes Gen?

Die transkribierte Region besteht aus:

codierender Leserahmen

vom Start zum Stop-Codon

5' untranslatierte Region (5' UTR)

von Transkriptionsstart bis ATG

3' untranslatierte Region (3' UTR)

vom Stop zum Terminations/polyA Signal

Introns

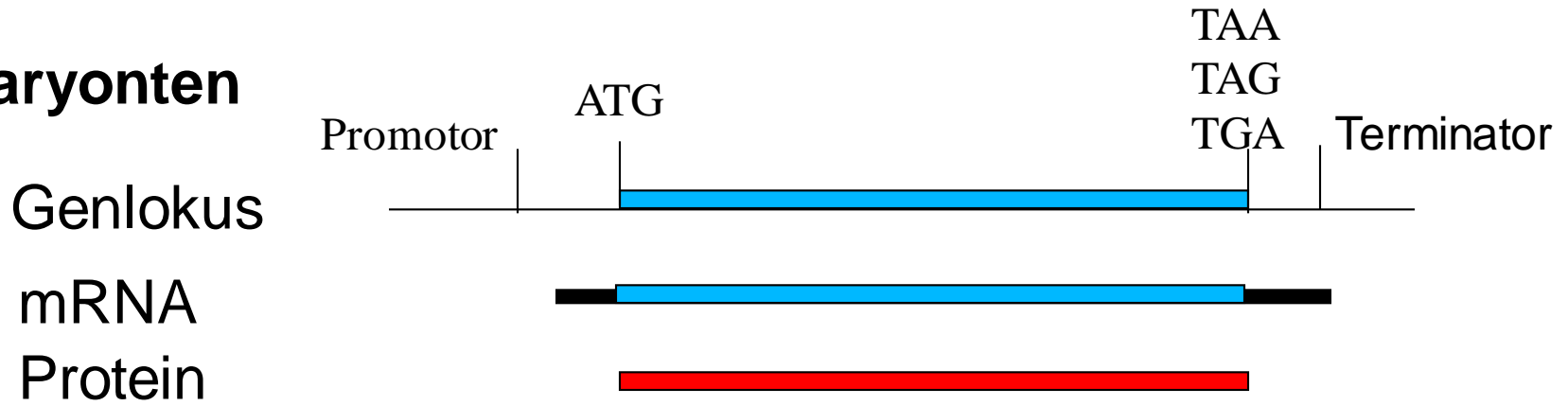
vom Donor zum Akzeptor

Beachte: Trans-Splicing führt dazu, dass verschiedene, weit von einander entfernte Regionen ein Gen bilden können!

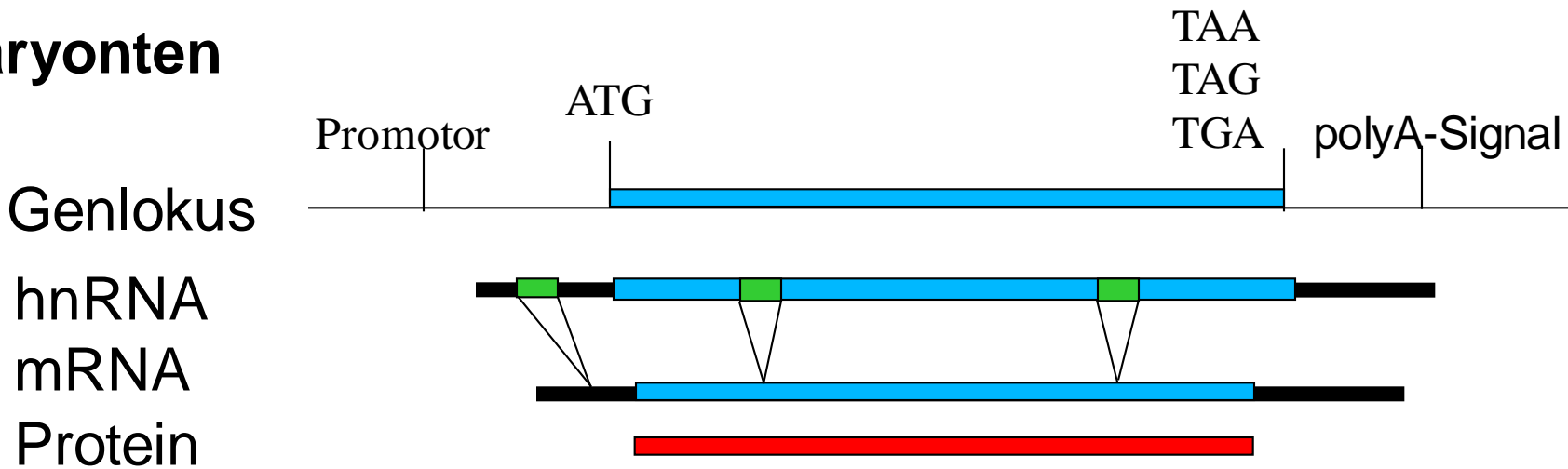


Organisation von Protein-kodierenden Genen

Prokaryonten



Eukaryonten



■ = Intron

Signale in DNA



Die Sinnsuche in DNA

Signal-Detektion

Probleme:

- schwache Signale

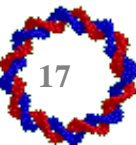
konservierte Motive sind sehr kurz

(z.B. Splice-Donoren und Akzeptoren)

Konservierung ist nicht zu 100 % gegeben

- spezies-spezifische Signale

Strukturkomponenten der Signale sind schwer zu berechnen



Gen- Signale in DNA

CpG Inseln
Methylierungsmuster
Promotoren

Polymerasebindungsstelle
Kofaktorbindestelle
Enhancer
Silencer

Transkriptionsstart
Introns

Donoren
Akzeptoren
Branch point
Splice Enhancers/Silencers

Polyadenylierungssignal
Terminationssignal

Unterschied Sequenz – Signal - Motiv

Sequenz

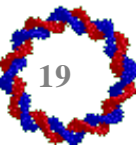
Eine eindeutige Anzahl von Basen.
Position und Reihenfolge stehen fest.

Signal

Z.B. positionsabhängige Häufigkeitsverteilungen
von Basen, wenn mehrere Sequenzen betrachtet werden.

