



MikroRNA-Bioinformatik

Annalisa Marsico

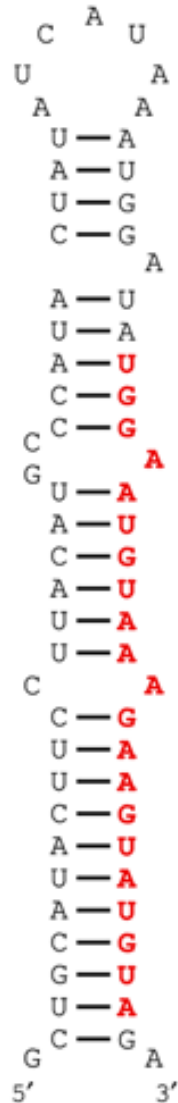
High-throughput Genomics, FU Berlin

07.12.2016



- Einführung in die Welt der nicht-kodierenden RNAs
 - Insbesondere mikroRNA
 - Molekulare Funktionen von microRNAs
- Was sind die Regeln der mikroRNA-Regulation
- MicroRNA Zielvorhersagealgorithmen
 - Der **miRanda** Algorithmus
 - Wie finden Wir oder priorisieren microRNA Ziele

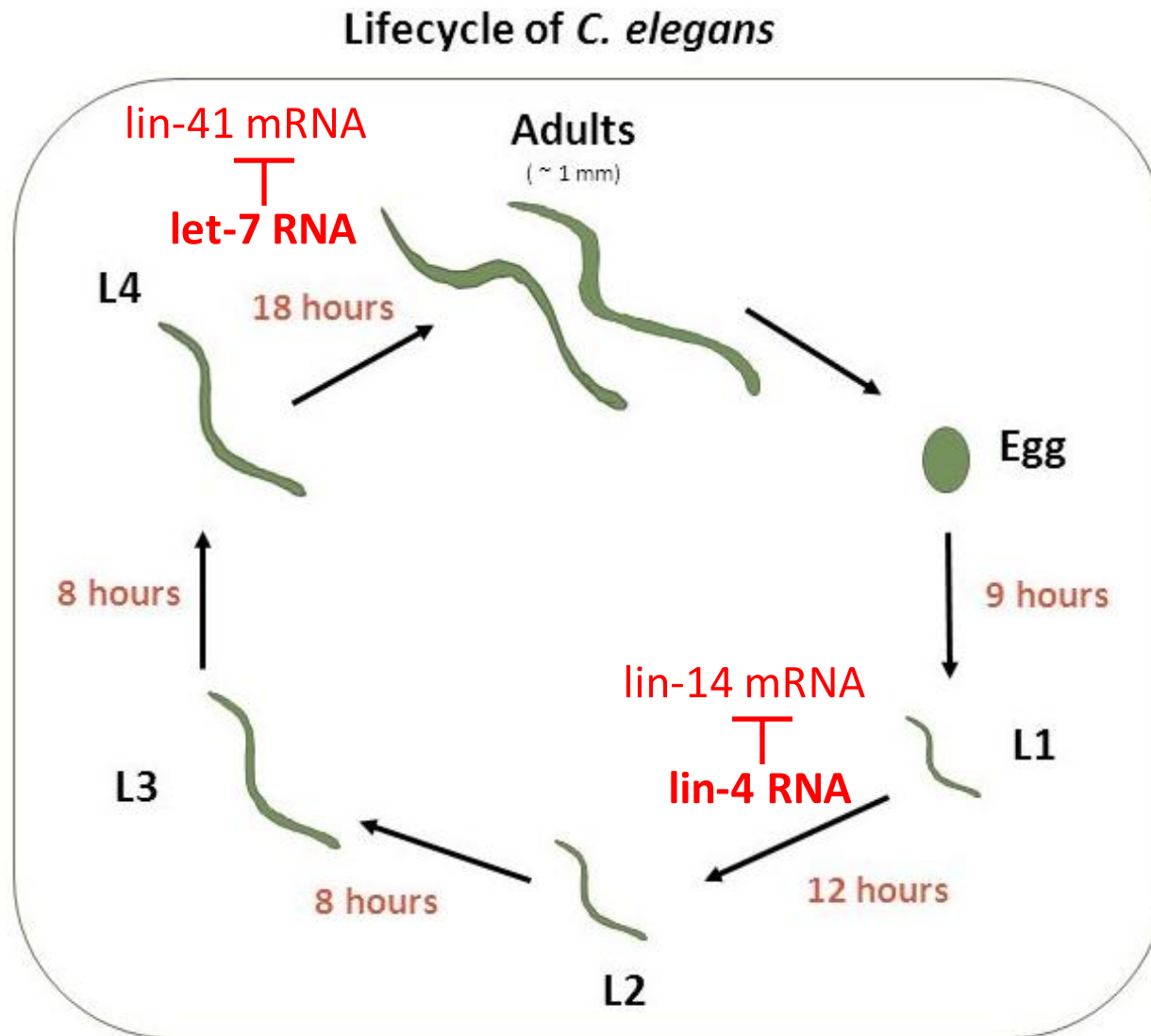
MikroRNA - Beispiel



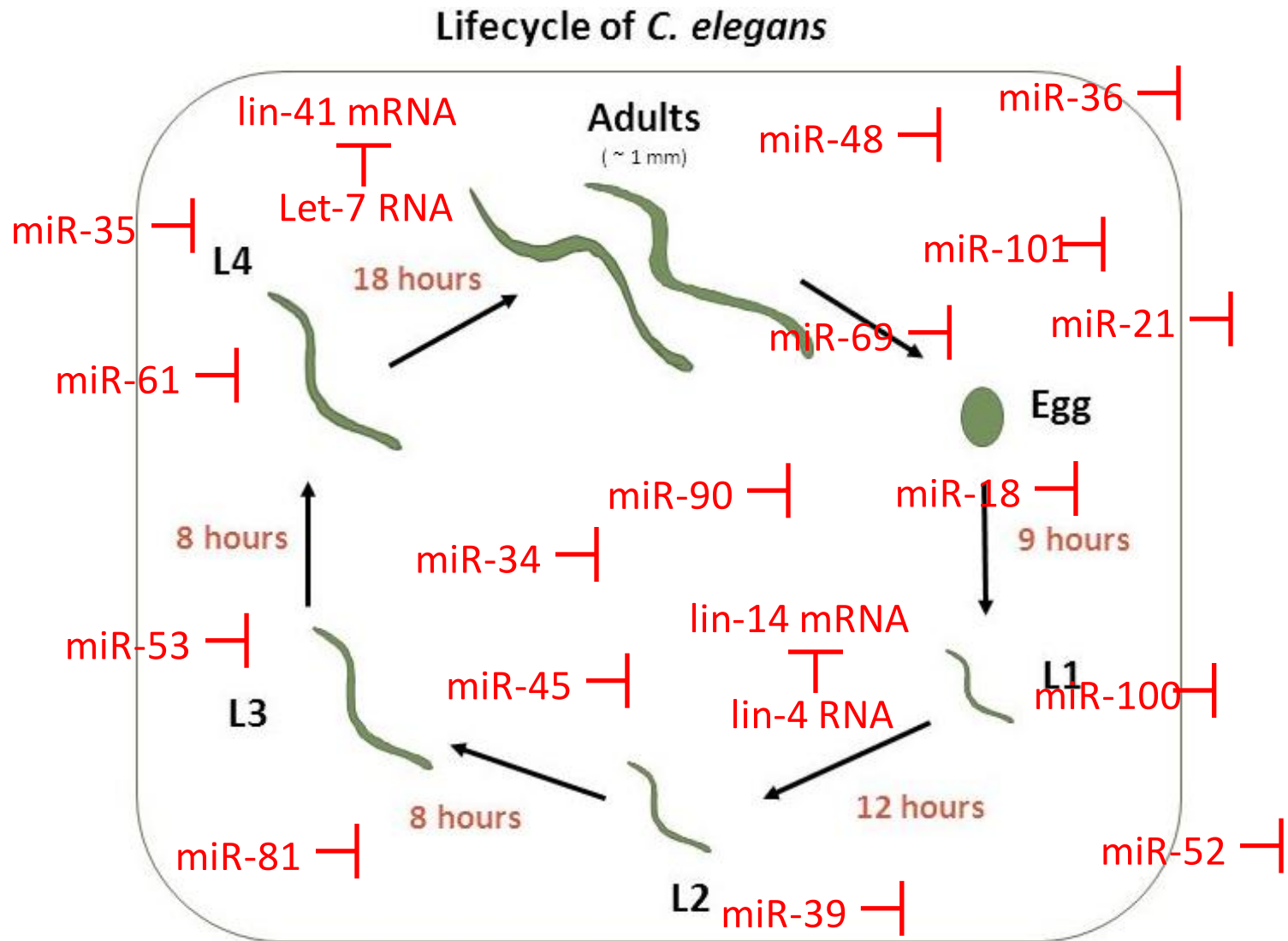
miR-1

~ 22 Nukleotiden lang

Die ersten entdeckten microRNAs



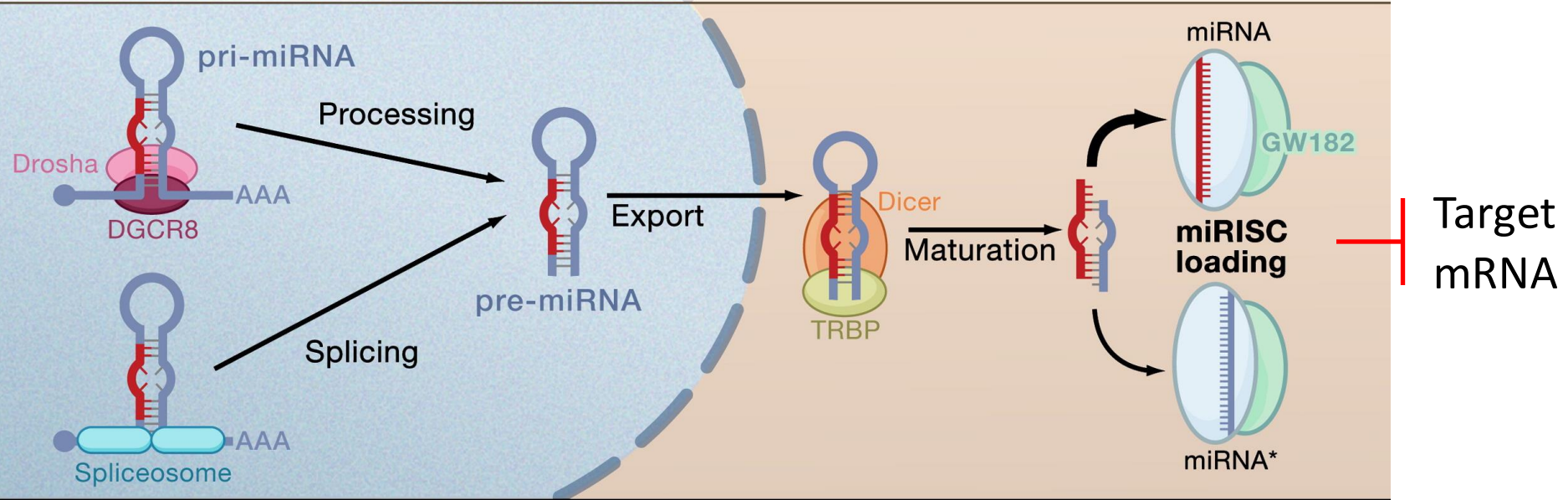
Die ersten entdeckten microRNAs



MicroRNA-Biogenese bei Tieren



Animals



MicroRNA-Biogenese bei Tieren



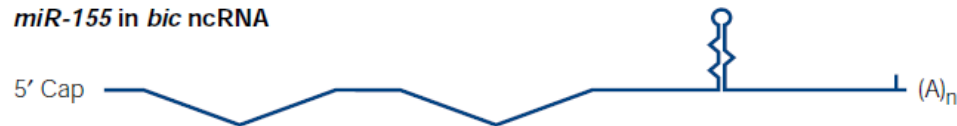
