Konservierte und Konsensus RNA Strukturen

Caroline Thurner

Institut für Theoretische Chemie und Molekulare Strukturbiologie

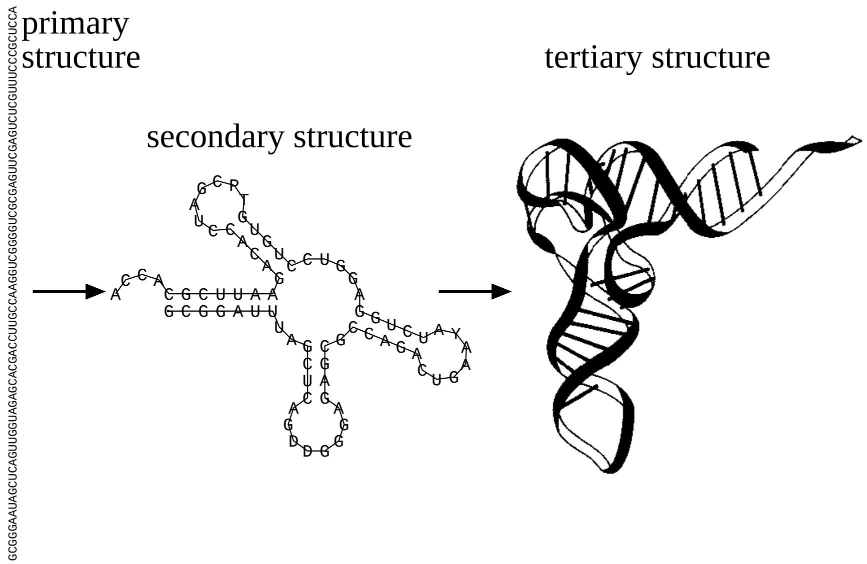
Universität Wien

Wien, 2004

Überblick

- Sekundärstrukturen von RNA
- Vorhersage von Sekundärstrukturen
- Anwendung auf Genome der Virenfamilie Flaviviridae
- Pseudoknoten
- Erweiterung des Algorithmus um Pseudoknoten vorherzusagen
- Anwendung auf RNA Beispiele die Pseudoknoten enthalten

RNA Sekundärstrukturen

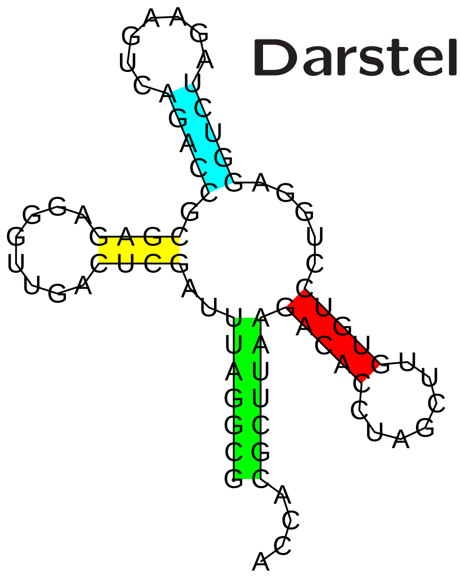


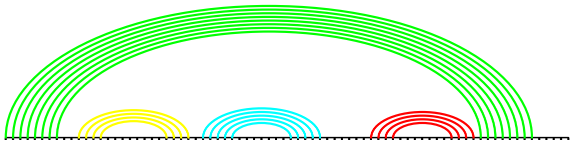
Sekundärstrukturen formen sich vor Tertiärstrukturen.

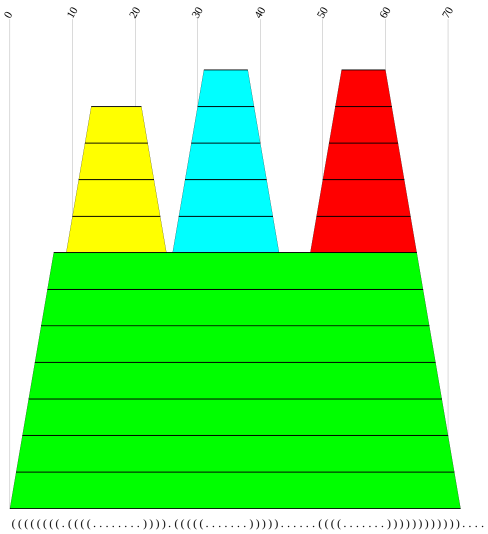
Eine Sekundärstruktur ist eine Liste von Basenpaaren, die folgende Bedingungen erfüllen:

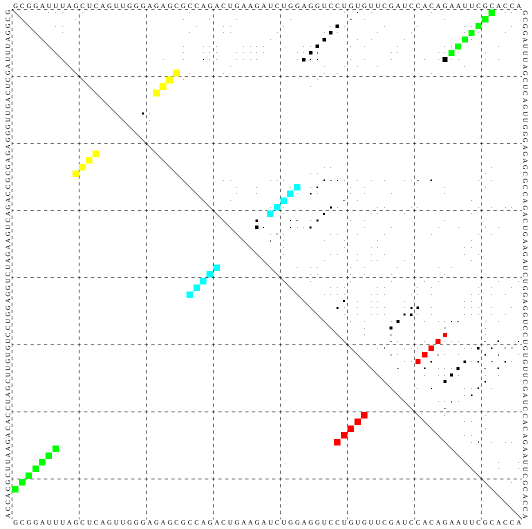
- Eine Base darf höchstens an einem Basenpaar teilnehmen.
- Basenpaare dürfen sich nicht überkreuzen, das heißt, keine zwei Paare (i.j) und (k.l) dürfen folgendermaßen angeordnet sein: $i < k < j < l$ (keine Pseudoknoten)

Darstellung von Sekundärstrukturen



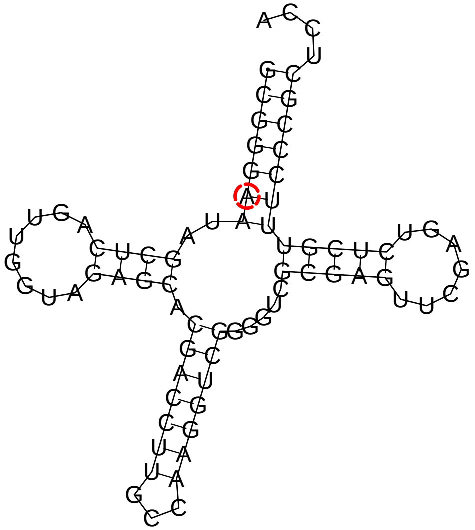




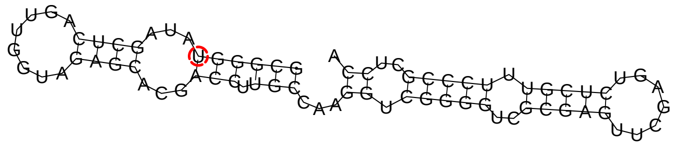


(((((((..((((........)))).(((((.......)))))......((((.......))))))))))))....

