

Bioinformatik-Strategien zur  
Analyse regulatorischer  
RNA-Motive und ihre  
experimentelle Validierung

Thomas Töller

Dipl.Biol. Thomas Töller  
AG Praktische Informatik  
Technische Fakultät  
Universität Bielefeld  
email: ttoeller@techfak.uni-bielefeld.de

Genehmigte Dissertation zur Erlangung des akademischen Grades  
Doktor der Naturwissenschaften (Dr. rer. nat.).  
Von Thomas Töller am 16. Juli 2003  
der Technischen Fakultät an der Universität Bielefeld vorgelegt.  
Am 15. Oktober 2003 verteidigt und genehmigt.

Prüfungsausschuß:

Prof. Dr. rer. nat. Robert Giegerich, Universität Bielefeld  
Dr. rer. nat. Thomas Schmitt-John, Universität Bielefeld  
Prof. Dr. rer. nat. Hermann Ragg, Universität Bielefeld  
Dr. rer. nat. Marc Rehmsmeier, Universität Bielefeld

Gedruckt auf alterungsbeständigem Papier nach ISO 9706.

# **Bioinformatik-Strategien zur Analyse regulatorischer RNA-Motive und ihre experimentelle Validierung**

Dissertation zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

der Technischen Fakultät der Universität Bielefeld

vorgelegt von

**Thomas Töller**

Juli 2003



**Ein Optimist ist in der Regel der Zeitgenosse,  
der am ungenügendsten informiert ist.**

(John Boynton Priestley, engl. Schriftsteller 1894 - 1984)



# Danksagung

An erster Stelle danke ich den beiden Betreuern meiner Arbeit, Prof. Dr. Robert Giegerich und PD Dr. Thomas Schmitt-John, daß sie mir die Möglichkeit gegeben haben, diese interdisziplinäre Arbeit anzufertigen. Ihre immer vorhandene Diskussionsbereitschaft sowie die vielfältigen Anregungen waren mir eine große Hilfe.

Ich danke allen Mitgliedern der AG Pi sowie von W7 und des Graduiertenkollegs Bioinformatik für eine außergewöhnlich gute Arbeitsatmosphäre. Einige von ihnen möchte ich jedoch besonders erwähnen:

Dank an Matthias Höchsmann, Carsten Meyer und Dirk Strothmann für die gute Zusammenarbeit bei den Bioinformatik-Projekten. Ohne sie wäre diese Arbeit so sicher nicht möglich gewesen.

Ralf Palmisano danke ich für die Zusammenarbeit bei den FRET Experimenten, Nadine Küpper für die Durchführung der MALDI-Analysen und Melanie Ronsiek sowie Silvia Naus für die allgemeine Unterstützung im Laboralltag.

Ein besonderer Dank gilt meiner Bürokollegin Jomuna Choudhuri. Ihr brasilianisches Temperament hat meinen Büroalltag immer wieder abwechslungsreich gestaltet. „Eu te desejo tudo de bom.“

Carsten Drepper und Dr. Marc Rehmsmeier danke ich für zahlreiche interessante und spannende Diskussionen.

<http://www.VielenDankAnJoern@techfak.uni-bielefeld.de>. Ohne die vielen nützlichen L<sup>A</sup>T<sub>E</sub>X-Ratschläge von Jörn Clausen wäre diese Arbeit wohl immer noch nicht zu Papier gebracht.

Meinen Laborkolleginnen und Kollegen Anja Fritzenkötter, Sonja Fuchs und Michael Ulbrich sei für eine äußerst positive Stimmung im Labor gedankt. Besonders bedanke ich mich bei Anke Mußmann für die vielen, in der Regel nicht-wissenschaftlichen Gespräche, die eine willkommene Ablenkung von so manchem fehlgeschlagenen Experiment waren.

Ohne Übertreibung kann ich sagen, daß mir die ständige Hilfsbereitschaft von Alex Sczyrba und seine Geduld bei der Beantwortung sämtlicher Fragen die Arbeit am Rechner enorm erleichtert hat. Dank auch für das Korrekturlesen dieser Arbeit.

Zu guter letzt danke ich meiner Familie für ihre immer vorhandene Unterstützung, die vieles einfacher gemacht hat.





# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>vii</b>
<b>Abbildungsverzeichnis</b>	<b>xi</b>
<b>Tabellenverzeichnis</b>	<b>xiii</b>
<b>1. Zusammenfassung</b>	<b>1</b>
<b>2. Einleitung</b>	<b>3</b>
2.1. Genregulation . . . . .	3
2.2. Transkriptionelle Genregulation . . . . .	3
2.3. Posttranskriptionelle Genregulation . . . . .	4
2.3.1. Posttranskriptionelle Genregulation im Nukleus . . . . .	4
2.3.2. Posttranskriptionelle Genregulation im Cytoplasma . . . . .	6
2.4. Untranslatierte Regionen und posttranskriptionelle Genregulation	7
2.4.1. Posttranskriptionelle Genregulation durch micro-RNAs . . . . .	10
2.4.2. Iron Responsive Elements (IREs) . . . . .	11
2.5. RNA-Sekundärstrukturvorhersage . . . . .	13
2.5.1. RNA-Sekundärstrukturvorhersage durch vergleichende Sequenzanalyse . . . . .	13
2.5.2. Thermodynamische RNA-Sekundärstrukturvorhersage . . . . .	14
2.6. Computergestützte Analyse regulatorischer RNA-Motive . . . . .	15
2.7. Ziel dieser Arbeit . . . . .	16
<b>3. Material und Methoden</b>	<b>17</b>
3.1. Material . . . . .	17
3.1.1. Mäusestämme . . . . .	17
3.1.2. Zelllinien . . . . .	17
3.1.3. Verwendete Chemikalien . . . . .	17
3.1.4. Allgemeines zu Puffer und Lösungen . . . . .	17
3.1.5. DNA/RNA Oligonukleotide . . . . .	19
3.2. Gelelektrophoretische Methoden . . . . .	19
3.2.1. nicht-denaturierende PAA-Gelelektrophorese . . . . .	19
3.2.2. Proteingelelektrophorese nach Laemmli . . . . .	20

## INHALTSVERZEICHNIS

3.2.3.	Coomassie-Färbung von Proteingelen . . . . .	20
3.3.	Molekularbiologische und biochemische Methoden . . . . .	21
3.3.1.	Herstellung von Proteinextrakten . . . . .	21
3.3.2.	Verzögerungsgelelektrophorese . . . . .	21
3.3.3.	Northern-Transfer . . . . .	22
3.3.4.	Chemolumineszenz-Nachweis Digoxigenin- oder Biotin- markierter Oligonukleotide . . . . .	22
3.3.5.	Autoradiographie . . . . .	25
3.3.6.	mRNA Detektion durch FRET . . . . .	25
	<i>in vivo</i> Hybridisierung . . . . .	26
	konfokale Laser-Scanning-Mikroskopie . . . . .	27
3.3.7.	Proteinaufreinigung . . . . .	28
	RNA-Oligonukleotidbindung an Dynabeads . . . . .	29
	Proteinaufreinigung mit Dynabeads . . . . .	30
3.3.8.	Proteinidentifikation . . . . .	30
	MALDI-TOF Spektrometrie . . . . .	31
3.4.	Informatikmethoden . . . . .	31
3.4.1.	Datenbanken . . . . .	31
	EMBL . . . . .	31
	UTRdb und UTRsite . . . . .	32
3.4.2.	Sequenzalignment . . . . .	33
	DIALIGN2 . . . . .	33
	ClustalW . . . . .	33
3.4.3.	RNA-Sekundärstrukturvorhersage . . . . .	33
	mfold . . . . .	33
	RNAfold . . . . .	34
	RNAsubopt . . . . .	34
3.4.4.	RNA-Strukturalignment . . . . .	34
3.4.5.	Musterdefinition durch Signifikanzberechnungen . . . . .	37
3.4.6.	Motiv-Beschreibung und Suche . . . . .	37
	RNAMotif . . . . .	37
	HyPa . . . . .	38
<b>4.</b>	<b>Ergebnisse</b> . . . . .	<b>39</b>
4.1.	Strategieentwicklung . . . . .	39
4.1.1.	Testsysteme . . . . .	39
	Testsystem 1: <i>Iron Responsive Element</i> . . . . .	39
	Testsystem 2: modifizierte IREs und UTRs . . . . .	40
	Testsystem 3: Histon Stemloop . . . . .	41
4.2.	Strategien zur Vorhersage von RNA-Strukturmotiven . . . . .	41
4.2.1.	Strategie 1: Sequenzalignment mit nachfolgender Struktur- vorhersage . . . . .	41
	Sequenzalignment . . . . .	41