a *miR-23a~27a~24-2*



miR-21



miR-155 in *bic* ncRNA



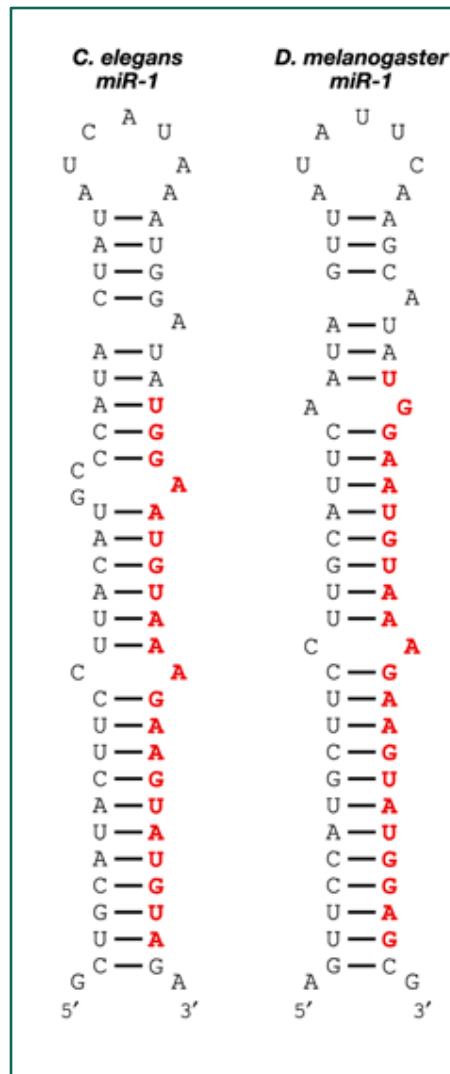
b *miR-15a~16-1* in *DLEU2* ncRNA



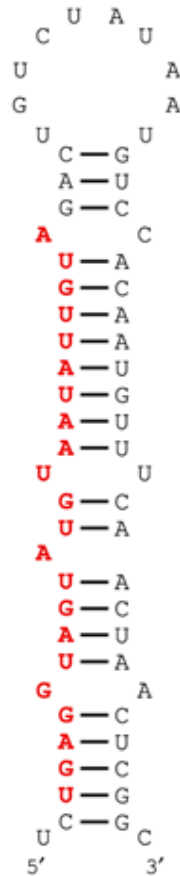
c *miR-106b~93~25* in *MCM7* pre-mRNA



mikroRNA Konservierung



C. elegans miR-84



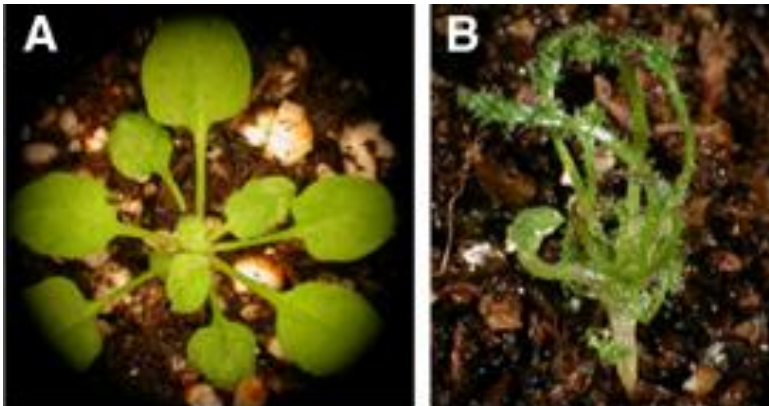
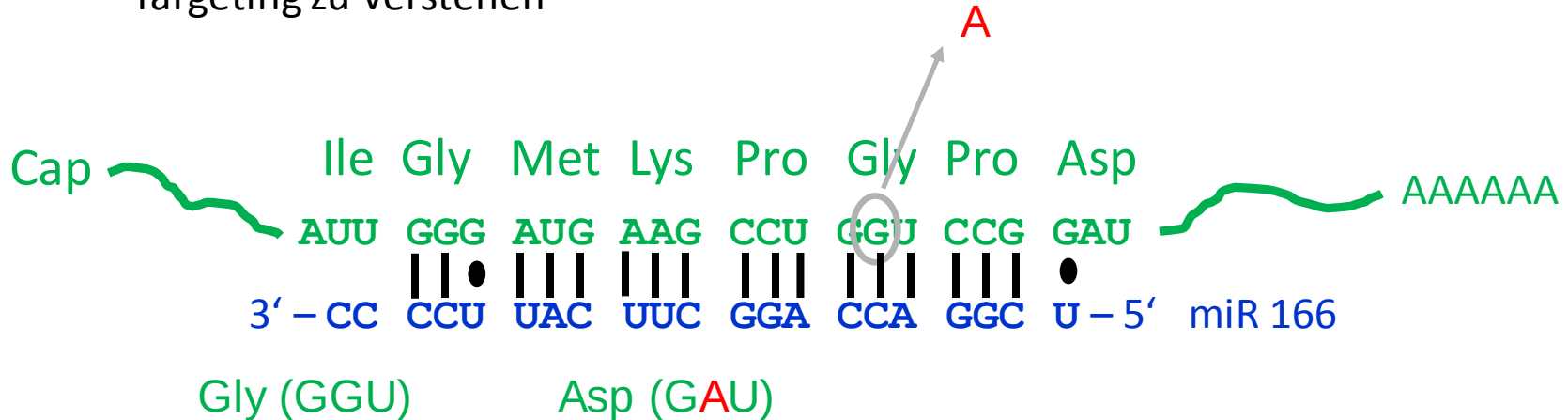
miR-1 orthologs