Motiv (kann auch aus mehreren Signalen bestehen)

Regional begrenzte Basenabfolge unbestimmter Länge,
aber mit erkennbarem Muster.



Wie können Signale beschrieben werden?

Das Motiv:

ATN₁₋₃G₂CGTN_xTGA_{4,5}

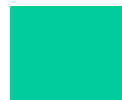
kann in viele mögliche Sequenzen übersetzt werden:

ATGGGCGTAGAGGAGACTTTATGAAAAA

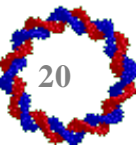
ATCGGCGTTGAAAA

ATAAGGCGTGAGTGAAAAA

etc.



= stabile Motivanteile



IUPAC Beschreibung von Signalen

a	a; adenine
c	c; cytosine
g	g; guanine
t	t; thymine in DNA; uracil in RNA
m	a or c
r	a or g
w	a or t
s	c or g
y	c or t
k	g or t
v	a or c or g; not t
h	a or c or t; not g
d	a or g or t; not c
b	c or g or t; not a
n	a or c or g or t

Beispiel eines Motivs in IUPAC Code

Beispielmotiv mit IUPAC wobble

NNGTAWBTSRWM

AA

CC C

GG AG CAAA

TTGTATTTGGTC

Detektion eines Motivs

NNGTAWBTSRWM

„Positional Weight Matrix“ Berechnung

- Erstellung einer ‚position frequency matrix‘ PFM

Normalisierung ergibt Wahrscheinlichkeiten für eine Base an einer Position

- normalisierte PFM

- Berechnung der Wahrscheinlichkeit einer Sequenz,
ein Profil widerzuspiegeln

Das Produkt der Wahrscheinlichkeit jeder Base an der entsprechenden Position

- Für die Computeranalyse:

Umrechnung in Log Werte mit Korrekturfaktor

(Korrekturfaktor sehr unterschiedlich, z.B. Quadratwurzel der Anzahl an Positionen)

+Korrektur für die Nukleotidfrequenzen

=PWM

‚Score values‘: Summierung aller PWM Werte über die
Länge des Profils



Positional Weight Matrix (PWM)

```
# scale = ln (frequency)
# frequency counts are based on
N(seq) = 21
```

#	A	C	G	T
1	1.50	1.25	2.25	1.50
2	1.70	0.92	2.01	1.87
3	2.01	0.92	1.50	2.01
4	-0.69	-0.69	-0.69	2.80
5	2.35	-0.69	1.87	-0.69
6	0.41	0.41	2.67	-0.69
7	-0.69	2.44	-0.69	1.70
8	0.41	0.41	1.50	2.35
9	0.41	2.25	0.41	1.70
10	2.80	-0.69	-0.69	-0.69
11	0.41	-0.69	2.67	0.41
12	0.92	0.41	-0.69	2.60
13	0.41	2.01	-0.69	2.14
14	-0.69	-0.69	2.80	-0.69
15	-0.69	-0.69	2.80	-0.69
16	0.92	0.92	-0.69	2.53
17	2.53	0.92	0.41	1.70
18	1.87	0.92	2.25	1.25
19	2.35	0.92	1.25	1.70

Je kleiner die Differenz bei den Werten
desto kleiner der Informationsgehalt

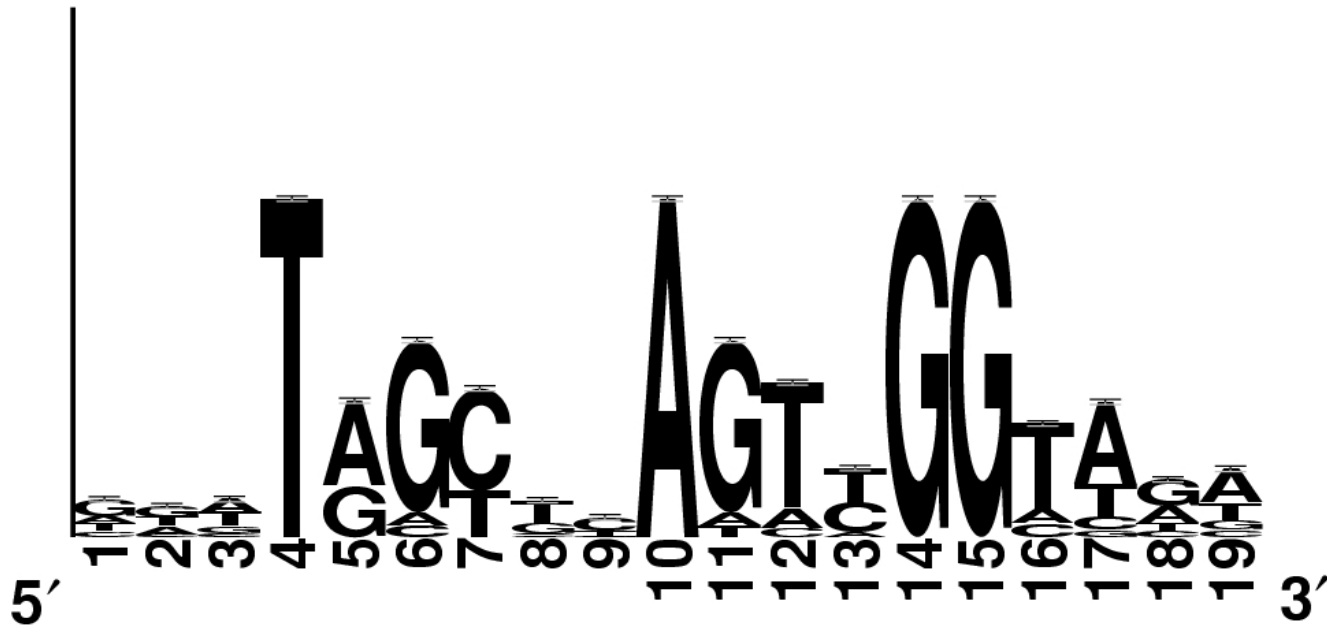
Je größer die Differenz bei den Werten
desto größer der Informationsgehalt

Hergestellt aus einem Alignment von
21 Sequenzen des A-Box Motivs
in *D. discoideum*



PWM Graphics

A-Box Motif from tRNAs in *D. discoideum*



Wie suche ich Sequenz-Motive?