Strukturvorhersage . . . . .	43
Anwendung der 1. Vorhersagestrategie auf IREs . . . . .	43
Strategiebewertung . . . . .	49
4.2.2. Strategie 2: Strukturvorhersage mit nachfolgendem Strukturalignment . . . . .	49
Paarweise lokale Strukturalignments mit RNAforester . . . . .	49
Multiple globale Strukturalignments mit RNAforester . . . . .	59
Strategiebewertung . . . . .	62
4.3. Validierung durch Musterdefinition . . . . .	63
4.3.1. Musterdefinition durch Signifikanzberechnungen . . . . .	63
4.3.2. Musterbeschreibung und Datenbanksuche . . . . .	65
Musterbeschreibung und Datenbanksuche mit RNAMotif . . . . .	65
Musterbeschreibung und Datenbanksuche mit HyPa . . . . .	67
4.4. Strategieanwendung . . . . .	69
4.4.1. Die Rab1A 3'UTR . . . . .	69
Strukturmotiv-Vorhersage . . . . .	72
Musterdefinition . . . . .	80
Datenbanksuche . . . . .	81
4.4.2. Die BMP7 3'UTR . . . . .	84
Strukturmotivvorhersage . . . . .	84
Musterbeschreibung und Datenbanksuche . . . . .	94
4.5. Experimente . . . . .	98
4.5.1. Gelverzögerung . . . . .	98
4.5.2. Proteinisolation und Identifizierung . . . . .	102
4.5.3. <i>in vivo</i> Lokalisationsuntersuchungen . . . . .	109
4.5.4. Zusammenfassung der experimentellen Ergebnisse . . . . .	111
<b>5. Diskussion</b>	<b>113</b>
5.1. Vorhersagestrategien im Vergleich . . . . .	113
5.1.1. Kombination von Sequenzalignment und Strukturvorhersage	114
5.1.2. RNA-Strukturalignment als universelle Methode zur Strukturmotivvorhersage . . . . .	115
5.1.3. Vergleiche mit anderen Vorhersagemethoden . . . . .	117
Assoziierung von Sequenzmotiven mit funktionellen Genklassen . . . . .	117
"Genetische Programmierung" zur RNA-Strukturvorhersage	118
Optimierung von Sequenz und Struktur: der FOLDALIGN Ansatz . . . . .	119
5.2. Datenbanksuche zur Analyse regulatorischer RNA-Motive . . . . .	120
5.2.1. RNAMotif und HyPa im Vergleich . . . . .	120
5.3. Alternativ gespleißte UTRs: die BMP7 3'UTR . . . . .	122
5.4. Die Rab1A 3'UTR . . . . .	124
5.4.1. Die Vorhersage eines potentiell regulatorischen Stemloops . . . . .	124

## INHALTSVERZEICHNIS

Das Rab1A Stemloop-Suchmuster und die Datenbanksuche	125
5.4.2. Die biologische Bedeutung des Stemloops . . . . .	126
Diskussion der Gelverzögerungsexperimente . . . . .	127
Diskussion der Proteinisolierungsexperimente . . . . .	128
Generelle Bewertung der MALDI-TOF Analysen . . . . .	128
Funktionelle Bedeutung der identifizierten Proteine . . . . .	129
5.4.3. Der Rab1A Stemloop in der Sirtuin 3'UTR . . . . .	131
5.5. Ausblick . . . . .	133
<b>Literaturverzeichnis</b>	<b>135</b>
<b>A. Sequenzen</b>	<b>147</b>
Histon 3'UTRs . . . . .	147
Ferritin 5'UTRs . . . . .	147
modifizierte Ferritin 5'UTRs . . . . .	148
Rab1 3'UTRs . . . . .	149
BMP7 3'UTRs . . . . .	149
weitere UTRs . . . . .	150
<b>B. Abkürzungsverzeichnis</b>	<b>153</b>

# Abbildungsverzeichnis

2.1. Aufbau einer eukaryotischen mRNA . . . . .	8
2.2. mRNA-Lokalisation: <i>cis</i> -regulatorische Elemente in UTRs . . . . .	9
2.3. Posttranskriptionelle Regulation der Ferritin und Transferrinrezeptor mRNAs . . . . .	12
2.4. Eukaryotische Konsensusstruktur des <i>Iron Responsive Element</i> . . . . .	13
3.1. Prinzip der Gelverzögerung. . . . .	23
3.2. Prinzip des Fluoreszenz Resonanz Energie Transfers . . . . .	26
3.3. Prinzip der Proteinisolierung mit Dynabeads . . . . .	28
3.4. Sekundärstruktur der <i>E.coli</i> tRNA für Alanin. . . . .	35
3.5. Baum-Repräsentation der <i>E.coli</i> tRNA Sekundärstruktur . . . . .	35
4.1. Eukaryotische Konsensusstruktur des <i>Iron Responsive Element</i> . . . . .	40
4.2. Konsensusstruktur des Histon Stemloops . . . . .	41
4.3. Strukturvorhersage eines Bereichs der Histon 3'UTR . . . . .	43
4.4. Strukturvorhersage eines Bereichs der Ferritin 5'UTR . . . . .	45
4.5. Strukturvorhersage für die Ferritin 5'UTR des Schweins . . . . .	46
4.6. Strukturvorhersage für die Ferritin 5'UTR des Schweins, mit IRE . . . . .	47
4.7. Sekundärstrukturen der <i>E.coli</i> tRNAs für Alanin und Leucin . . . . .	50
4.8. Globales Strukturalignment der Alanin und Leucin tRNAs . . . . .	51
4.9. Strukturvorhersage für die Ferritin 5'UTR des Menschen . . . . .	52
4.10. Strukturvorhersage für die Succinatdehydrogenase 5'UTR von <i>Drosophila melanogaster</i> . . . . .	53
4.11. Globales Strukturalignment der Ferritin und Succinatdehydrogenase UTRs . . . . .	54
4.12. Lokales Strukturalignment der Ferritin und Succinatdehydrogenase UTRs . . . . .	55
4.13. Strukturvorhersage für die Transferrinrezeptor 3'UTR . . . . .	56
4.14. Lokales Strukturalignment der Ferritin 5'UTR und der Transferrinrezeptor 3'UTR . . . . .	57
4.15. Sekundärstrukturen von zwei modifizierten UTR Sequenzen . . . . .	58
4.16. Lokales Strukturalignment der modifizierten UTR Sequenzen . . . . .	58
4.17. Multiples Strukturalignment der 20 tRNAs aus <i>E.coli</i> . . . . .	60