Human	UGGAAUGUAAAGAAGUAUGUA _u
Fly	UGGAAUGUAAAGAAGUAUG _g A
Worm	UGGAAUGUAAAGAAGUAUGUA

In Pflanzen mikroRNA-Ziele umfassende Komplementarität



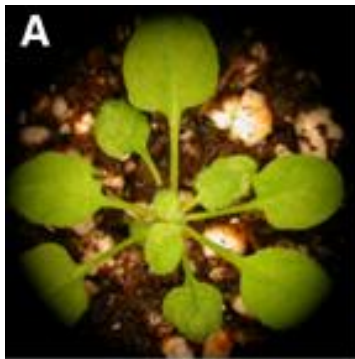
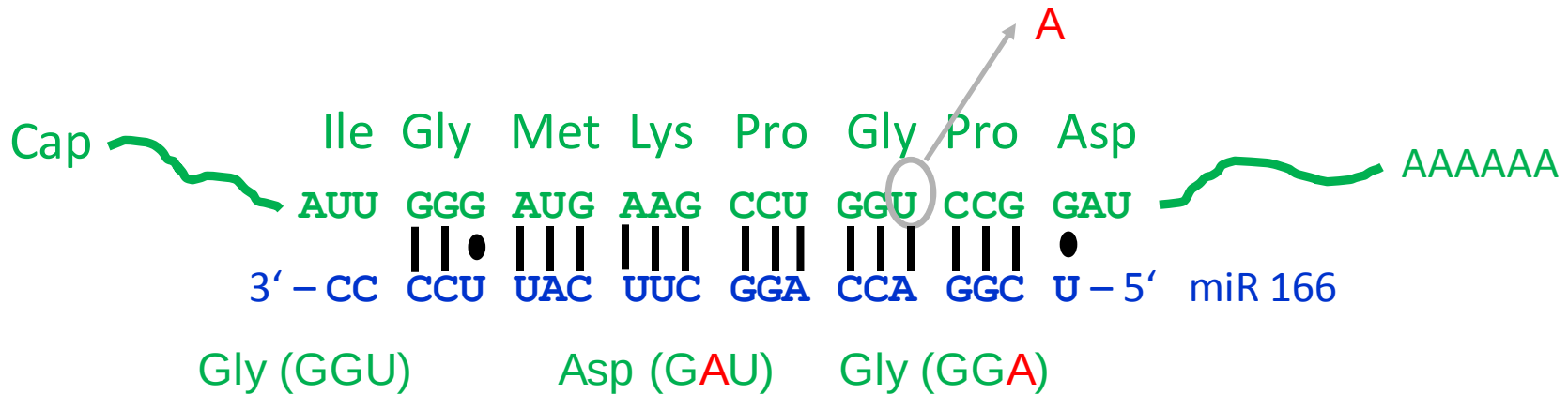
Beispiel für das PHABULOSA-Gen, um die microRNA-Funktion und microRNA-mRNA Targeting zu verstehen



In Pflanzen mikroRNA-Ziele umfassende Komplementarität



Example of the PHABULOSA gene to understand microRNA function and microRNA-mRNA targeting