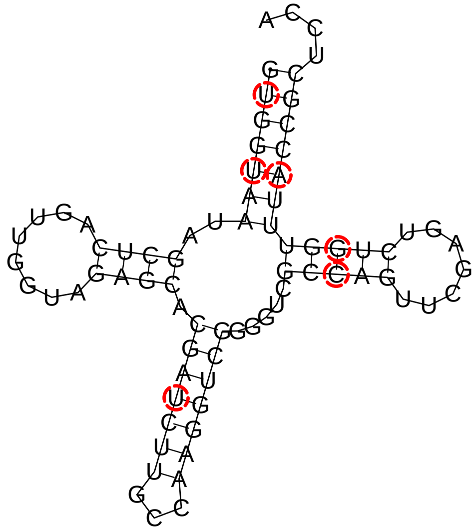
Der Effekt von Mutationen



original Struktur



Einzelmutation
- grosser Effekt in Struktur



Einige kompensatorische Mutationen
- gleiche Struktur

Was sagt uns das über konservierte Strukturen?

- 10% Zufallsmutationen in der Sequenz führen mit großer Sicherheit zu verschiedenen Strukturen.

- Gemeinsame Sekundärstrukturen in einer Gruppe von Sequenzen mit weniger als 95 % durchschnittlicher paarweiser Identität sind sehr wahrscheinlich das Ergebnis einer stabilisierten Selektion.

- Wenn Selektion Strukturelemente aufrechterhält, dann müssen diese eine Funktion haben.

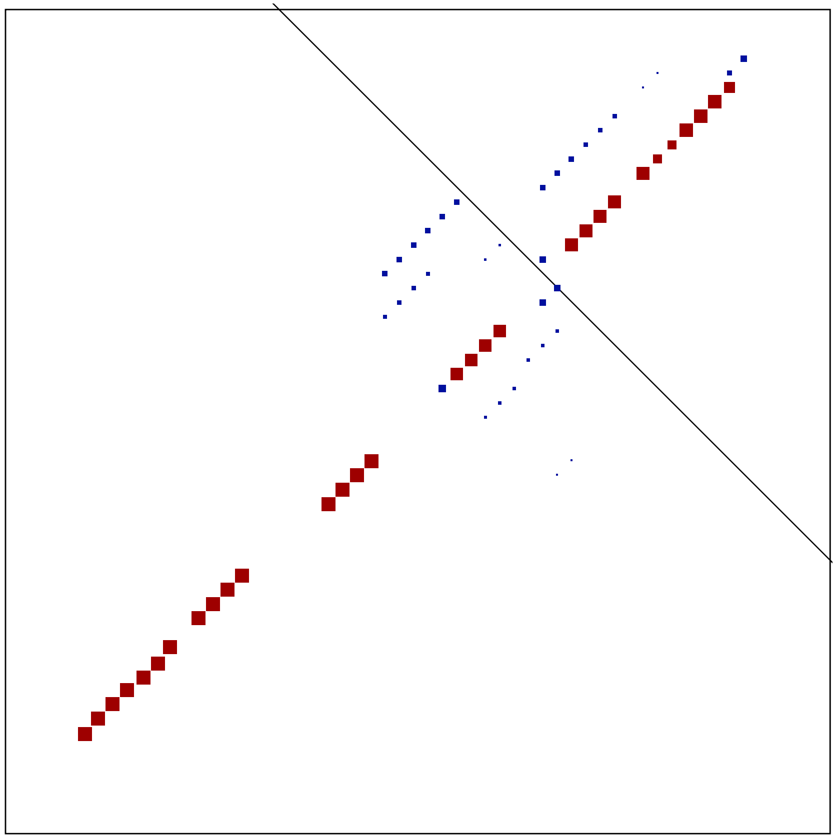
Suche nach konservierten Sekundärstrukturen

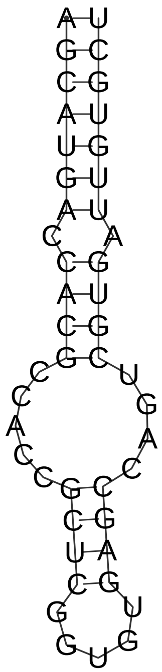
- Multiples Sequenzalignment (ClustalW)
- Ermittlung von Basenpaar-Wahrscheinlichkeits Matrix
- Verknüpfung von Strukturvorhersage und Sequenzalignment zur Erstellung einer Liste möglicher Kandidaten konservierter Strukturen
- Kompensatorische und inkonsistente Mutationen geben Auskunft, ob eine Struktur angenommen oder verworfen wird.

Free Software, `C` source code and fold servers available at
`http://www.tbi.univie.ac.at/~ivo/RNA/`

Thermodynamisches Falten: RNAfold

Vorhersage der Basenpaar Wahrscheinlichkeits Matrix basierend auf McCaskills Algorithmus zur Bestimmung von Zustandsfunktionen