## ABBILDUNGSVERZEICHNIS

4.18. Alternative Darstellung des multiplen Strukturalignments der 20 tRNAs . . . . .	60
4.19. Multiples globales Strukturalignment von Ferritin und SLC11A3 UTRs . . . . .	61
4.20. Lokalisation und Funktion der Rab Proteine . . . . .	70
4.21. ClustalW Alignment von Rab1A 3'UTRs . . . . .	71
4.22. Strukturvorhersagen für sequenzähnliche Bereiche der Rab1A 3'UTR . . . . .	74
4.23. Strukturvorhersage für die Rab1A 3'UTR des Menschen . . . . .	75
4.24. Strukturvorhersage für die Rab1A 3'UTR des Zitterrochen . . . . .	76
4.25. Lokales Strukturalignment der Rab1A 3'UTRs . . . . .	77
4.26. Basenpaarwahrscheinlichkeiten der Rab1A 3'UTR des Menschen . . . . .	78
4.27. Basenpaarwahrscheinlichkeiten der Rab1A 3'UTR des Zitterrochen . . . . .	79
4.28. Muster des Rab1A Stemloops . . . . .	80
4.29. Multiples Strukturalignment von suboptimalen Strukturen der BMP7 3'UTR (Variante 1) . . . . .	88
4.30. Multiples Strukturalignment von suboptimalen Strukturen der BMP7 3'UTR (Variante 2) . . . . .	89
4.31. Basenpaarwahrscheinlichkeiten der BMP7 3'UTR (Variante 1) . . . . .	90
4.32. Basenpaarwahrscheinlichkeiten der BMP7 3'UTR (Variante 2) . . . . .	91
4.33. Basenpaarwahrscheinlichkeiten der BMP7 3'UTR (UTRdb) . . . . .	92
4.34. Globales paarweises Strukturalignment der BMP7 3'UTRs (Varianten 1 und 2) . . . . .	93
4.35. Muster des verzweigten BMP7 Stemloops . . . . .	94
4.36. RNA-Oligonukleotide für Gelverzögerungsexperimente . . . . .	99
4.37. Autoradiographie des ersten Gelverzögerungsexperimentes . . . . .	100
4.38. Autoradiographie des zweiten Gelverzögerungsexperimentes . . . . .	101
4.39. SDS-Proteingel mit Coomassiefärbung . . . . .	102
4.40. MALDI-Spektrum 1 . . . . .	105
4.41. MALDI-Spektrum 2 . . . . .	106
4.42. MALDI-Spektrum 3 . . . . .	107
4.43. MALDI-Spektrum 4 . . . . .	108
4.44. FRET-Experiment: Fluorescein Kontrolle . . . . .	109
4.45. FRET-Experiment: Texas Red Kontrolle . . . . .	110
4.46. FRET-Experiment: Förster-Transfer . . . . .	111
5.1. Vorhergesagter Stemloop in der Rab1A 3'UTR . . . . .	125
5.2. Stemloop in der Sir2 $\alpha$ 3'UTR . . . . .	127

# Tabellenverzeichnis

3.1. Bezugsquellenverzeichnis der Chemikalien . . . . .	18
3.2. Verwendete DNA/RNA Oligonukleotide . . . . .	19
4.1. Größe und Basenzusammensetzung der UTRdb . . . . .	64
4.2. Ergebnis der Datenbanksuche für das IRE Muster . . . . .	67
4.3. Vergleich von RNAMotif und HyPa für die Mustersuche nach dem IRE Muster . . . . .	68
4.4. Ergebnis der Datenbanksuche für das Rab1 Muster . . . . .	82
4.5. Vergleich von RNAMotif und HyPa für die Mustersuche nach dem Rab1 Muster . . . . .	84
4.6. Ergebnis der Datenbanksuche für das BMP7 Muster . . . . .	96
4.7. Vergleich von RNAMotif und HyPa für die Mustersuche nach dem BMP7 Muster . . . . .	96

*TABELLENVERZEICHNIS*



# 1. Zusammenfassung

Regulatorische Motive in untranslatierten Regionen spielen eine entscheidende Rolle bei der posttranskriptionellen Genregulation von mRNAs. In dieser Arbeit wurden bioinformatische Strategien zur Vorhersage und Validierung potentiell regulatorischer RNA-Motive entwickelt. Die Berechnung globaler und lokaler Strukturalignments mit dem Programm `RNAforester` führte zur Entwicklung einer sehr flexiblen Strategie zur Vorhersage regulatorischer RNA-Motive. Diese Strategie ermöglichte die Detektion von RNA-Motiven unabhängig von ihrer Sequenz und Position innerhalb verschiedener UTRs. Im Vergleich zu anderen Vorhersagemethoden konnten Motive auch in sehr großen Sekundärstrukturen gefunden werden.

Für die Validierung vorhergesagter Motive wurden Datenbanksuchen durchgeführt. Dazu wurden ausgehend von den Vorhersagen Muster definiert, die allgemein genug sein mußten, um möglichst viele Vorkommen des Musters zu finden, die aber auch spezifisch genug sein mußten, um möglichst wenige falsch positive Treffer zu liefern. Um Muster mit solchen Eigenschaften definieren zu können, wurde ein Verfahren zur Musterevaluation mit Hilfe von Signifikanzberechnungen eingesetzt. Für die Datenbanksuche wurden die Programme `RNAMotif` und `HyPa` verwendet, die speziell für die Suche nach kombinierten RNA Sequenz- und Strukturmustern geeignet sind. Darüber hinaus wurden die beiden Programme in dieser Arbeit vergleichend analysiert.

Die Anwendung der entwickelten Strategien hat zur Vorhersage eines potentiell regulatorischen Stemloops in der Rab1A 3'UTR von Mensch und Zitterrochen geführt. Die Datenbanksuche hat ergeben, daß dieser Stemloop in den Rab1A 3'UTRs verschiedener Vertebraten und der Sirtuin 2 $\alpha$  3'UTR von Mensch und Maus vorkommt. Neben diesem Motiv konnte auch eine verzweigte Stemloopstruktur in einer alternativen Spleißvariante der BMP7 3'UTR des Krallenfroschs entdeckt werden.

Experimentelle Analysen im Rahmen dieser Arbeit konnten Hinweise auf eine tatsächliche biologische Funktion des vorhergesagten Stemloops in der Rab1A 3'UTR liefern. Die Ergebnisse von Gelverzögerungs- und Proteinisolierungsexperimenten legten die Interaktion eines größeren Proteinkomplexes mit dem RNA-Stemloop nahe. MALDI-TOF Analysen mehrerer Proteinbanden ergaben, daß ein potentielles RNA-Bindeprotein und der "Rab GDP-dissociation-inhibitor" in dem Proteinkomplex enthalten sein könnten. Die Durchführung von *in vivo* Lokalisationsuntersuchungen mittels FRET-Technologie zeigte eine Lokalisation der Rab1A mRNA an wenigen Positionen innerhalb des Cytoplasmas.

## 1. Zusammenfassung

## 2. Einleitung

### 2.1. Genregulation

Die Komplexität höherer Eukaryonten, wie etwa Pflanzen oder Tiere, erfordert eine aufwendige Regulation der Umsetzung der genetischen Information. Diese Regulation findet sowohl auf transkriptioneller als auch auf posttranskriptioneller Ebene statt. Die transkriptionelle Genregulation umfaßt alle Schritte bis zur Bildung des Primärtranskripts. Die posttranskriptionelle Genregulation setzt nach dessen Bildung an. Zu ihr gehört die Prozessierung des Primärtranskripts. Dazu zählt das Splicing, also die Entfernung der Introns aus dem Primärtranskript, sowie das RNA-Editing, bei dem die Nukleotidabfolge in der mRNA nachträglich geändert werden kann. In einem weiteren posttranskriptionellen Regulationsschritt wird der mRNA-Export aus dem Nukleus kontrolliert. Im Cytoplasma kann es dann zu einer Kontrolle der mRNA-Lokalisation, des mRNA-Abbaus oder auch der Translationseffizienz kommen. Wie später beschrieben wird, spielen die untranslatierten Regionen der mRNAs für diese drei Regulationsereignisse eine entscheidende Rolle.

### 2.2. Transkriptionelle Genregulation

Die Regulation der Transkription spielt eine wichtige Rolle für die spezifische Genexpression und ist in den letzten Jahrzehnten ausführlich untersucht worden. Bei der transkriptionellen Genregulation wird zwischen *cis*- und *trans*-regulatorischen Faktoren unterschieden. Bei den *cis*-regulatorischen Elementen handelt es sich in der Regel um kurze Sequenzabschnitte innerhalb der DNA. Dazu zählen sowohl Promotoren als auch Enhancer und Silencer. Die Transkription wird von zwei Klassen von Transkriptionsfaktoren reguliert. Generelle Transkriptionsfaktoren bilden durch Bindung an den Kern-Promotor, der sich direkt vor dem Transkriptionsstartpunkt befindet, und Assemblierung mit der RNA-Polymerase den sogenannten Präinitiationskomplex (Green, 2000). Zu dieser Klasse von Transkriptionsfaktoren gehören unter anderem die hoch konservierten Faktoren TFIIB, TFIID, TFIIE, TFIIF und TFIIH, die für eine korrekte Transkription *in vivo* allesamt wichtig sind (Holstege et al., 1998). Im Gegensatz zu den generellen Transkriptionsfaktoren binden die sequenzspezifischen Transkriptionsfaktoren an regulatorische DNA-Elemente, die sich in unterschiedlichem Abstand zum Transkriptionsstartpunkt befinden können (Müller, 2001). Diese Faktoren sind ent-

## 2. Einleitung

scheidend für die genspezifische Transkription. Es wird angenommen, daß zum Beispiel in *Drosophila* 5%-7% aller Gene für solche spezifischen Transkriptionsfaktoren codieren (Adams et al., 2000; Rubin et al., 2000). Die Funktionen dieser spezifischen Transkriptionsfaktoren nehmen einen weiten Bereich ein. Manche von ihnen wirken zum Beispiel gewebespezifisch, da sie nur in ganz speziellen Zelltypen exprimiert werden (Keller et al., 1998). Andere wiederum werden zeitabhängig exprimiert, um zum Beispiel den zirkadianen Rhythmus zu bestimmen (Darlington et al., 1998). Wieder andere werden als Reaktion auf Umwelteinflüsse exprimiert. Ein Beispiel hierfür ist der Hitzeschock-Transkriptionsfaktor (Wu, 1995). Neuere Untersuchungen zeigen, daß neben diesen direkten DNA-Transkriptionsfaktor Wechselwirkungen auch Histonmodifikationen zur Transkriptionskontrolle beitragen. So ist zum Beispiel die Acetylierung der Histone ein wichtiger transkriptioneller Regulationsschritt (Berger, 2002).