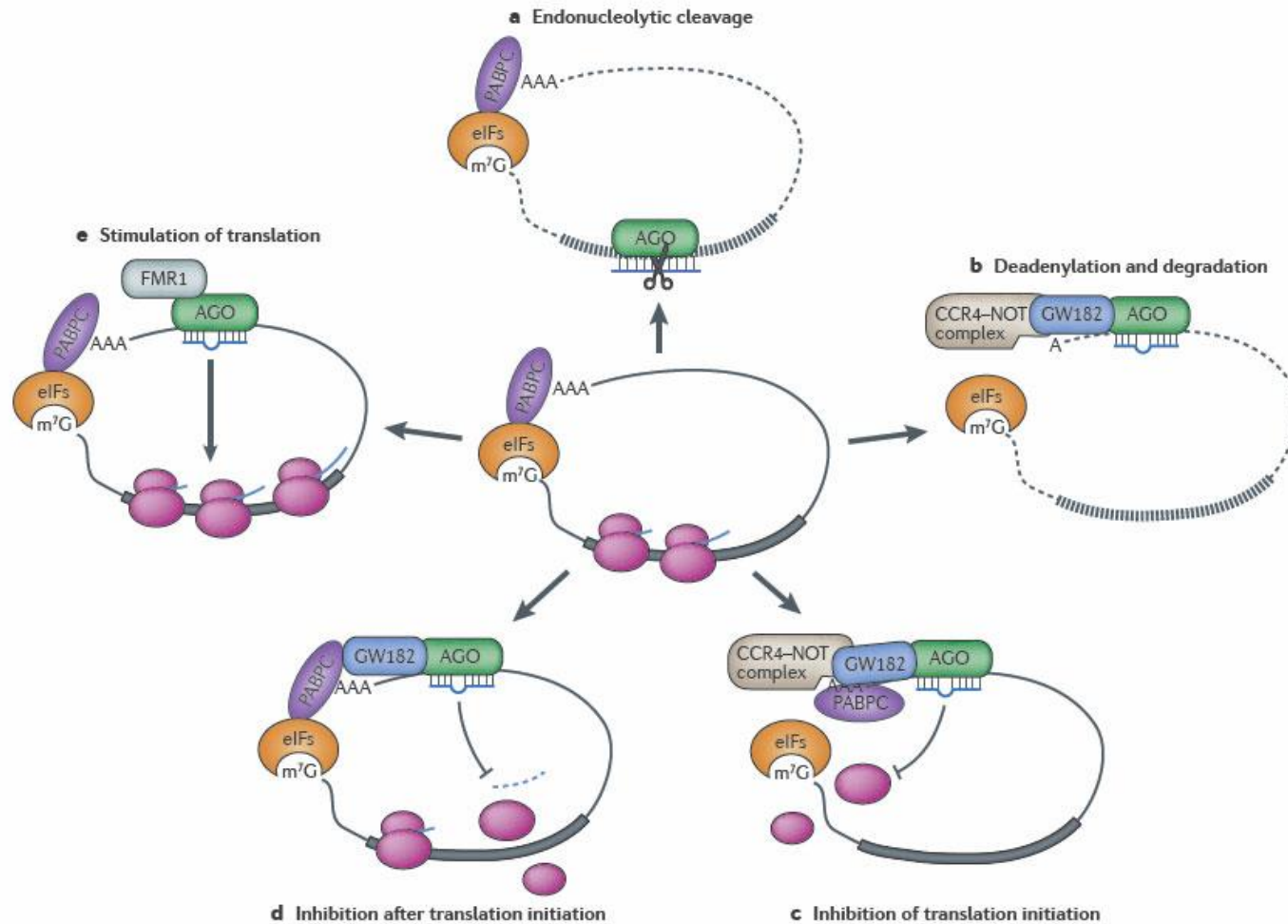
MiRNA-Wirkmechanismen und offene Fragen



- Die jeweilige MikroRNA (miRNA) wird in einen aus mehreren Proteinen bestehenden Komplex, den “RNA-induced silencing complex” (RISC) integriert
- Der RISC-Komplex bindet an die Ziel-mRNA an eine zur miRNA komplementäre Basensequenz in deren 3'- UTR-Region
- Diese Bindung führt zur Hemmung der Proteinsynthese und/oder Degradation der Ziel-mRNA



MiRNA-Wirkmechanismen und offene Fragen



Nature Reviews | Genetics

MiRNAs: Bindung an Ziel-mRNA

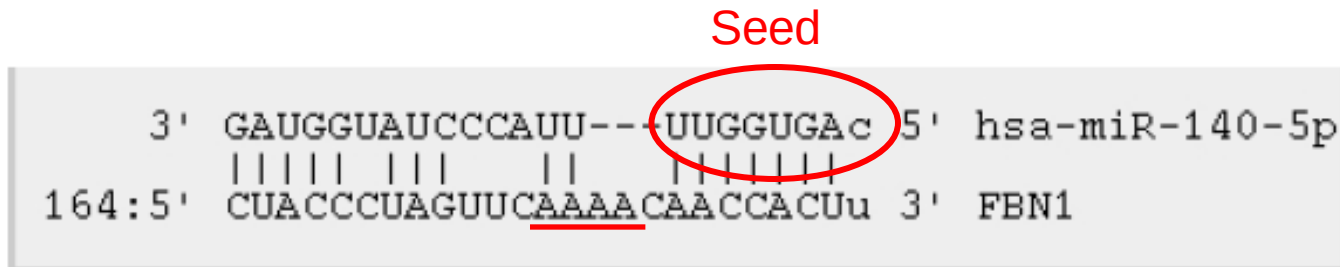


- Aber woher wissen wir, ob eine bestimmte mRNA Ziel einer mikroRNA darstellt? Ideen?
- Vorhersagealgorithmen für microRNA-Targets **optimieren die Sequenzkomplementarität zwischen einem Teil der microRNA und einem Teil der Zielsequenz**
- Was sind die Regeln für „matches“ zwischen mRNA und Ziel-mRNAs? Sie sind von vielen Experimenten abgezogen worden..

Grundlagen der Vorhersage von miRNA-Zielgenen: Das 5'-Ende



- 5'-Ende der miRNA: 'Saat' (seed)
- 6–8 Nukleotide, (fast)perfekte Basenpaarung mit Ziel
- Das 5'-Ende der miRNA ist typischerweise ungepaart, oder beginnt mit Uracil und enthält i.d.R. keine G:U-Wobble-Paare
- Die Bindungsstelle im Ziel ist oft von Adenosinresten flankiert.