...CACGGTAAATTATTA...

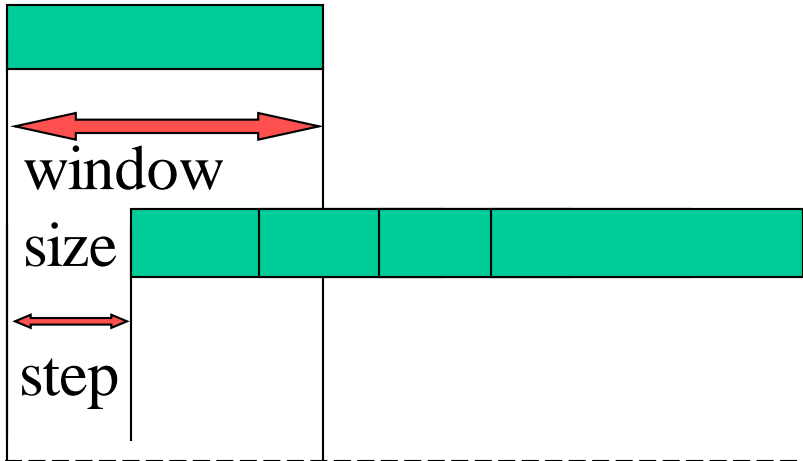


↓
Bewertungsmaßstab ist die Häufigkeit eines Nukleotides an einer bestimmten Stelle

Sliding Windows

Sequenz

Fenster

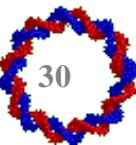


55 58 59 54 56

Ergebnis (GC%)



Signale in DNA: Introns



Introns

Selbstsplicende Introns in Prokaryonten bzw. Organellen

Zwei verschiedene Klassen in Eukaryonten

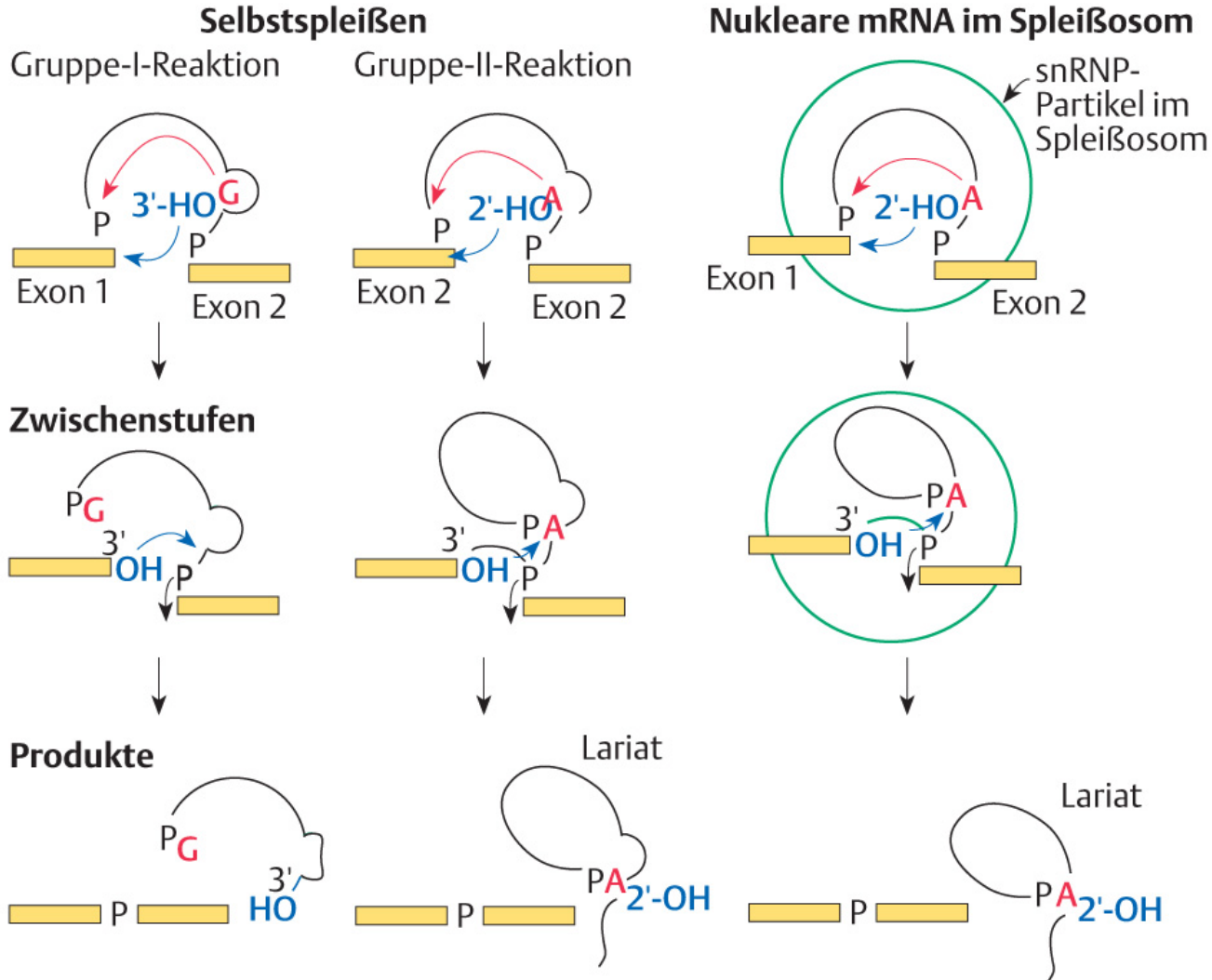
kurze Introns ~50 bis ~500 bp
in allen Eukaryonten vorhanden
(Ausnahmen: sehr kompakte Genome)
durchschnittlich ~100 bp, ‚klare‘ Signale

lange Introns bis zu mehrere 100 kb
nicht in Eukaryonten mit kleinen Genomen

Es besteht eine direkte Korrelation von
Intron-Größe mit Genomgröße!