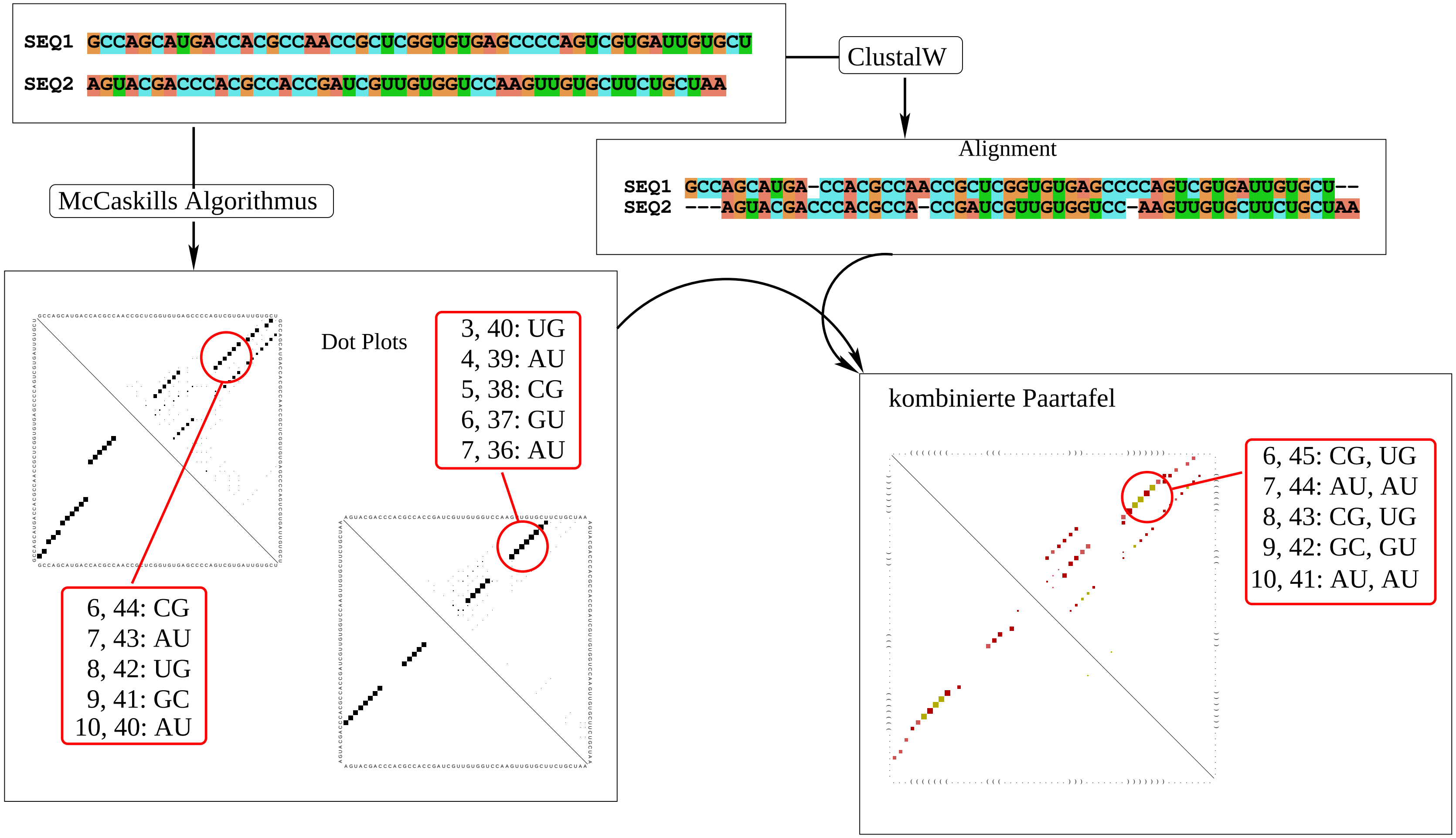


Thermodynamisches Gleichgewicht:

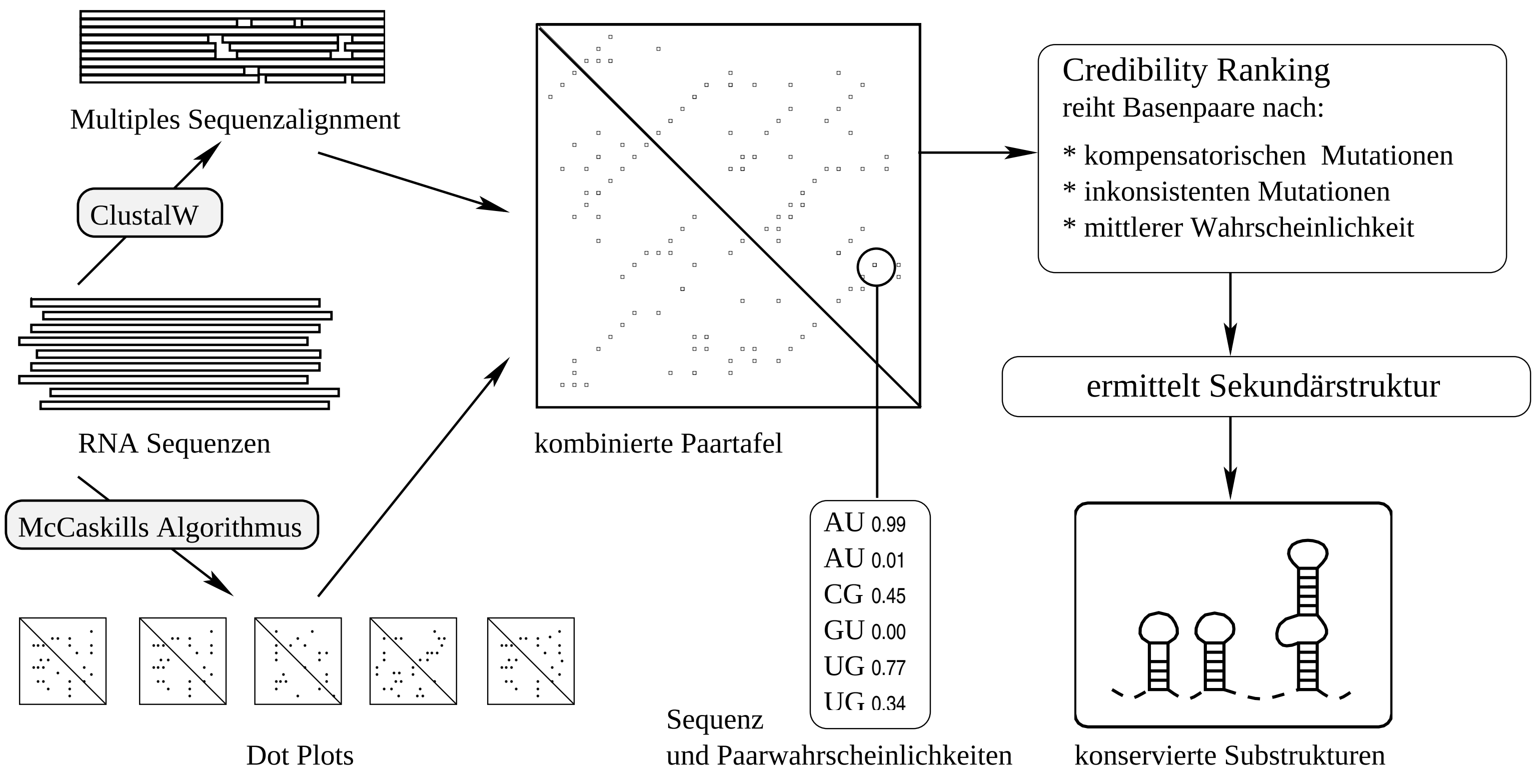
alternative niedrige Energiezustände

sind besetzt

Verbindung von Strukturvorhersage und Sequenz Alignment



Alidot

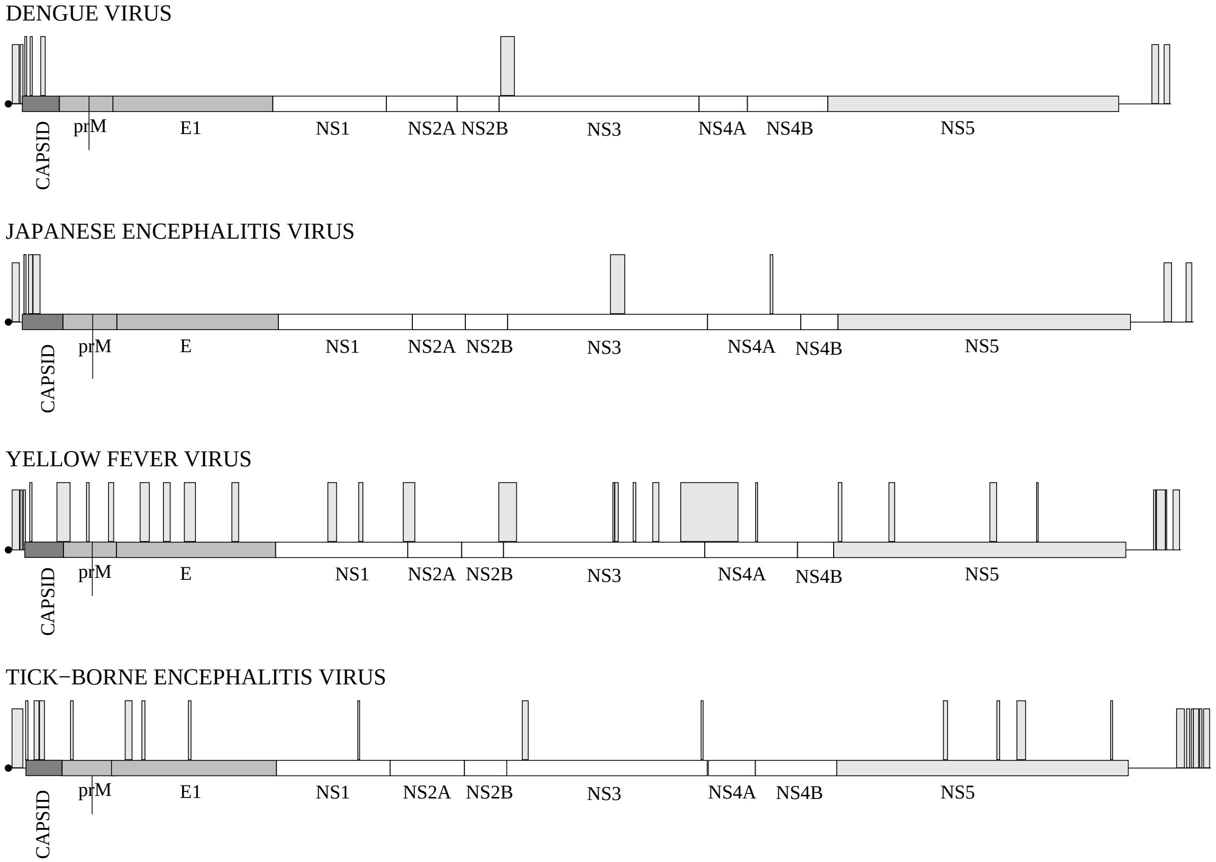


Multiples Sequenzalignment
ClustalW
RNA Sequenzen
McCaskills Algorithmus
Dot Plots
kombinierte Paartafel
AU 0.99
AU 0.01
CG 0.45
GU 0.00
UG 0.77
UG 0.34
Sequenz
und Paarwahrscheinlichkeiten
Credibility Ranking
reiht Basenpaare nach:
* kompensatorischen Mutationen
* inkonsistenten Mutationen
* mittlerer Wahrscheinlichkeit
ermittelt Sekundärstruktur
konservierte Substrukturen

Familie Flaviviridae

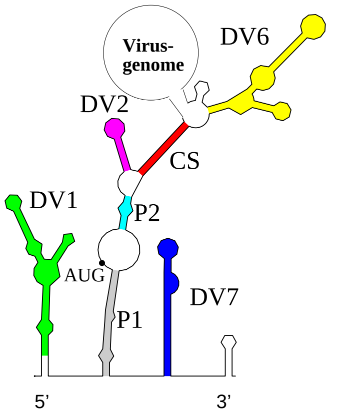
Genus	Species	homol.	Transl.
Hepatitis G Virus		89.8	IRES
Hepatitis C Virus		87.1	IRES
Pestivirus		74.9	IRES
Flavivirus	Dengue Virus	85.4	CAP
	Jap. Enceph. Virus	95.6	CAP
	Yellow Fever Virus	91.7	CAP
	Tick-borne Enceph. Virus	69.8	CAP

Flavivirus: Genomorganisation

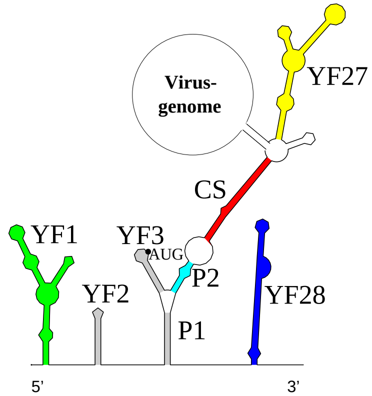


Genomkarte von Flaviviren mit konservierten Strukturelementen

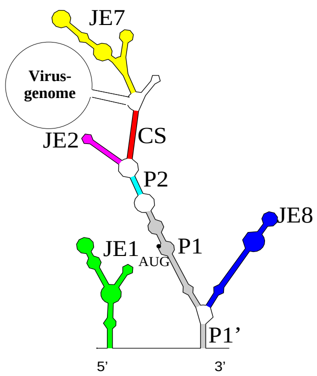
Konservierte Strukturen im Genus Flavivirus