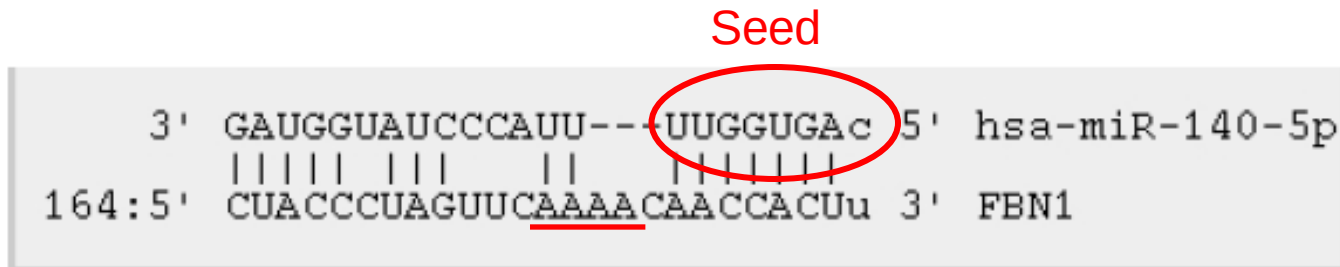


Bindungsstelle für miRNA-140 in *FBN1*

Grundlagen der Vorhersage von miRNA-Zielgenen: Das 3'-Ende

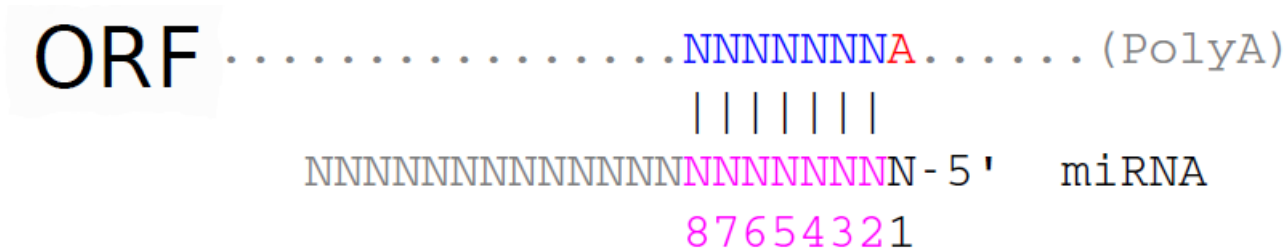


- Perfekte Basenpaarung am 3'-Ende kann eine imperfekte Basenpaarung am 5'-Ende kompensieren (**5' – 3' Asymmetrie**)
- Viele miRNA-Bindungsstellen haben einen 'Buckel' (Bulge) im mittleren Bereich



Bindungsstelle für miRNA-140 in *FBN1*

- **8mer:** Genaue Übereinstimmung der Nukelotide 2-8 der reifen miRNA (Seed + Position 8) gefolgt von einem A
- die stärkste Bindung



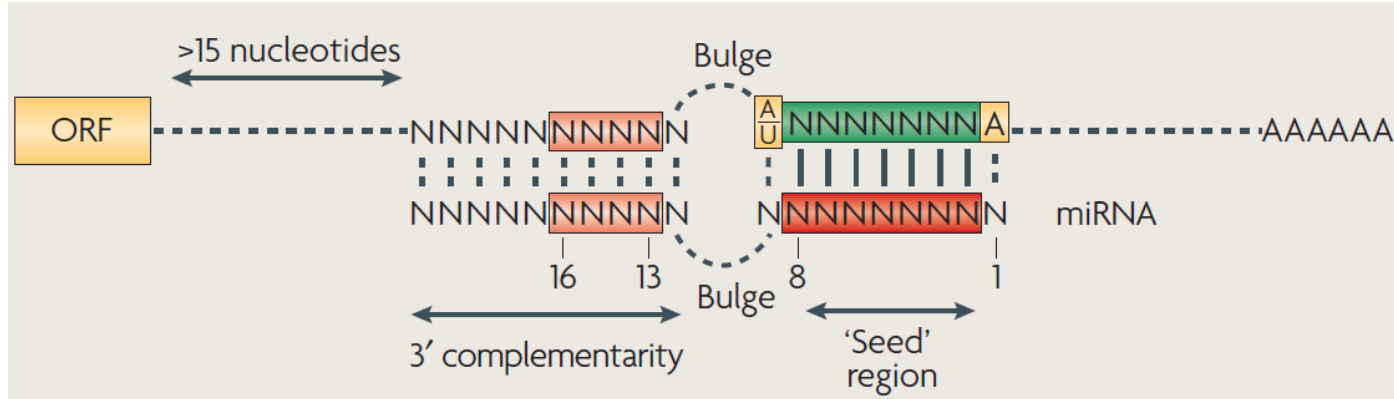
- ORFNNNNNNA.....(PolyA)
 |||||
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5' miRNA
 87654321

- ORF NNNNNN (PolyA)
 |||||
NNNNNNNNNNNNNNNNNNNNNNNNNNNN-5' miRNA
 87654321

Klassifikation der miRNA-Bindungsstellen (nur zum Info)



- Fehlpaarungen (“bulge”) im mittleren Abschnitt des miRNA–mRNA Duplexes
- teilweise Komplementarität des 3′ Bereiches der miRNA, insbesondere Nukleotide 13–16





Aber..Ist es nur die Sequenzkomplementarität?

- Reicht es zur Identifikation von miRNA-regulatorischen Elemente (MREs) nach Sequenzkomplementarität zu suchen?
- Überschlagsrechnung
 - typische 3 UTR hat eine Länge von 700 nt
 - ca. 20.000 proteinkodierende Gene (nehmen wir hier 1 Transkript pro Gen an), daher 20.000x700nt 3UTR Sequenz im Genom
 - kürzeste MRE-Sequenz umfasst 6 nt, d.h., und Wir haben $4^6 = 4096$ möglichkeiten
 - Für eine bestimmte miRNA erwarten wir also $\frac{20000 \cdot 700 \text{nt}}{4096} \approx 3148$ Hits per Zufall
- Man findet also viele MRE-“Kandidaten” per Zufall
- Die Bioinformatik versucht daher auch andere Eigenschaften zu berücksichtigen um die Spezifität der Vorhersagen zu erhöhen.