Intronarten



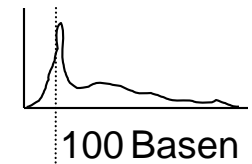
Introns II

Zwei verschiedene Spliceosomvarianten in Eukaryonten

U2 - abhängig die meisten Introns

viele kurze Introns

‚short intron peak‘



U12 - abhängig nur wenige Introns

gleichmäßige

Längenverteilung



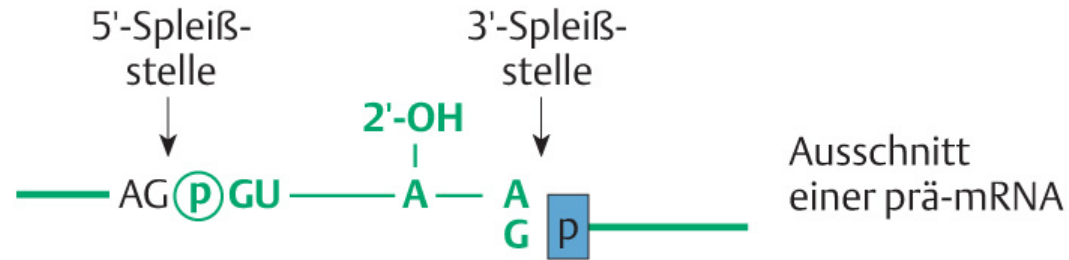
Verschiedene Spliceosomen

(Benennung nach dem ‚small nuclear Ribonucleoprotein‘, snRNP)

U5 ist die einzige gemeinsame snRNA,

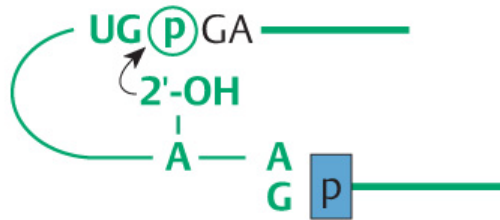
aber es gibt einige gemeinsame Proteinfaktoren

Splicingmechanismus



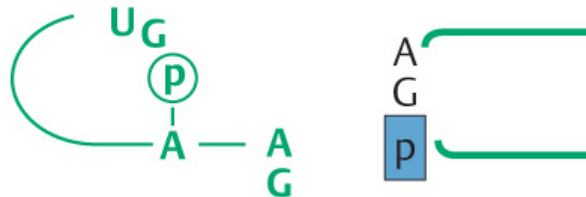
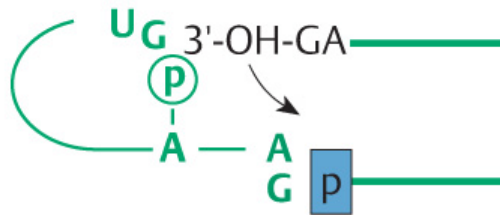
erster Schritt:

Öffnung der 5'-Spleißstelle
und Bildung der
Verzweigung

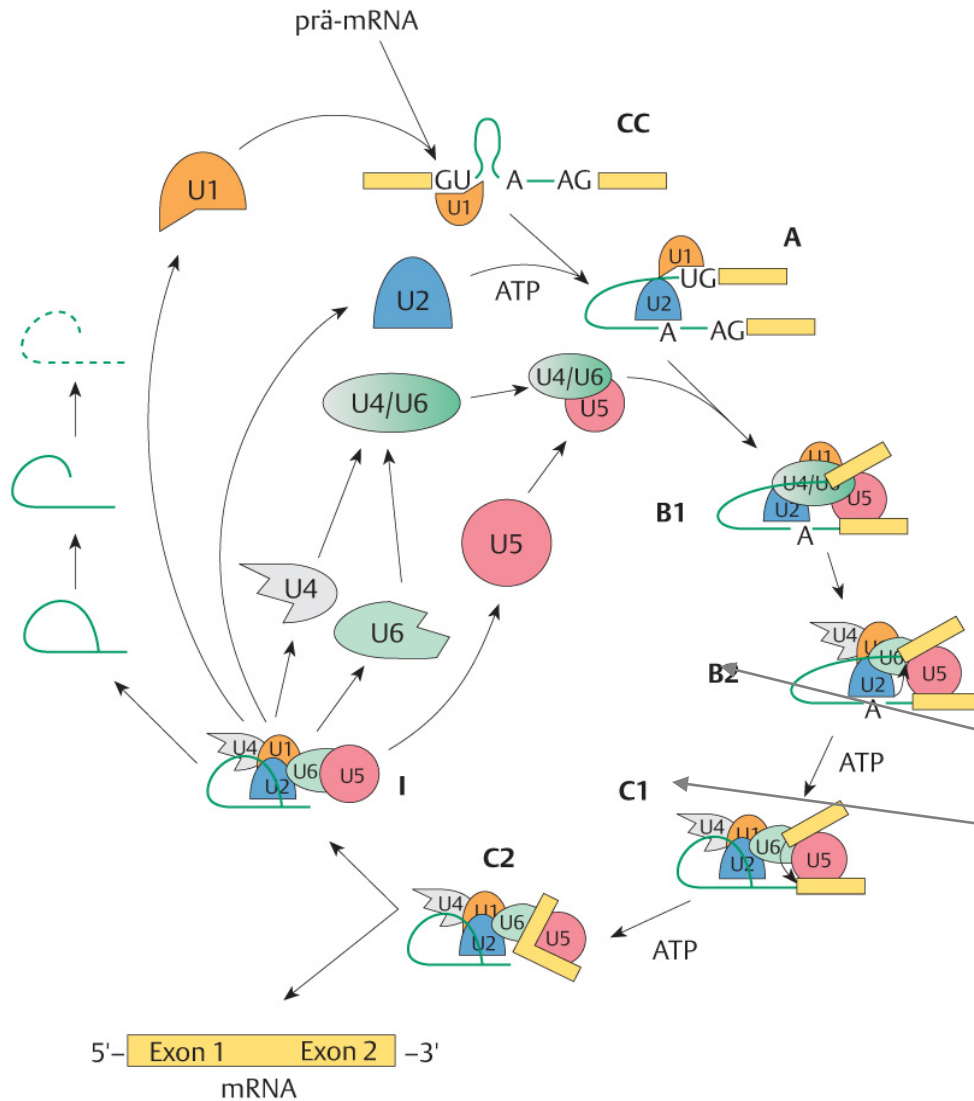


zweiter Schritt:

Öffnung der 3'-Spleißstelle
und Verknüpfen
der Exons



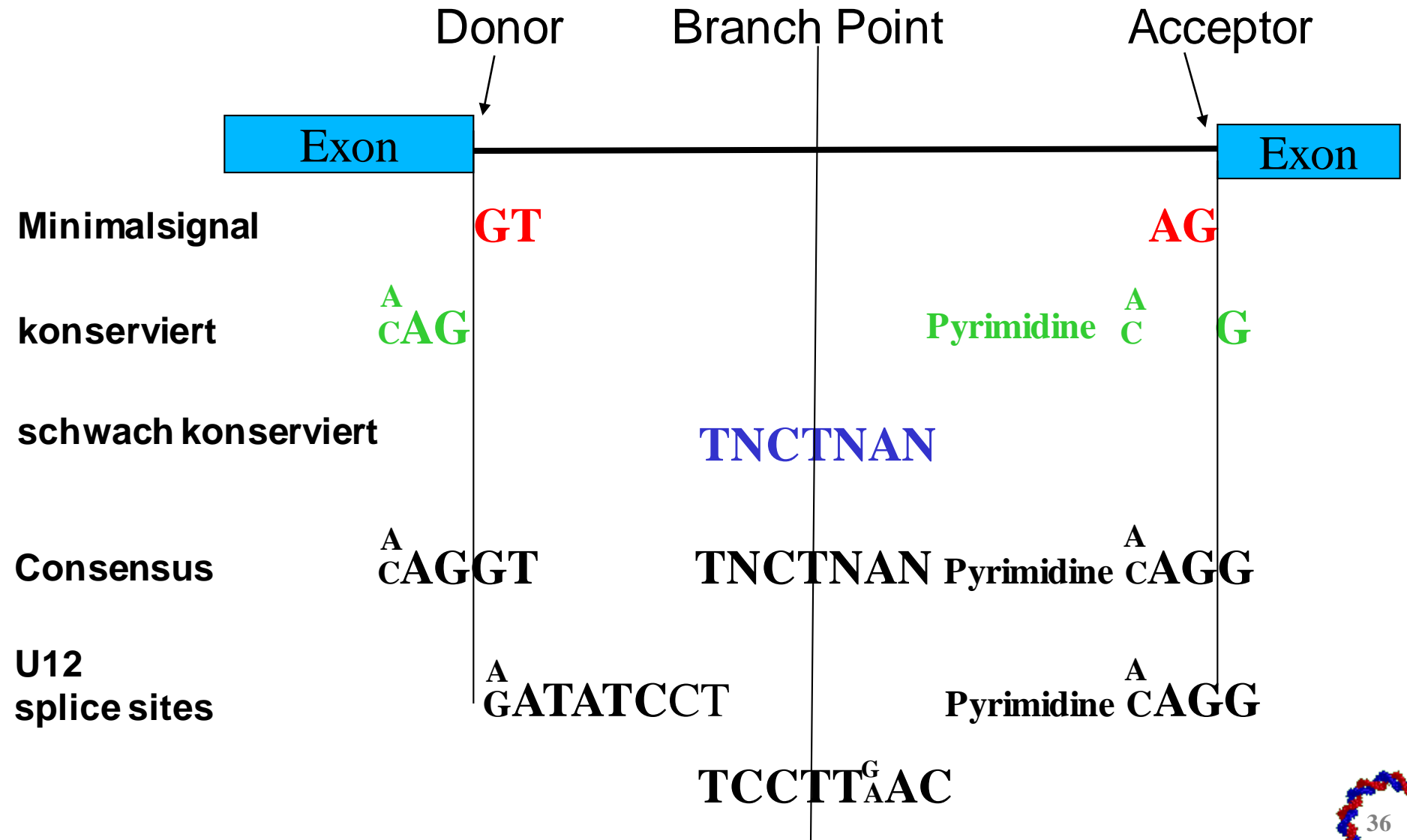
Splicing-Maschinerie



Rekrutierung der Splice Faktoren

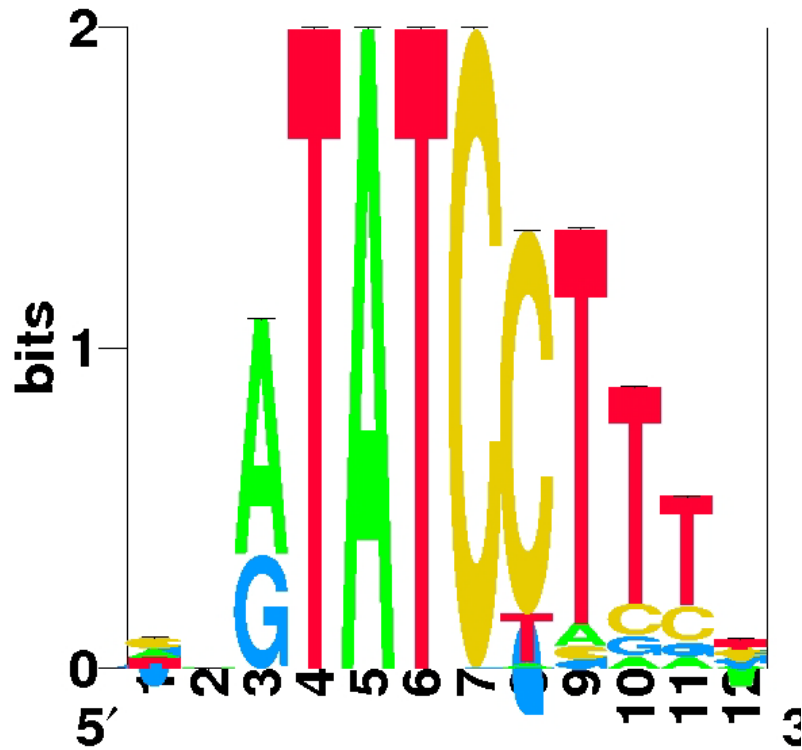