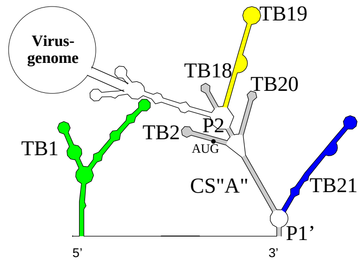
Dengue Virus



Yellow Fever Virus

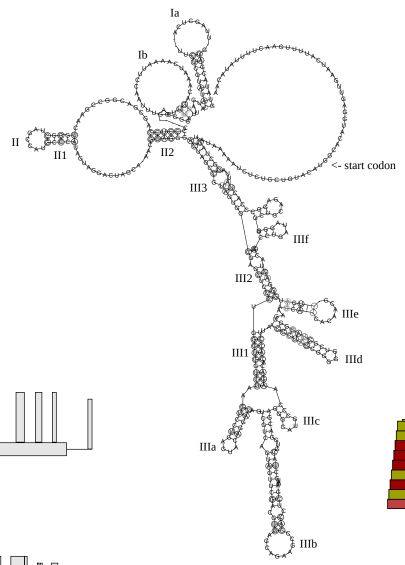


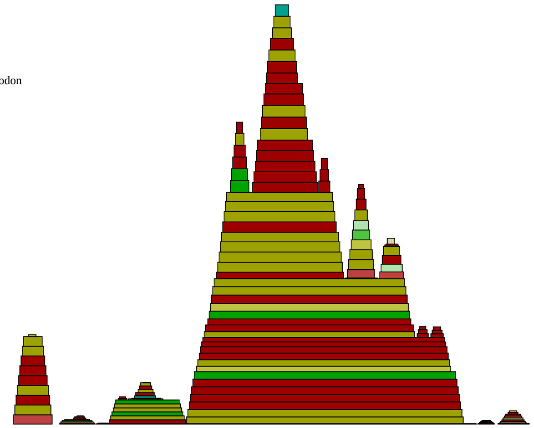
Japanese Encephalitis Virus

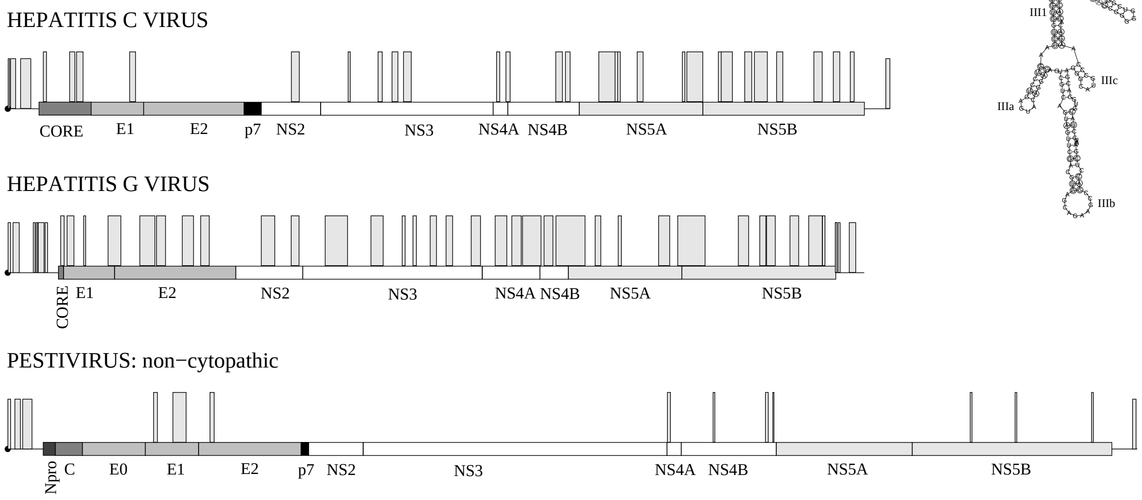


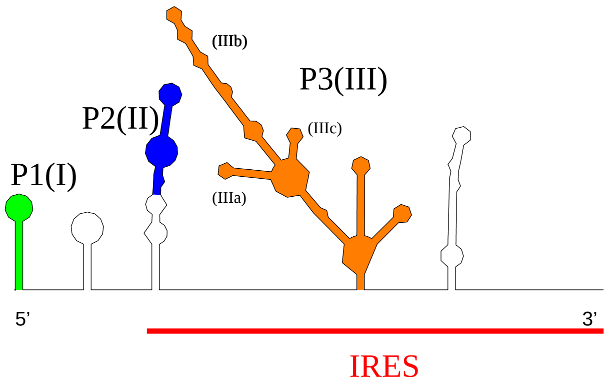
Tick-borne Encep. Virus

Hepatitis G Virus, Hepatitis C Virus und Pestivirus

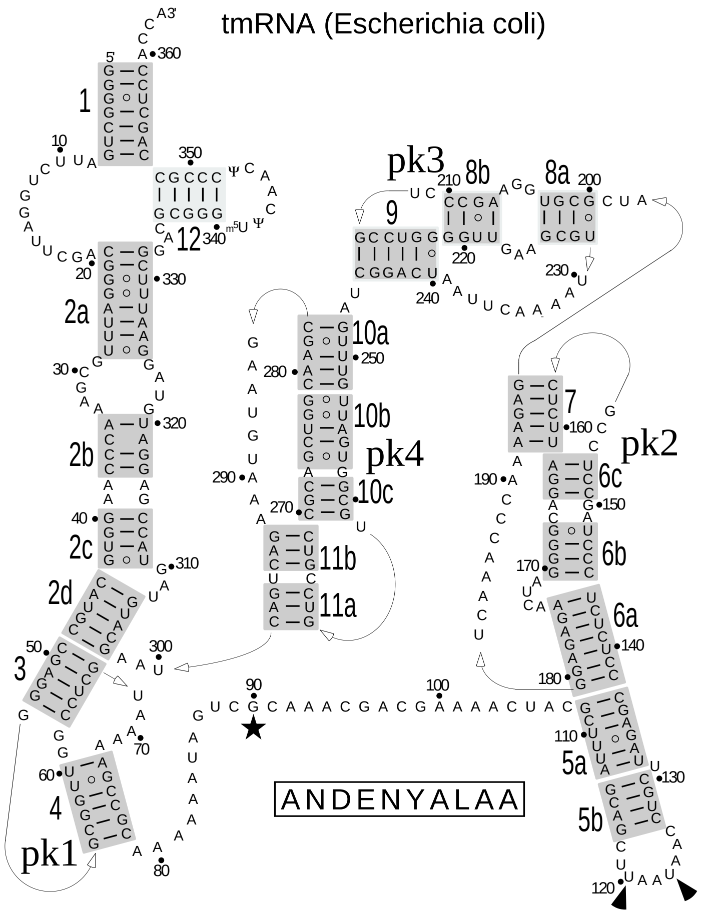