## 2.3. Posttranskriptionelle Genregulation

Die posttranskriptionelle Genregulation läßt sich sowohl räumlich als auch funktionell in zwei Bereiche gliedern. Beim im Nukleus ablaufenden RNA-Editing sowie RNA-Splicing und hier vor allem durch das alternative Splicing, wird die endgültige genetische Information der mRNA festgelegt. Dadurch wird bestimmt, welches Protein letztendlich von der mRNA codiert wird. Dagegen ändert die posttranskriptionelle Genregulation im Cytoplasma die codierende Information der mRNA nicht mehr. Nachdem durch die Kontrolle des mRNA-Exports aus dem Nukleus reguliert worden ist, welche mRNAs überhaupt in das Cytoplasma gelangen, wird durch die Lokalisation einer mRNA im Cytoplasma sichergestellt, daß das von dieser mRNA codierte Protein nur an einem definierten Ort in der Zelle gebildet wird. Die Regulation der mRNA-Stabilität beeinflusst die Zeitspanne, in der eine mRNA translatiert werden kann und die Translationsregulation wiederum bestimmt, ob das von der mRNA codierte Protein tatsächlich gebildet wird.

### 2.3.1. Posttranskriptionelle Genregulation im Nukleus

Viele eukaryontische Gene sind aus Exons und Introns aufgebaut, wobei zum Beispiel beim Menschen die durchschnittliche Größe eines Exons 150 Nukleotide und die eines Introns 3500 Nukleotide beträgt (Deutsch und Long, 1999). Introns können allerdings bis zu 500000 Nukleotide groß sein (Rowen et al., 2002). Durch einen als Splicing bezeichneten Prozeß werden mittels Spliceosomen die Introns aus dem Primärtranskript entfernt und die Exons zu der sogenannten reifen mRNA verbunden. Spliceosomen sind große Komplexe aus RNAs und Proteinen (Reed, 2000; Stevens et al., 2002), welche die Spleißstellen, die sich an den Grenzen zwischen Exons und Introns befinden, erkennen und anschließend in einem

### 2.3. Posttranskriptionelle Genregulation

mehrstufigen Prozeß die reife mRNA bilden (Staley und Guthrie, 1998). Durch alternatives Splicing ist es möglich, aus einem Primärtranskript viele verschiedene mRNAs zu bilden, die dann für viele unterschiedliche Proteine codieren (Maniatis und Bosiljka, 2002). Alternatives Splicing kann durch Einfügen oder Entfernen von Exons, aber auch durch Verlängern oder Kürzen von Exons erreicht werden (Cartegni et al., 2002). Kontrolliert wird dies durch die Bindung regulatorischer Proteine an die Spleißstellen oder aber an spezielle Regionen (sogenannte “splicing enhancers” oder “splicing silencers”) innerhalb der Exons oder Introns (Cartegni et al., 2002; Graveley, 2000; Hastings und Krainer, 2001). Das alternative Splicing gilt als Hauptgrund für die Vergrößerung des Proteoms in Metazoen.

Ein weiterer Prozeß, bei dem nachträglich die genetische Information des Primärtranskriptes geändert wird, ist das sogenannte RNA-Editing. Man unterscheidet beim RNA-Editing zwischen der Insertion beziehungsweise Deletion und der Änderung von Nukleotiden, wobei letzteres bei höheren Eukaryonten am häufigsten auftritt (Keegan et al., 2002). Ein gut untersuchtes Beispiel bei Säugetieren ist das Apolipoprotein B, das in einer langen und einer kürzeren Form vorkommt. Der Unterschied der beiden Formen wird durch die Umwandlung eines Cytosins in ein Uracil erreicht, wodurch in der mRNA ein Stop-Codon erzeugt wird, das für die verkürzte Form des Apolipoproteins B verantwortlich ist (Powell et al., 1987). Katalysiert werden die Nukleotidmodifikationen durch Deaminasen, die entweder Cytosin in Uracil oder Adenin in Inosin umwandeln (Keegan et al., 2002). Der Prozeß des RNA-Editing tritt zeitlich meistens vor dem RNA-Splicing auf und kann somit auch Auswirkungen auf dieses haben, indem zum Beispiel durch das RNA-Editing eine neue Spleißstelle entsteht (Rueter et al., 1999). RNA-Editing ist nicht auf den codierenden Bereich einer mRNA beschränkt, sondern kann auch die UTRs betreffen (Morse und Bass, 1999), was zum Beispiel Auswirkungen auf die mRNA Stabilität haben könnte.

Nachdem die reife mRNA fertiggestellt worden ist, muß sie aus dem Kern in das Cytoplasma transportiert werden, um dort translatiert werden zu können. Der Kernexport der mRNA stellt den letzten posttranskriptionellen Genregulationsschritt innerhalb des Nukleus dar. Dabei muß die Exportmaschinerie zwischen noch nicht vollständig prozessierten und reifen mRNAs unterscheiden. Die große Mehrheit der mRNAs wird durch einen allgemeinen mRNA-Exportmechanismus aus dem Kern in das Cytoplasma transportiert (Lei und Silver, 2002). Es kann jedoch auch zum spezifischen mRNA-Export kommen, zum Beispiel als Reaktion auf einen Hitzeschock (Saavedra et al., 1997). Darüber hinaus gibt es Beispiele, bei denen der Kernexport an den Prozeß des RNA-Splicing gekoppelt ist (Reed und Hurt, 2002), was ein gutes Beispiel dafür ist, wie die verschiedenen Genregulationsmechanismen miteinander zusammenhängen können.

### 2.3.2. Posttranskriptionelle Genregulation im Cytoplasma

Nachdem die vollständig prozessierten mRNAs aus dem Kern in das Cytoplasma transportiert worden sind, kann es dort zu weiteren Regulationsschritten bei der Umsetzung der in den mRNAs enthaltenen genetischen Informationen kommen. Die Stabilisierung oder Destabilisierung einer mRNA stellt einen wichtigen Kontrollpunkt der Genexpression dar. Sie entscheidet darüber, welche Menge der jeweiligen mRNA für die Translation zur Verfügung steht. Dabei wird die Stabilität einer mRNA durch spezifische Interaktionen zwischen *cis*-regulatorischen Elementen der mRNA und *trans*-regulatorischen Faktoren reguliert, wobei sowohl Proteine als auch RNA als *trans*-regulatorische Faktoren wirken können (Guhaniyogi und Brewer, 2001). Wenn ein Gen stark exprimiert werden muß, kann das zum einen durch eine hohe Transkriptionsrate, aber auch durch eine besondere Stabilität der mRNA erreicht werden, wie zum Beispiel beim  $\beta$ -Globin (Russel et al., 1997). Im Gegensatz dazu sind mRNAs von Genen, die Proteine codieren, welche nur in einem kurzen Zeitraum gebildet werden sollen, oft instabil (Stanton et al., 2000). Es gibt verschiedene Abbauwege für eukaryotische mRNAs, die unterschiedlichen Kontrollmechanismen unterliegen. Der Abbau der meisten polyadenylierten mRNAs erfolgt exonukleolytisch und wird durch eine Verkürzung des poly(A)-Stranges eingeleitet (Korner et al., 1998). Es gibt jedoch auch Beispiele, bei denen der Abbau unabhängig von einer Deadenylierung erfolgt und durch endonukleolytische Schnitte, meist in der 3'UTR, eingeleitet wird (Beelman und Parker, 1995). Abweichungen bei der Kontrolle der mRNA-Stabilität werden mit der Entstehung schwerer Krankheiten, wie zum Beispiel bestimmter Tumorerkrankungen oder Alzheimer, in Verbindung gebracht (Guhaniyogi und Brewer, 2001).