Namen der Komplexe mit Protein

Splice Signale

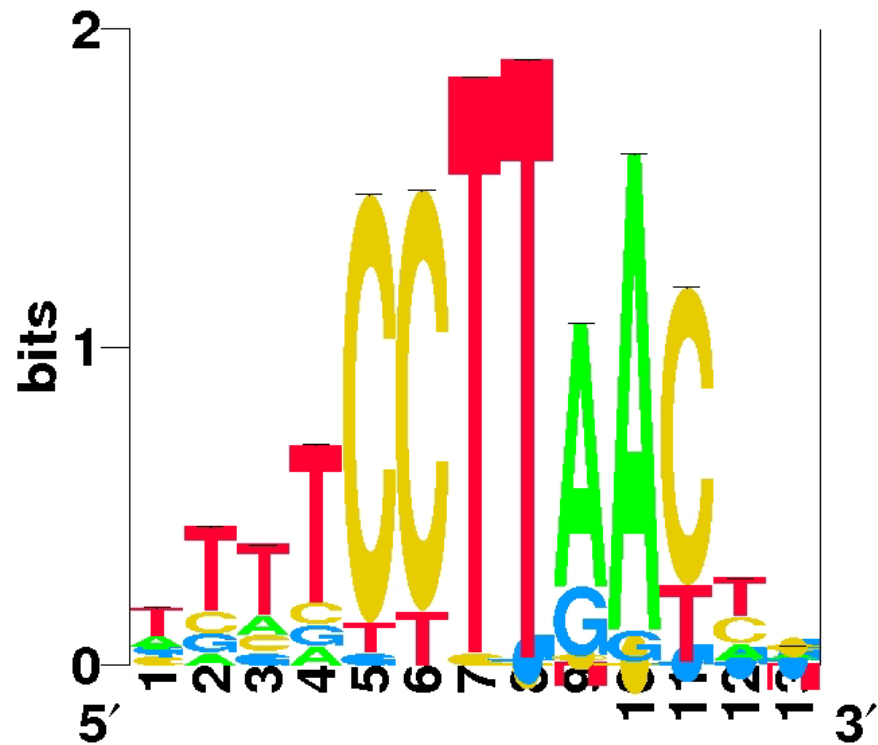


Intronlogo U12

donor



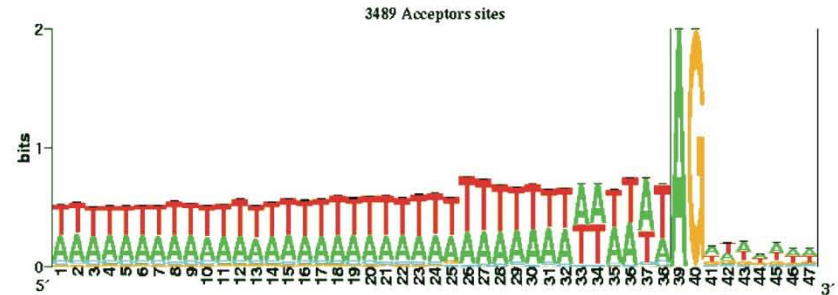
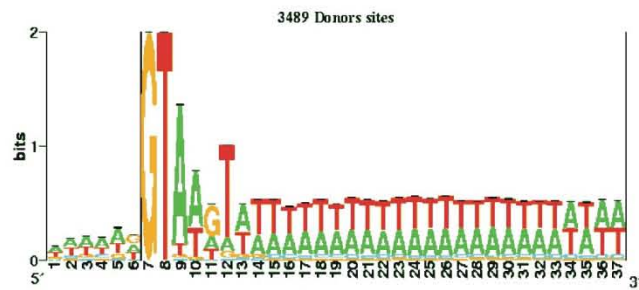
branch point