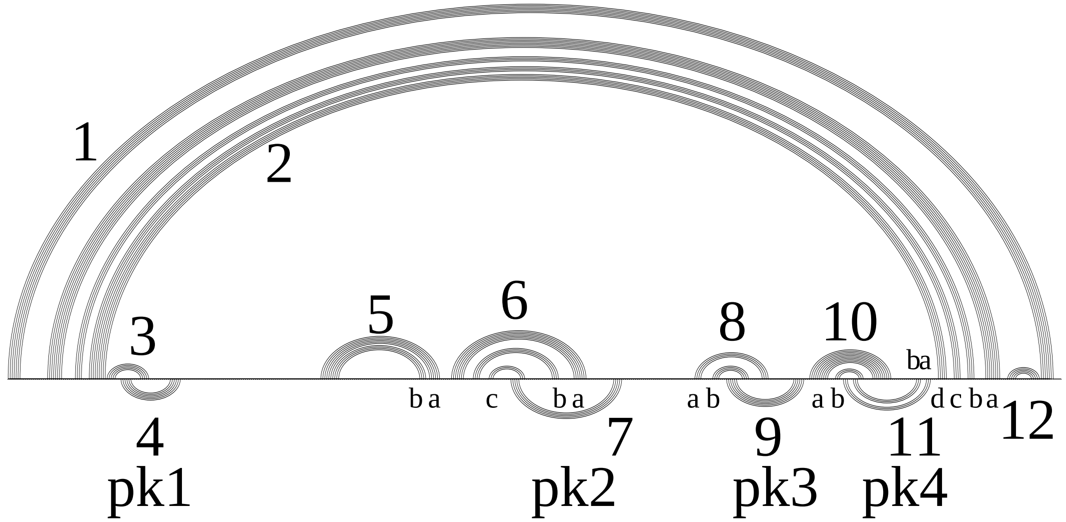






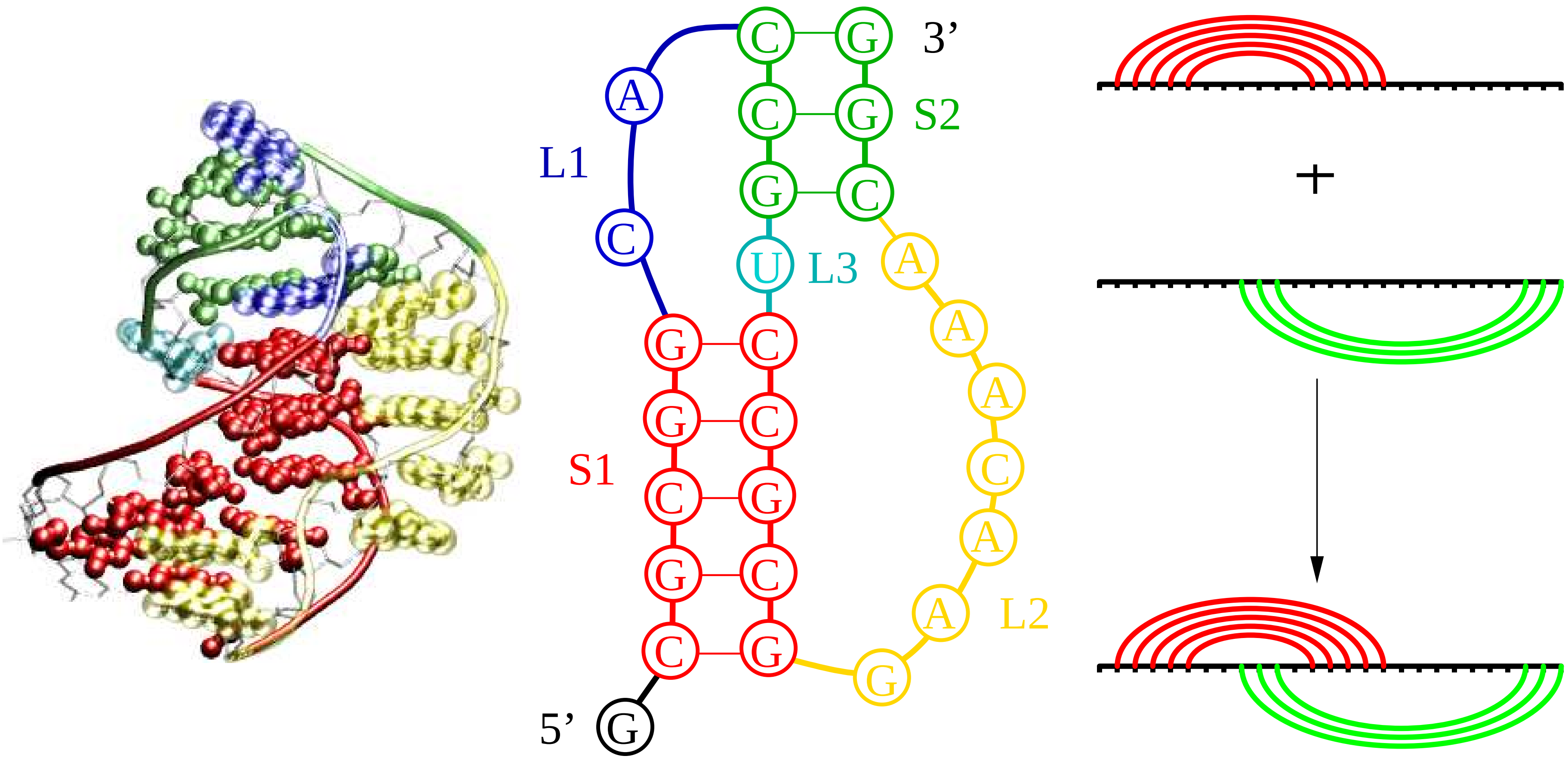
Pseudoknoten in tmRNA von E. coli



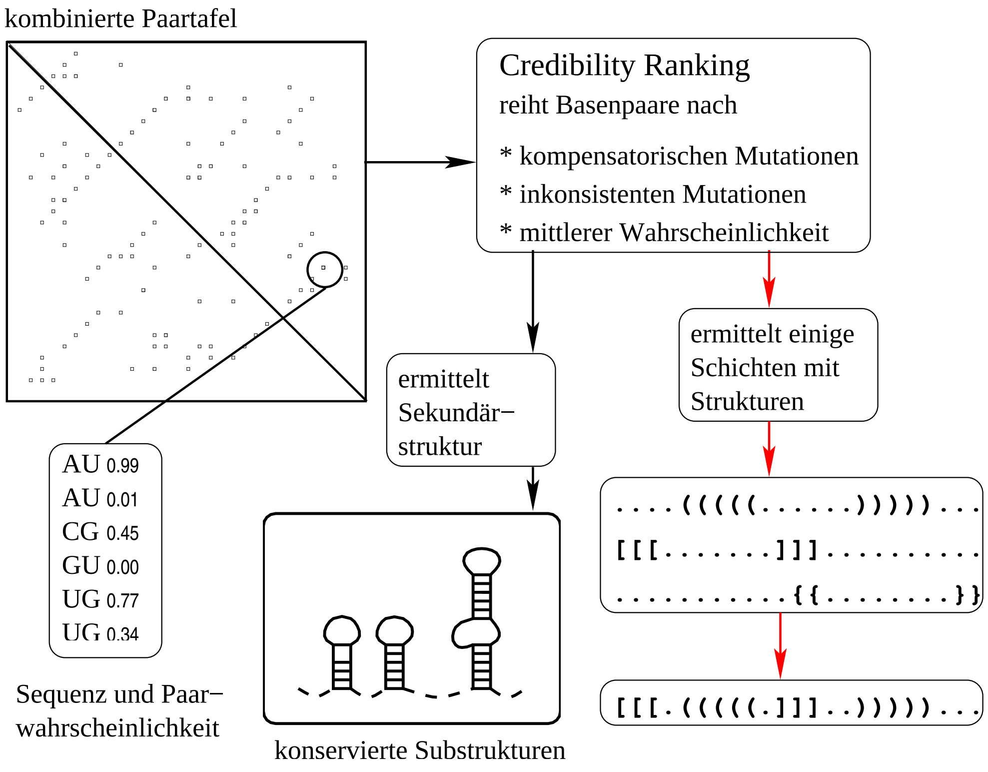


Zwieb, C. et al., NAR (1999) 27(10):2063

Was sind Pseudoknoten?



aliknot

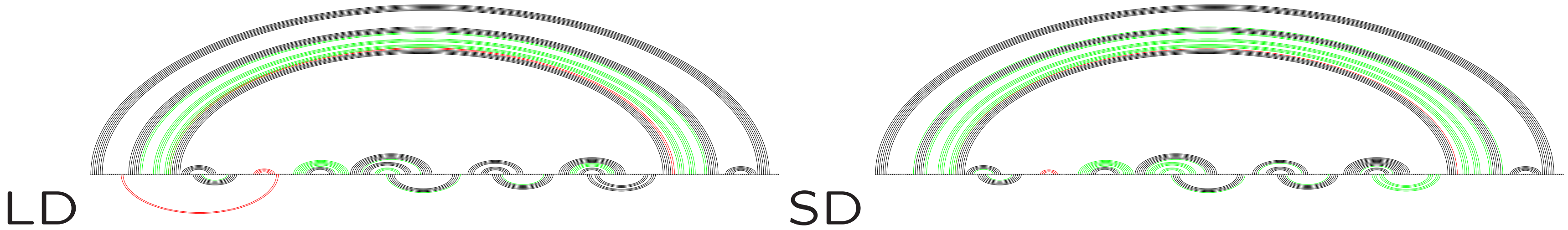


Zwei Wege:

1. Zuerst ursprüngliche Sekundärstruktur in erste Zeile schreiben, dann andere Paare nach ihrem Ranking in höhere Zeilen füllen (LD).

2. Zuerst alle möglichen Stacks finden und bewerten, dann ganze Stacks in Zeilen füllen, wobei die erste Schicht wieder die wahrscheinlichste Sekundärstruktur enthält (SD).

Ergebnisse von aliknot



tmRNA verglichen zu tmRNA aus E. coli

	N	korrekte Basenpaare	korrekte Stems	falsch positive Helices	korrekte Pseudoknoten
SRP RNA	7	90.1%	6/6	0	1/1
RNase P RNA	8	74.2%	16/18	3	1/2
tmRNA	8	78%	12/12	2	4/4