Die Lokalisation von mRNAs an bestimmten Orten in der Zelle ist ein weiterer wichtiger Genregulationsschritt im Cytoplasma. Sie führt dazu, daß das von der lokalisierten mRNA codierte Protein nur an einer ganz bestimmten Position in der Zelle gebildet wird. Das kann zum Beispiel für die richtige Entwicklung eines Organismus entscheidend sein (Driever und Nüsslein-Volhard, 1988). Es gibt verschiedene Mechanismen der mRNA-Lokalisation, die durch die Bindung *trans*-regulatorischer Faktoren an *cis*-regulatorische Elemente, die meistens in der 3'UTR der mRNA liegen, kontrolliert werden (Kloc et al., 2002). Zu diesen Mechanismen gehört der aktive, gerichtete Transport durch Elemente des Cytoskeletts, die ungerichtete cytoplasmatische Diffusion kombiniert mit der Verankerung der mRNA an bestimmten Stellen, der vektorielle Transport aus dem Kern zu einem bestimmten Ziel und die lokale Stabilisierung der zu lokalisierenden mRNA (Kloc et al., 2002; Lipshitz und Smibert, 2000). Die biologischen Funktionen der mRNA Lokalisation sind vielfältig. So ist es zum Beispiel energetisch effizienter, ein Protein, das an einer bestimmten Stelle in der Zelle in großer Anzahl benötigt wird, direkt an diesem Ort zu bilden, als es jedesmal dorthin zu transportieren (Jansen, 2001). Die Lokalisation von mRNAs, die für Proteine codieren, welche nur an ganz

## 2.4. Untranslatierte Regionen und posttranskriptionelle Genregulation

bestimmten Stellen im Cytoplasma auftreten dürfen, stellt eine Art Sicherheitsmechanismus dar. Dabei wird verhindert, daß diese Proteine zu schwerwiegenden Schäden führen (Gavis und Lehmann, 1992).

Die Regulation der Translation ist der dritte wichtige posttranskriptionelle Kontrollmechanismus im Cytoplasma. Die Translationsregulation kann mRNA unspezifisch erfolgen, indem zum Beispiel die Menge an Translationsinitiationsfaktoren begrenzt wird (Gingras et al., 1999). Für die spezifische Genregulation ist jedoch die Translationskontrolle bestimmter mRNAs entscheidend. Wie schon die mRNA-Stabilität und Lokalisation, wird auch die Translation durch die Interaktion von *trans*-regulatorischen Faktoren und *cis*-regulatorischen Elementen reguliert. So spielt zum Beispiel die cytoplasmatische Polyadenylierung, die durch die Bindung bestimmter Proteine an ein regulatorisches Element in der 3'UTR ausgelöst wird, eine wichtige Rolle für die Aktivierung der Translation verschiedener mRNAs (Wickens et al., 2000). Die Regulation der Translation kann darüber hinaus zum Beispiel an die Lokalisation einer mRNA gekoppelt sein (Kloc et al., 2002). Das macht deutlich, daß die verschiedenen Mechanismen der posttranskriptionellen Genregulation eng miteinander verbunden sein können.

## 2.4. Untranslatierte Regionen und posttranskriptionelle Genregulation

Eukaryotische mRNAs haben eine dreigeteilte Struktur. Die codierende Region wird von zwei untranslatierten Regionen flankiert. Vor dem Startcodon liegt die 5'UTR, die mit der sogenannten 5'-Kappe beginnt. Hinter dem Stopcodon liegt die 3'UTR, die bei den meisten mRNAs am Ende polyadenyliert ist. Dieser PolyA-Strang enthält in der Regel zwischen 100 und 250 Adenin Reste und wird nach dem Splicing an das letzte Exon angefügt. Die durchschnittliche Länge der 5'UTR ist bei den meisten Organismen ähnlich und liegt zwischen 100 und 200 Nukleotiden (Pesole et al., 2001). Die durchschnittliche Länge der 3'UTR ist in verschiedenen Organismen sehr unterschiedlich und reicht von 200 bei Pflanzen und Pilzen bis zu 800 bei Menschen und anderen Vertebraten, wobei die Länge sowohl der 5'UTR als auch der 3'UTR innerhalb eines Organismus stark variieren kann (Pesole et al., 2001). Es konnte zum Beispiel gezeigt werden, daß ein einziges Nukleotid als 5'UTR in einem *in vitro* System zur Translationsinitiation ausreichen kann (Hughes und Andrews, 1997). Neben der Länge unterscheiden sich 5'UTRs und 3'UTRs auch in ihrer Basenzusammensetzung. Dabei ist der G+C-Gehalt in der 5'UTR höher als in der 3'UTR (Pesole et al., 1997).

Die UTRs spielen eine wichtige Rolle bei der cytoplasmatischen posttranskriptionellen Genregulation. Sowohl die 5'UTR als auch die 3'UTR können regulatorische Motive enthalten, die Auswirkungen auf die mRNA-Lokalisation, mRNA-Stabilität und Translationseffizienz haben können. Abbildung 2.1 zeigt einige Bei-

## 2. Einleitung

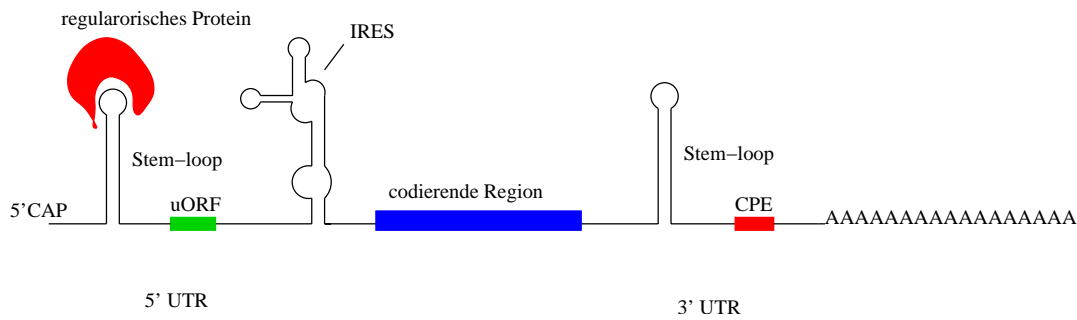


Abbildung 2.1.: Eukaryotische mRNA. Die dargestellten regulatorischen Elemente können die Lokalisation, Stabilität und Translation der mRNA regulieren. uORF: upstream open reading frame; IRES: internal ribosome entry site; CPE: cytoplasmic polyadenylation element.

spiele von verschiedenen regulatorischen Motiven, die in UTRs vorhanden sein können. Im Gegensatz zu regulatorischen Motiven in der DNA, welche reine Sequenzmotive sind, sind die regulatorischen Motive in UTRs häufig eine Kombination aus Sequenz und Struktur (Dandekar und Hentze, 1995).

Die Stabilität einer mRNA wird in der Regel durch *cis*-regulatorische Elemente innerhalb der 3'UTR reguliert. Bei diesen Elementen kann es sich sowohl um Sequenzmotive als auch um Strukturmotive handeln. Ein weit verbreitetes Beispiel für Sequenzmotive sind die sogenannten AU-reichen Elemente (ARE) (Guhaniyogi und Brewer, 2001). Als Sequenzmotiv kommen hier AUUUA Pentamere vor, die sich in den 3'UTRs von mRNAs befinden, deren Abbau durch Bindung spezifischer Proteine an die AREs beschleunigt wird (Brewer, 2000; Carballo et al., 2000). Die 3'UTR der Transferrinrezeptor-mRNA enthält dagegen Stem-loop Strukturen, welche die Stabilität dieser mRNA abhängig vom Eisengehalt in der Zelle regulieren. Diese sogenannten *Iron Responsive Elements* (IRE) werden in Abschnitt 2.4.2 detailliert beschrieben.