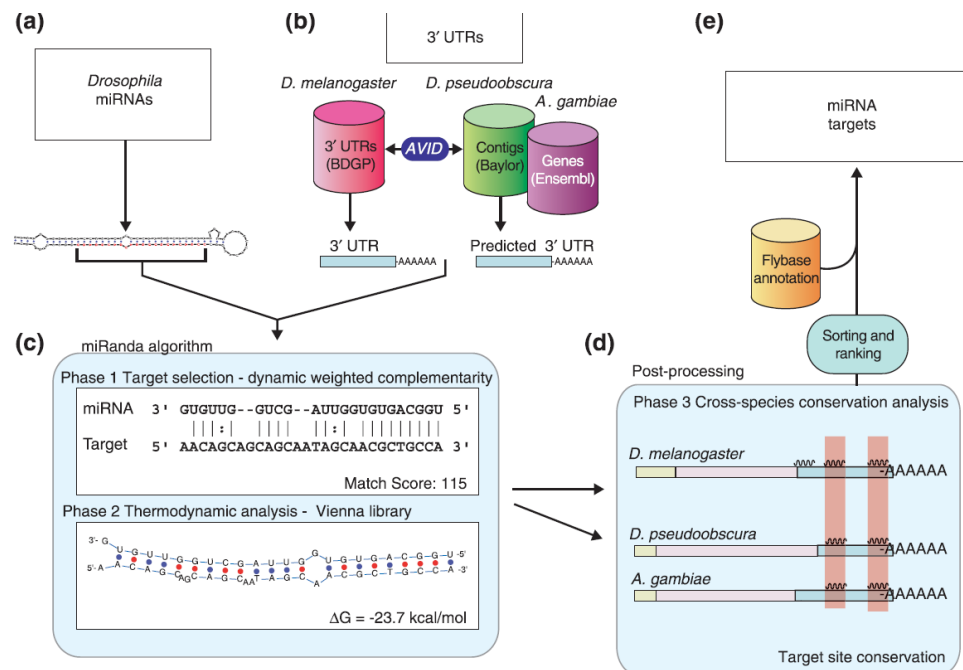
Eine Suche nach miRNA-regulatorischen Elementen (MREs) allein auf Grund einer Sequenz Komplementarität findet zahlreiche falsch positive MREs. Die biologischen Regeln der mikroRNA:mRNA Interaktionen **sind noch nicht vollständig geklärt!**

- Stärke: 8mer > 7mer >> 6mer
- Sequenzkontext (AU reich)
- Sekundärstruktur (“Zugänglichkeit”) -> freie Energie
- Konservierung von MREs
- Bindungen anderer Sequenzabschnitte

miRanda



- Wir werden also eine vereinfachte Version des miRanda-Algorithmus vorstellen
 1. dynamic programming-Alignment (miRNA vs. 3UTR)
 2. thermodynamische Analyse
 3. Konservierung





miRanda: Der miRanda-Algorithmus besitzt eine Ähnlichkeit mit dem Smith-Waterman-Algorithmus. Statt jedoch das Alignment anhand von übereinstimmenden Nukleotiden zu berechnen (z.B. A–A, oder G–G), bewertet miRanda die Komplementarität der Nukleotide (A–U oder G–C) bzw. auch Wobble-Paare (G–U)

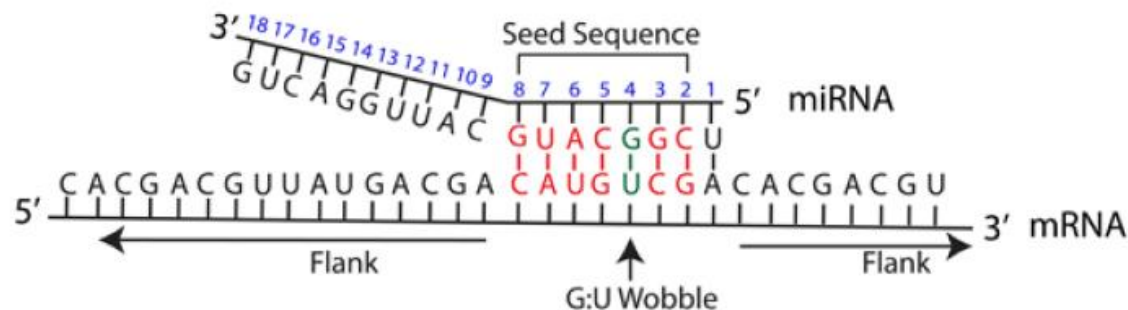
Score-Matrize $S(i,j)$:

- $G \equiv C$: $S(i,j) = +5$
- $A = U$: $S(i,j) = +5$
- $G-U$: $S(i,j) = +2$
- $S(i,j) = -3$ für alle anderen

miRanda: DP



- Affine Gap-penalties (linear in der Anzahl der Nukleotide in der Gap): Gap-Eröffnen: -8, Gap-Verlängern: -2
- Der Score in den ersten 11 Positionen der miRNA wird durch zwei multipliziert, um die besondere Wichtigkeit des 5'-Bereichs der miRNA zu berücksichtigen





miRanda: DP

Der DP-Algorithmus von miRanda funktioniert ähnlich wie bei Smith Waterman indem die Matrix M Schritt für Schritt berechnet wird:

$$S(i, j) = \begin{cases} 0 & \text{Local alignment} \\ S(i-1, j-1) + s(i, j) & \text{match / mismatch} \\ \max_{k \geq 1} \{S(i-k, j) + G_k\} & \text{deletion} \\ \max_{l \geq 1} \{S(i, j-l) + G_l\} & \text{insertion} \end{cases}$$