Vergleich mit anderen Algorithmen

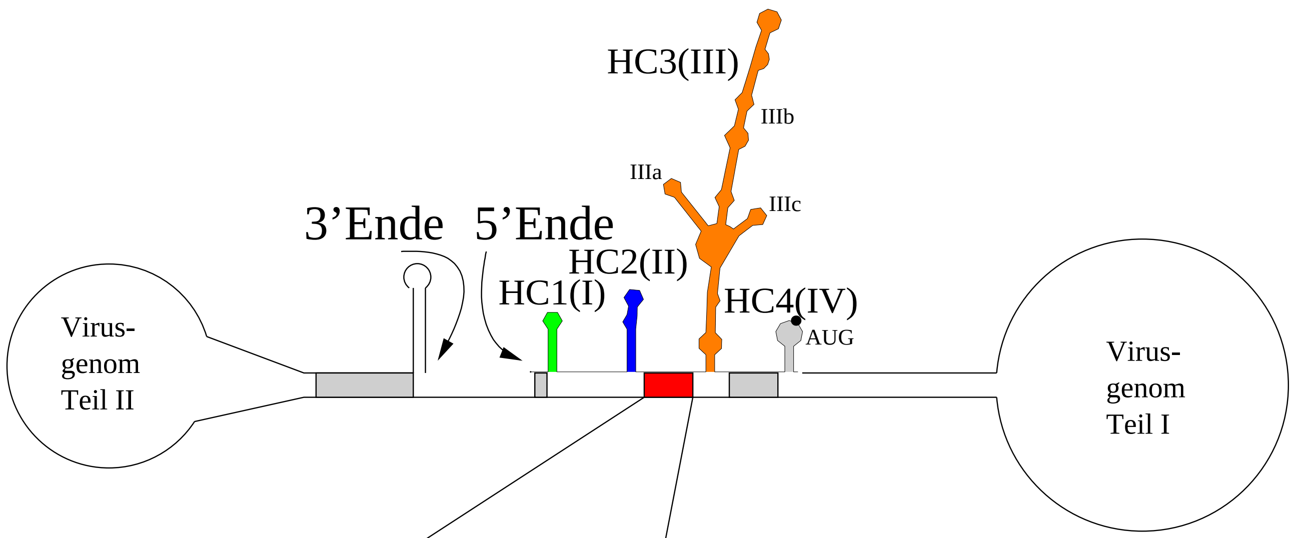
Am Beispiel SRP RNA berechnet mit SD				
	korrekte Basenpaare	korrekte Helices	falsch positive Helices	korrekte Pseudoknoten
aliknot	86 %	8/8	0	1/1
hxmach	91.9%	8/8	0	1/1
ilm	86%	7/8	1	0/1

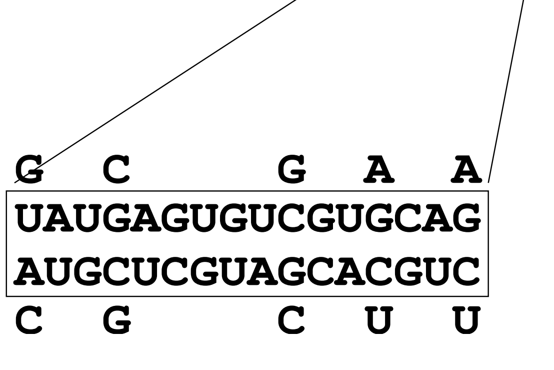
Danke:

- Peter F. Stadler
- Ivo L. Hofacker
- Christina Witwer, Christoph Flamm
- und allen meinen Freunden am TBI

Acknowledgements an Andrea Tanzer

Mutationen zur Überprüfung der vorhergesagten Struktur

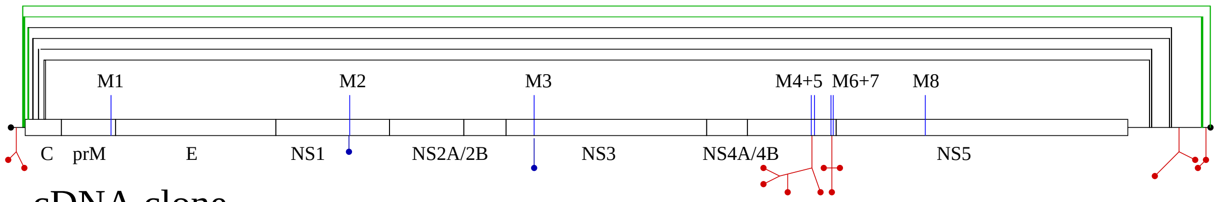




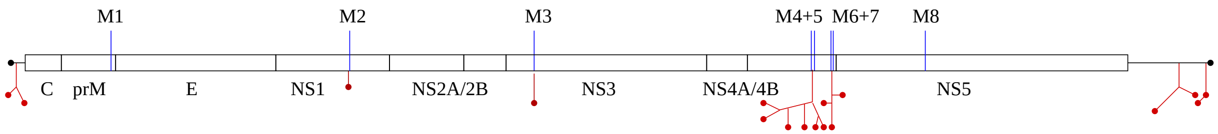
Pos	orig	m1	m2	m3
95	U	G	G	
98	G	C	C	
104	C	G	G	
107	G	A	A	
110	G	A	A	
8528	C		U	U
8531	C		U	U
8534	G		C	C
8540	C		G	G
8543	A		C	C
Energie	23.9	8.1	19.9	8.8

Strang Neudoerfl, ein Tick-borne Encephalitis Virus

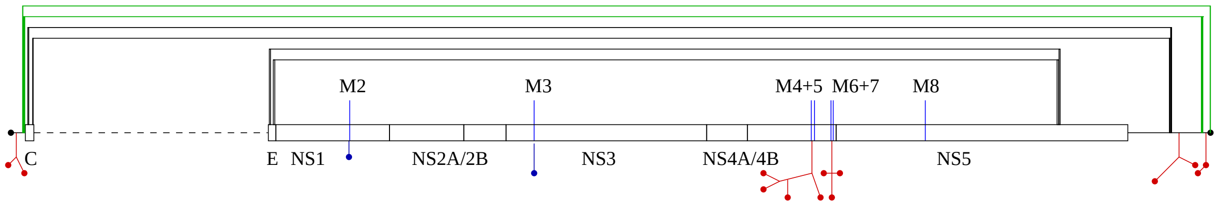
Neudoerfl: U27495



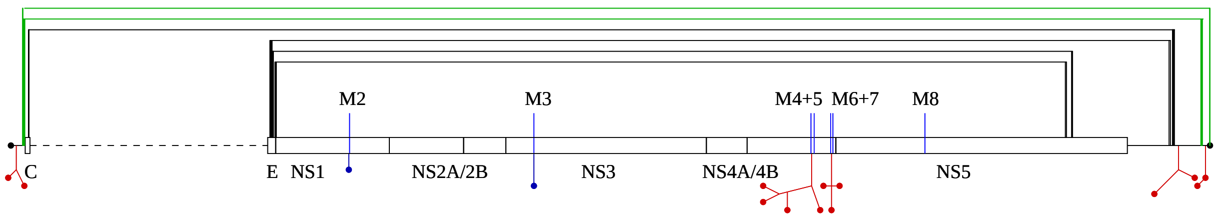
cDNA clone



dCME



C15



C17