So wie die mRNA-Stabilität, wird auch die Lokalisation einer mRNA in erster Linie durch *cis*-regulatorische Elemente, die sich innerhalb der 3'UTR der lokalisierten mRNA befinden, reguliert. Auch hier lassen sich Sequenzmotive von komplexen RNA-Sekundärstrukturmotiven unterscheiden (siehe Abbildung 2.2). Die Lokalisation der Vg1 mRNA von *Xenopus laevis* wird beispielsweise durch die Proteinbindung an kurze, sich wiederholende Sequenzmotive innerhalb der 3'UTR reguliert (Mowry und Melton, 1992). Die bicoid mRNA wird in den Eizellen von *Drosophila melanogaster* lokalisiert, wobei zwischen einer frühen und einer späten Lokalisation unterschieden wird. Die für diese Lokalisationsereignisse verantwortlichen Strukturmotive sind sequentiell innerhalb der bicoid 3'UTR angeordnet (Abbildung 2.2) (Jansen, 2001; Macdonald et al., 1993).

Die Translation wird durch regulatorische Motive sowohl in der 5'UTR als auch in der 3'UTR gesteuert. Eine Regulierung der Translation erfolgt vor allem bei



## 2.4. Untranslatierte Regionen und posttranskriptionelle Genregulation

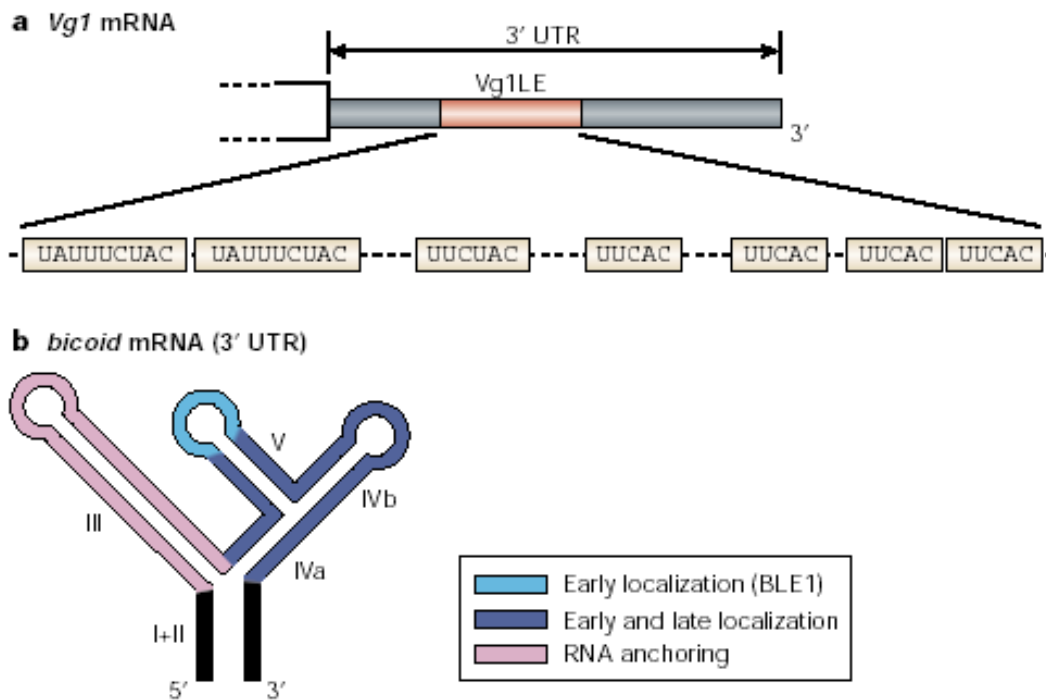


Abbildung 2.2.: *cis*-regulatorische Elemente zur mRNA-Lokalisation. a) Für die Lokalisation der *Vg1* mRNA bei *Xenopus* sind wiederholte Sequenzmotive in der 3'UTR verantwortlich. b) Die *bicoid* 3'UTR von *Drosophila* enthält ein komplexes Strukturmotiv, dessen verschiedene Elemente für unterschiedliche Phasen der Lokalisation von Bedeutung sind (aus (Jansen, 2001)).

## 2. Einleitung

mRNAs, die für Entwicklungsprozesse eine Bedeutung haben, wie zum Beispiel Wachstumsfaktoren und Proto-Onkogene (Mignone et al., 2002). Diese mRNAs haben meistens deutlich längere 5'UTRs als der Durchschnitt (Kozak, 1987). Die Translation beginnt in der Regel am ersten Startcodon in der mRNA. Durch das Vorkommen von kurzen offenen Leserastern in der 5'UTR kann die Produktion des eigentlichen Proteins jedoch verhindert oder verringert werden (Mignone et al., 2002). Innerhalb der 5'UTRs können sehr stabile Sekundärstrukturen (Svitkin et al., 2001) oder Proteine, die spezifisch an Struktur motive binden, die Translation blockieren. Beispiele hierfür findet man im Eisenmetabolismus der Zelle, in Form der bereits erwähnten *Iron Responsive Elements* (Klausner et al., 1993). Bei verschiedenen Entwicklungsgenen von *Caenorhabditis elegans* konnte eine Translationsinhibierung nachgewiesen werden. Diese wurde durch die Bindung kleiner RNA Moleküle an Stellen innerhalb der 3'UTRs der zu regulierenden mRNAs verursacht. Diese Art der posttranskriptionellen Genregulation durch sogenannte micro-RNAs (miRNAs) wird in Abschnitt 2.4.1 beschrieben.

Wie oben bereits erwähnt worden ist, kann die cytoplasmatische Polyadenylierung einer mRNA zur Aktivierung der Translation führen. Um diese Adenylierung zu starten, müssen spezifische Proteine an ein Sequenzmotiv in der 3'UTR, das sogenannte cytoplasmatische Polyadenylierungselement (CPE), binden (Wickens et al., 2000). Einen alternativen Weg zur Translationsinitiation bieten die sogenannten Internal Ribosome Entry Sites (IRES), die als erstes bei Picornaviren entdeckt wurden (Pelletier et al., 1988). Die IRES haben eine komplexe Sekundärstruktur und befinden sich innerhalb der 5'UTR in der Nähe des Startcodons. Sie dienen dazu, mRNAs auch dann translatieren zu können, wenn die normale Translationsmaschinerie der Zelle ausgeschaltet ist.

### 2.4.1. Posttranskriptionelle Genregulation durch micro-RNAs

Die Genregulation durch kleine, ca. 22 Nukleotid große antisense RNAs, die sogenannten micro-RNAs (miRNA), konnte zunächst in der Larvenentwicklung des Nematoden *Caenorhabditis elegans* nachgewiesen werden. In heterochronen Mutanten kommt es dabei zu Störungen beim Übergang der Larvenstadien. Die miRNA lin-4 kontrolliert den Übergang vom Larvenstadium 1 in das Larvenstadium 2. Dabei bindet sie an spezifische Bindestellen in den 3'UTRs ihrer Ziel-mRNAs lin-14 und lin-28, wodurch die Translation dieser mRNAs verhindert wird (Lee et al., 1993). Dabei scheint die Ausbildung größerer und kleinerer Bulges bei der Duplexbildung charakteristisch für die Bindung von miRNAs an ihre Ziel-mRNAs zu sein. Der genaue Mechanismus der Translationsregulation durch lin-4 konnte bisher nicht aufgeklärt werden. Es wird jedoch vermutet, daß die Duplexe der miRNA mit der 3'UTR durch spezifische Proteine stabilisiert werden, und das diese Proteine für die Translationsregulation mitverantwortlich sind (Hu et al., 1996). Mit der miRNA let-7 wurde eine zweite antisense RNA gefunden, die für die Entwicklung von *C. elegans* entscheidend ist. Dabei kontrolliert let-7

## 2.4. Untranslatierte Regionen und posttranskriptionelle Genregulation

den Übergang vom Larvenstadium 4 ins Erwachsenenstadium (Reinhart et al., 2000). Im Gegensatz zu *lin-4* konnten *let-7* Homologe bei allen untersuchten bilateral symmetrischen Tieren nachgewiesen werden (Pasquinelli et al., 2000). Das macht eine Rolle der miRNAs für die posttranskriptionelle Genregulation bei vielen Organismen wahrscheinlich. Ein weiterer Hinweis auf die genregulatorische Bedeutung solch kleiner RNA Moleküle war die Entdeckung zahlreicher neuer potentieller miRNAs (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee und Ambros, 2001; Lagos-Quintana et al., 2002). Die Funktion dieser neu entdeckten miRNAs ist jedoch weitestgehend unbekannt, da es bis jetzt keine zuverlässige Methode zur Vorhersage potentieller Ziel-mRNAs gibt. Ursache hierfür ist das Ausbilden von Bulges bei der Duplexbildung von miRNAs mit ihren Ziel-mRNAs, was einen bioinformatischen Ansatz zur Vorhersage der Ziel-mRNAs erschwert. Eine Ausnahme spielen dabei die miRNAs in Pflanzen, die anscheinend nahezu perfekt komplementär an ihre Ziel-mRNAs binden. So ließen sich in *Arabidopsis thaliana* durch einen Bioinformatikansatz für 14 miRNAs 49 potentielle Ziel-mRNAs vorhersagen (Rhoades et al., 2002). Dabei wurde nach signifikanten Bindestellen mit maximal 3 Fehlern in mRNAs gesucht. In Tieren war dieser Ansatz jedoch nicht erfolgreich, so daß die Vorhersage potentieller Ziel-mRNAs in Tieren eine große Herausforderung in der miRNA Forschung ist.