Gewichtetes Splice Motif in *D. discoideum*

donor

acceptor



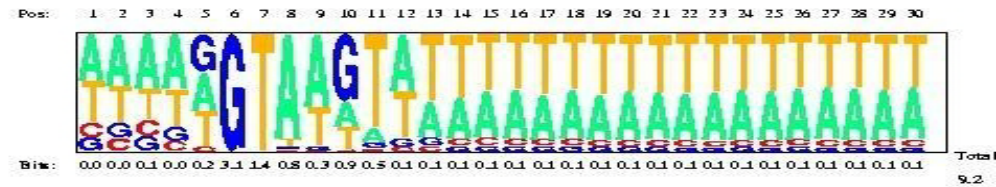
177 bases

Splice-Motive in A/T reichen Organismen

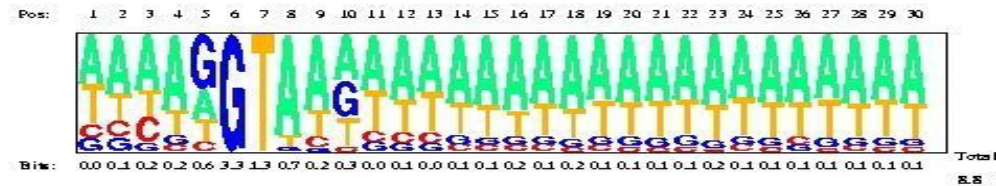
Splice sites

Donor site

Dictyostelium discoideum

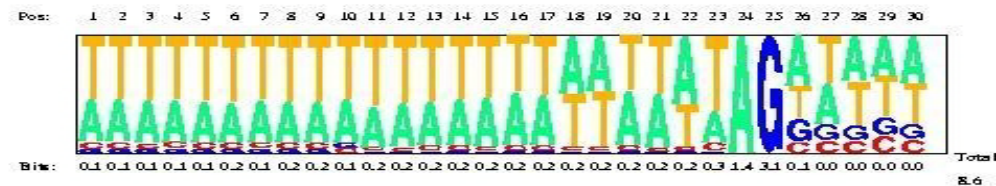


Plasmodium falciparum

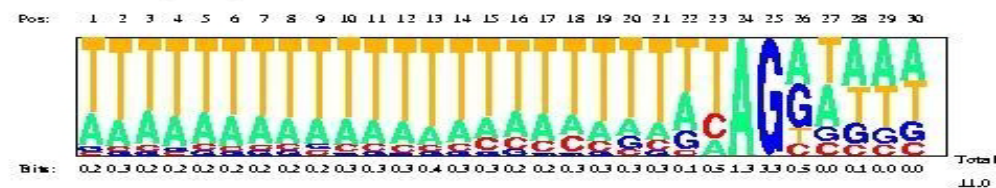


Acceptor site

Dictyostelium discoideum

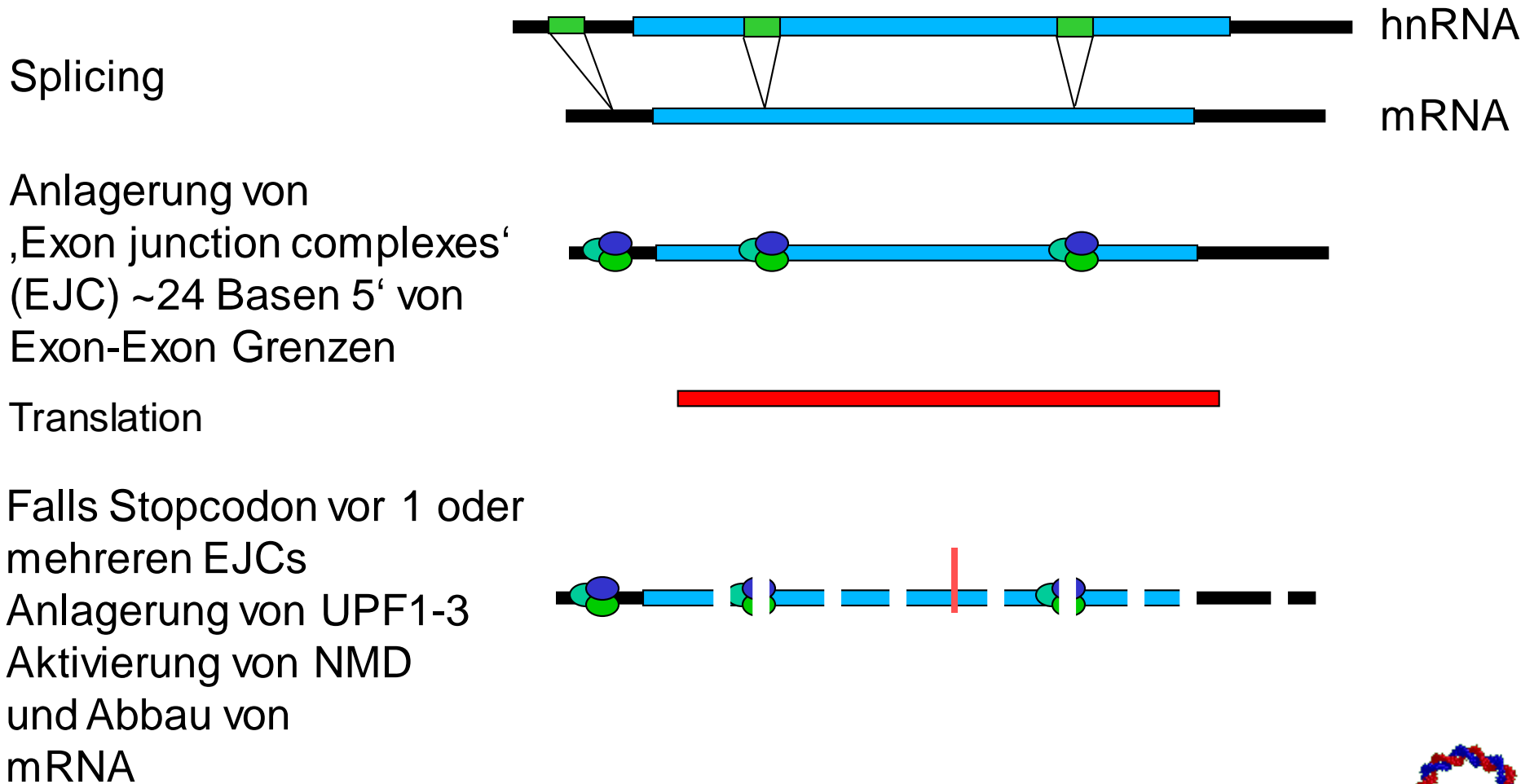


Plasmodium falciparum

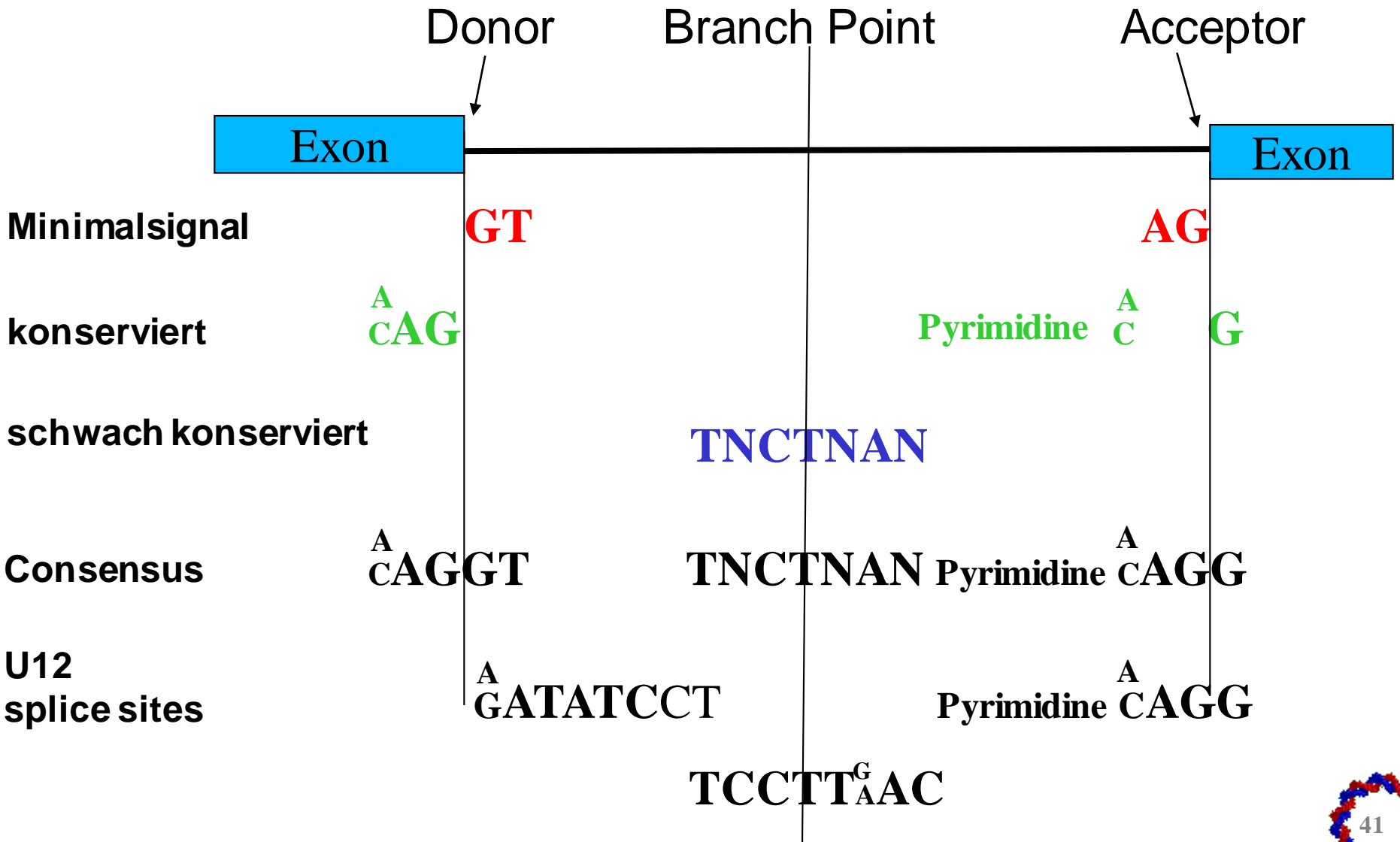


Erkennung fehlerhafter Transkripte

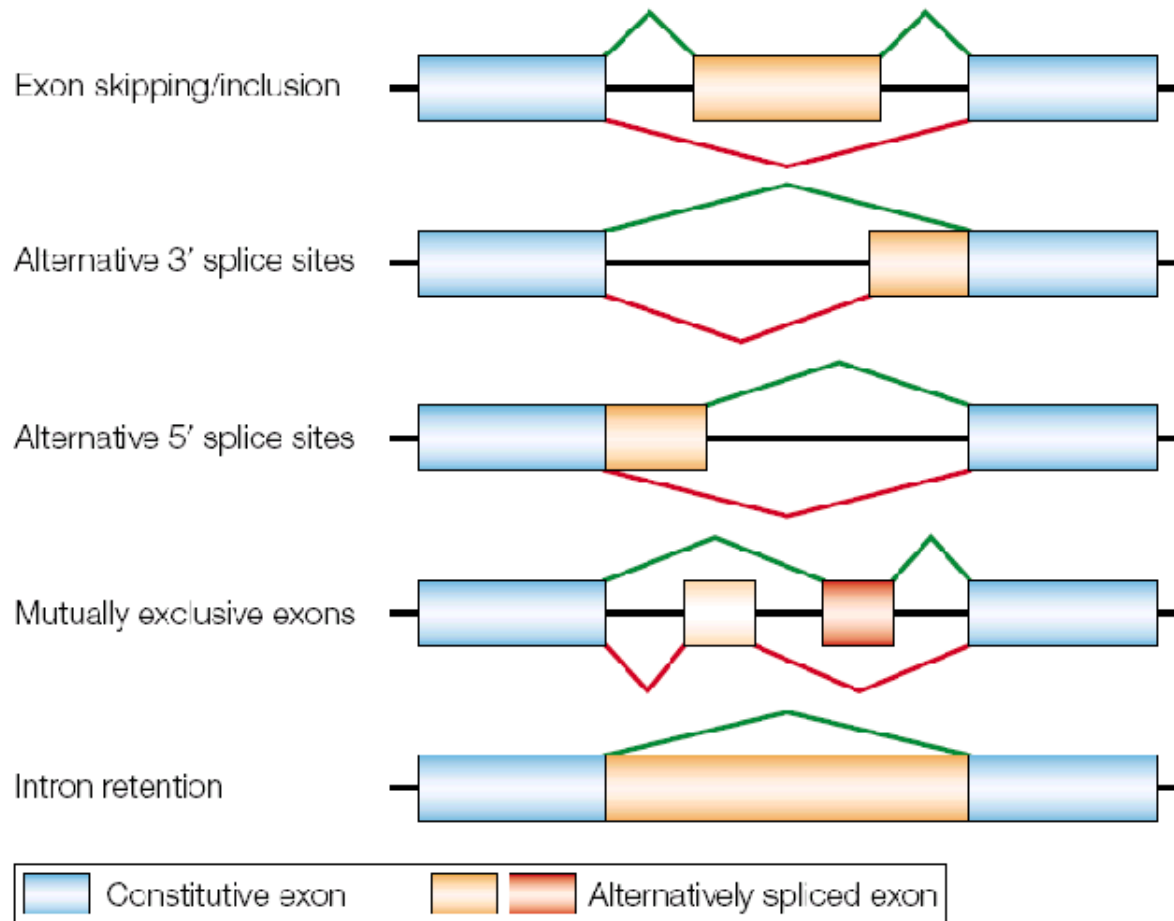
Nonsense mediated mRNA decay (NMD)



Splice Signale



Alternatives Splicing



Zusammenfassung I

DNA als genetisches Material
Verschlüsselung
Genorganisation
Gensignale, Detektion
Introns
Splicing



Fragen zum ersten Teil

- Was ist der IUPAC Code der DNA, welche Vor- und Nachteile besitzt er?
- Woraus wird eine Positional Weight Matrix (PWM) abgeleitet?
- Wie kann eine Sequenz auf Anwesenheit eines Motivs/Signals untersucht werden?
- Was sind die Mindestanforderungen für eine Splicestelle?
- Wie erfolgt Splicing?
- Welche Intronarten gibt es?



END

...