Hinweis: Es verwendet eine affine Lücke Strafe (an der Tafel)

$$G_k = h + gk$$

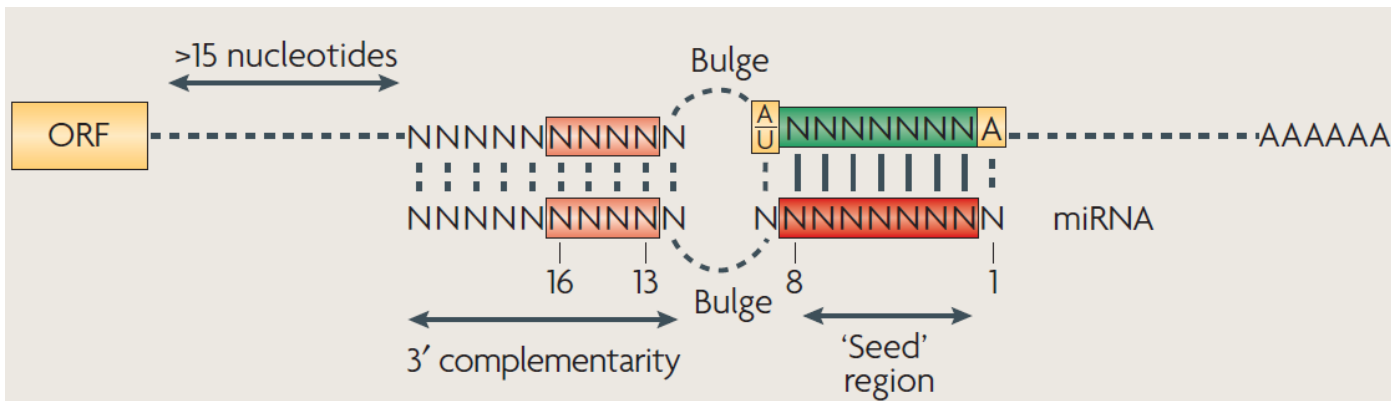
Strafe für die Gap Öffnung

Strafe für die Gap Verlängerung



Zum Schluss werden die folgenden vier empirischen Regeln angewendet

- Keine Mismatches an Position 2–4
- Weniger als 5 Mismatches zwischen Positionen 3–12
- Mindestens ein Mismatch zwischen Positionen 9 und L-5 (wo L: Gesamtlänge des Alignments)
- Weniger als 2 Mismatches in den letzten 5 Positionen des Alignments

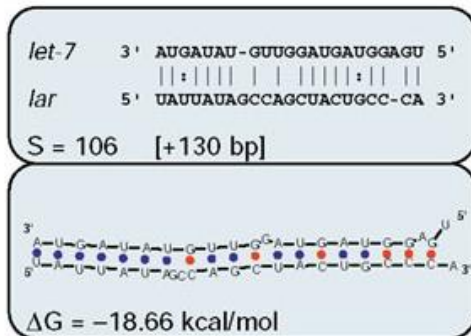


miRanda: freie Energie

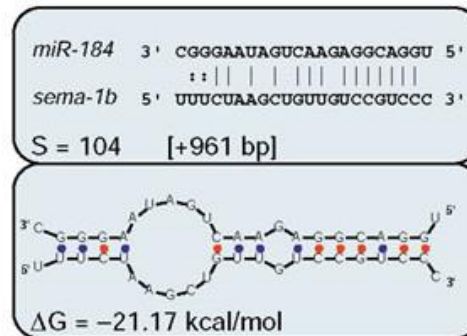


Die Berechnung der freien Energie des miRNA:mRNA-Alignment oder duplex erfolgt im Prinzip wie beim letztes Mal vorgestellten DP Algorithmus (Zuker's algorithmus)

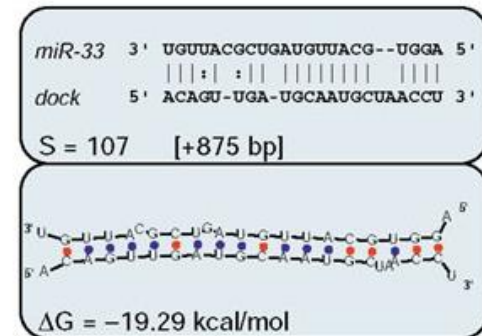
let-7 and lar



miR-184 and sema-1b



miR-33 and dock



Hier sehen wir drei verschiedene microRNAs, die auf drei verschiedene Gene abzielen und drei verschiedene Duplex bilden (mit unterschiedlichen freien Energien deltaG)



- Vienna 1.3 RNA secondary structure programming library (RNAlib)
- Die miRNA-Sequenz und die 3' UTR Sequenz werden zu einer einzelnen Sequenz durch einen 8nt-“Linker” (XXXXXXXX) kombiniert
- Die minimale freie Energie aller Strukturen wird berechnet



- All miRNA-Sequenzen werden dann gegen alle 3' UTR-Sequenzen im Genom verglichen
- Schwellen für Hit-Detektion:
 - Smith-Waterman Alignment $S > 80$
 - Freie Energie des miRNA:mRNA-Duplex $G \leq -14$ kcal/mol.
 - **Hits werden unter evolutionär verwandten Organismen verglichen (Mensch und Maus) oder Sie erhalten ein Konservierungscore von ein MSA**
 - ..und andere empirische Regeln



- Einführung in die Welt der nicht-kodierenden RNAs
 - Insbesondere mikroRNA
 - Molekulare Funktion von microRNAs
- Was sind die Regeln der microRNA-Regulation
- MicroRNA Zielvorhersagealgorithmen
 - Der **miRanda** Algorithmus
 - Wie zu finden oder zu priorisieren microRNA Ziele

Appendix - Local Alignment



		H	E	A	G	A	W	G	H	E	E
P	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	5	0	5	0	0	0	0	0
W	0	0	0	0	2	0	20	12	4	0	0
H	0	10	2	0	0	0	12	18	22	14	6
E	0	2	16	8	0	0	4	10	18	28	20
A	0	0	8	21	13	5	0	4	10	20	27
E	0	0	6	13	18	12	4	0	4	16	26

AWGHE

AW-HE

- Initialisierung: erste Zeile und erste Spalte mit Null initialisieren
- Traceback: Maximalwert von $S(i,j)$ finden. Es kann überall in der Matrix S sein.
- Stoppen wann Ihr zur eine Zelle mit Wert Null kommt