### 2.4.2. Iron Responsive Elements (IREs)

Ein besonders gut untersuchtes Beispiel für die Bedeutung der untranslatierten Regionen einer mRNA für die posttranskriptionelle Genregulation findet sich im Eisenmetabolismus vieler eukaryotischer Organismen. Die Zellen benötigen Eisen für die verschiedensten Aufgaben, zum Beispiel als Co-Faktor für Enzyme. Bei Säugetieren dienen Transferrinrezeptoren zur Eisenaufnahme in die Zelle (Harford et al., 1990) und Ferritin fungiert als Eisenspeichermolekül. Der Eisenmetabolismus der Zelle basiert auf einer koordinierten posttranskriptionellen Genregulation dieser beiden Moleküle (Harford und Klausner, 1990). Diese Genregulation erfolgt durch die spezifische Bindung regulatorischer Proteine (*Iron Regulatory Proteins*) an spezielle regulatorische Motive, die sogenannten *Iron Responsive Elements* in den UTRs dieser mRNAs. In der 5'UTR der mRNAs für die große und kleine Ferritin-Untereinheit befindet sich jeweils ein IRE in der Nähe des 5'-Endes der mRNA. Diese mRNAs werden auf der Translationsebene reguliert, da es bei Eisenmangel zu einer Bindung des regulatorischen Proteins an das IRE kommt, wodurch die Translation der mRNA verhindert wird (Hentze et al., 1987; Klausner et al., 1993). Eisenmangel führt ebenfalls zu der Proteinbindung an die IREs in der Transferrinrezeptor mRNA, wobei sich fünf IREs in der 3'UTR dieser mRNA befinden (Casey et al., 1988). Hier wird die mRNA durch diese Protein-RNA Interaktion jedoch stabilisiert, so daß bei Eisenmangel in der Zelle vermehrt Eisen aufgenommen werden kann (Casey et al., 1989). Diese gegensätzliche Regulation der Ferritin und Transferrinrezeptor mRNAs bei Eisenmangel ist in Abbildung

## 2. Einleitung

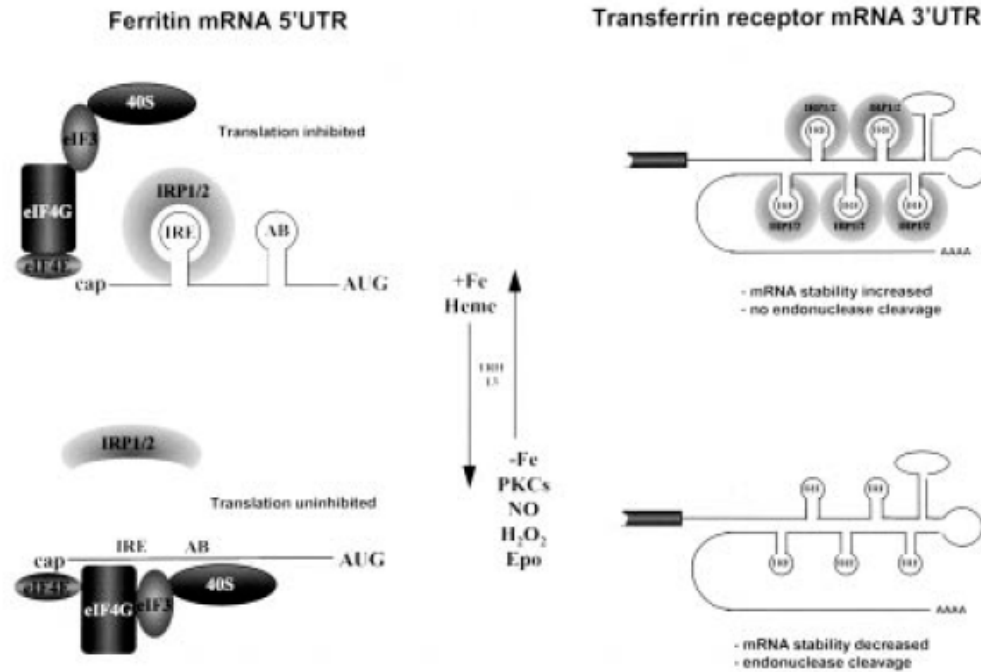


Abbildung 2.3.: Posttranskriptionelle Regulation der Ferritin (links) und Transferrinrezeptor (rechts) mRNAs. Die Bindung der regulatorischen Proteine IRP1 oder IRP2 an das IRE verhindert bei Eisenmangel die Bindung des Translationsinitiationskomplexes und damit die Translation der Ferritin mRNA. Die Bindung der regulatorischen Proteine an die 5 IREs in der Transferrinrezeptor 3'UTR blockiert endonukleolytische Schnittstellen und erhöht dadurch die mRNA-Stabilität (aus (Thomson et al., 1999)).

2.3 noch einmal ausführlich dargestellt.

*Iron Responsive Elements* haben einen charakteristischen Aufbau. Sie bestehen aus einem Hairpin-loop, der bei den meisten IREs eine konservierte Sequenz besitzt (CAGUGH). Die obere Helix besteht in der Regel aus fünf Basenpaaren, kann jedoch in der Länge variieren. Der aus einem Cytosin bestehende Bulge-loop kann zu einem Internalloop erweitert sein (z.B. beim IRE in der Ferritin 5'UTR des Menschen). Die Länge der unteren Helix ist variabel. Mutationen in den sequenzkonservierten Bereichen des IREs können zu schweren Erkrankungen führen (Kato et al., 2001). Aus bekannten IREs verschiedener Organismen lässt sich ein IRE-Konsensus ableiten (s. Abbildung 2.4) (Klausner et al., 1993). Die in diesem Konsensus erkennbare Kombination aus Sequenz- und Struktureigenschaften kommt bei regulatorischen Motiven in UTRs relativ häufig vor. Dies führt zu speziellen Anforderungen für die Untersuchung solcher regulatorischer Signale.

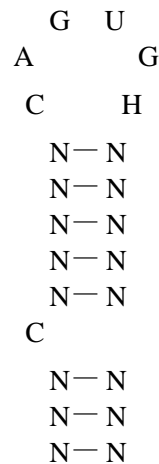


Abbildung 2.4.: *Iron Responsive Element*: eukaryotische Konsensusstruktur (H: A,C,U). Der Cytosin-Bulge kann zu einem Internalloop erweitert sein (z.B. Ferritin 5'UTR des Menschen).

## 2.5. RNA-Sekundärstrukturvorhersage

RNA Moleküle bilden charakteristische Sekundärstrukturen aus. Für die Nucleotide in einer RNA-Sekundärstruktur gibt es zwei Zustände. Entweder sie sind Teil einer Basenpaarung, oder sie sind ungepaart. Basenpaarungen in RNA-Sekundärstrukturen werden meistens durch die kanonischen Paare A-T und G-C sowie das nichtkanonische G-U Paar gebildet. Es gibt jedoch weitere Möglichkeiten, Basenpaarungen zu bilden. Beispiele hierfür findet man in dem SECIS Element. Dieses kommt in den 3'UTRs verschiedener mRNAs vor, die für Proteine kodieren, welche die Aminosäure Selenocystein enthalten (Walczak et al., 1998). Da die Sekundärstruktur für die Funktion vieler RNA Moleküle entscheidend ist, haben RNAs häufiger eine konservierte Sekundärstruktur als eine konservierte Sequenz (Gutell, 1994; Chen et al., 2000). Für die Vorhersage von RNA-Sekundärstrukturen gibt es verschiedene Verfahren.

### 2.5.1. RNA-Sekundärstrukturvorhersage durch vergleichende Sequenzanalyse

Die vergleichende Sequenzanalyse ist eine der verlässlichsten Methoden zur RNA Sekundärstrukturvorhersage (Pace et al., 1999). Sie erfordert jedoch das Vorhandensein eines Alignments von einer großen Anzahl an Sequenzen mit derselben Funktion. Der Ansatz basiert darauf, daß Mutationen in basengepaarten Bereichen, die funktionell wichtig sind, durch kompensatorische Mutationen ausgeglichen werden (Parsch et al., 2000). Ein Beispiel für eine solche kompensatorische Mutation wäre der Wechsel von einem G-C zu einem A-U Basenpaar innerhalb

## 2. Einleitung

einer basengepaarten Region in einer Sekundärstruktur. Der Ansatz eignet sich besonders für die Strukturvorhersage von evolutionär hoch konservierten RNAs wie Transfer-RNAs (Dirheimer et al., 1995) oder ribosomalen RNAs (Woese und Pace, 1993).

### 2.5.2. Thermodynamische RNA-Sekundärstrukturvorhersage

Im Gegensatz zur RNA-Sekundärstrukturvorhersage durch vergleichende Sequenzanalyse ist die thermodynamische Strukturvorhersage möglich, wenn nur eine einzige Sequenz vorhanden ist. Die Methode basiert darauf, daß jeder RNA Sekundärstruktur eine freie Energie  $\Delta G$  zugeordnet werden kann. Diese setzt sich wiederum aus den freien Energien einzelner Elemente einer Sekundärstruktur, wie zum Beispiel einer basengepaarten Region oder einer Loop-Region, zusammen. Diese Energiebeiträge können positiv oder negativ sein. Für viele verschiedene Strukturelemente sind sequenzabhängige Energieparameter experimentell bestimmt worden (Mathews et al., 1999). Während für die Basenpaarstapelwechselwirkungen (das sogenannte "stacking") sowohl ein Enthalpiebeitrag ( $\Delta H$ ) als auch ein Entropiebeitrag ( $\Delta S$ ) berücksichtigt wird, zählt für Loopregionen nur der Entropiebeitrag. Für die temperaturabhängige Berechnung der freien Energie einer basengepaarten Region ( $\Delta G_{bp}$ ) oder einer Loop-Region ( $\Delta G_{loop}$ ) ergeben sich folgende Formeln:

$$\Delta G_{bp} = \Delta H_{37, bp} - T \Delta S_{37, bp}$$

$$\Delta G_{loop} = -T \Delta S_{37, loop}$$

Der nach dem Prinzip der Dynamischen Programmierung (Bellman und Kabala, 1960) entwickelte Algorithmus von Zuker und Stiegler minimiert diese freie Energie, um so die energetisch günstigste Sekundärstruktur zu berechnen (Zuker und Stiegler, 1981). Eine Erweiterung dieses Algorithmus erlaubt die Berechnung suboptimaler Strukturen (Zuker, 1989). Diese Verbesserung ist wichtig, da die energetisch günstigste Struktur nicht die biologisch tatsächlich vorliegende Struktur sein muß.

Die Programme `mfold` von Zuker (Zuker et al., 1999) oder `RNAfold` aus dem Vienna RNA-Paket (Hofacker et al., 1994) ermöglichen die RNA-Strukturvorhersage durch Energieminimierung. Durch die Anwendung neuer Energieparameter konnte die durchschnittliche Anzahl korrekt vorhergesagter Basenpaarungen in der energetisch günstigsten Struktur von 64% auf 73% gesteigert werden (Mathews et al., 1999).

## 2.6. Computergestützte Analyse regulatorischer RNA-Motive

Für die Analyse regulatorischer Motive in der DNA werden häufig Programme eingesetzt, die Motive auf Sequenzebene vorhersagen. Diese Sequenzmotive beeinflussen normalerweise als Bindestellen für regulatorische Proteine die Transkription. Das von Apostolico entwickelte und auf der Datenstruktur des Suffixbaumes basierende Verfahren zur Suche überrepräsentierter Wörter (Apostolico et al., 2000) ist ein Beispiel für ein solches Programm. Andere statistische Methoden zur Suche nach DNA Sequenzmotiven wurden zum Beispiel erfolgreich zur Untersuchung regulatorischer Bereiche von Hefe-Genen eingesetzt (van Helden et al., 2000).

Die zunehmende Bedeutung der posttranskriptionellen Genregulation für das Verständnis verschiedenster biologischer Vorgänge (z.B. Entwicklungsprozesse) und die Entdeckung von regulatorischen Motiven in den untranslatierten Regionen der regulierten mRNAs hat zu verschiedenen computergestützten Strategien zur Vorhersage solcher RNA-Motive geführt. So konnte mit der Untersuchung von statistisch überrepräsentierten Sequenzmotiven in einer Sammlung von menschlichen 3'UTRs zum Beispiel das in Abschnitt 2.4 beschriebene AU-reiche Destabilisierungselement (ARE) gefunden werden (Pesole et al., 1994). In einem anderen Ansatz wurden bestimmte Nukleotidmuster innerhalb von 3'UTRs mit funktionellen Proteinklassen in Zusammenhang gebracht (Conclin et al., 2002). Hier wurde das bereits bekannte Auftreten von ARE-Elementen in den 3'UTRs von Cytokinen bestätigt und einige weitere potentielle regulatorische Sequenzmotive vorhergesagt.

Da die regulatorischen Motive in UTRs aber häufig durch eine bestimmte Sekundärstruktur funktionell definiert sind, eignen sich die beschriebenen Methoden zur Suche nach Sequenzmotiven nur sehr begrenzt für die Vorhersage regulatorischer Motive in UTRs. Methoden, die auch die Sekundärstruktur berücksichtigen, sind hierfür sehr viel besser geeignet. Das Programm FOLDALIGN optimiert gleichzeitig das Sequenzalignment und die Struktur in einer Sammlung mehrerer Sequenzen (Gorodkin et al., 1997). Dadurch ist es zum Beispiel möglich, ein gemeinsames Strukturmotiv in mehreren UTRs zu finden. Das Programm basiert auf den Prinzipien des von Sankoff entwickelten Algorithmus zum optimalen Falten und Alignieren mehrerer RNA Sequenzen (Sankoff, 1985). Dieser Algorithmus hat eine Zeitkomplexität von  $O(L^{3N})$  für N Sequenzen der Länge L und ist damit für mehrere Sequenzen nicht praktikabel. Für FOLDALIGN wurde der Sankoff Algorithmus daher vereinfacht, um eine bessere Zeitkomplexität zu erhalten. Während der Sankoff Algorithmus gleichzeitig die Gesamtkosten des Alignments und die Energie der Strukturen minimiert, maximiert FOLDALIGN einen Score, der auf einer Kombination aus Sequenzähnlichkeit und Struktur basiert. Dabei wird die Energie der Struktur nicht minimiert, sondern die Anzahl der Basenpaare maxi-

## 2. Einleitung

miert (Gorodkin et al., 1997). Da verzweigte Strukturen nicht betrachtet werden, sinkt die Zeitkomplexität auf  $O(L^4)$  für zwei Sequenzen. Um in  $N$  Sequenzen nach einem lokalen gemeinsamen Motiv zu suchen, wurde durch mehrere weitere Vereinfachungen eine Zeitkomplexität von  $O(L^4N^4)$  erreicht, was den Ansatz nur für kurze Sequenzen (bis 300 Nukleotide) praktikabel macht. Durch die Kombination mit dem Programm COVE (Eddy und Durbin, 1994) wurde der Ansatz erweitert (Gorodkin et al., 2001) und praktikabler gestaltet. Die Funktionalität des Ansatzes wurde unter anderem an modifizierten Ferritin 5'UTRs, in denen stark veränderte *Iron Responsive Elements* gefunden werden konnten, dargestellt.

## 2.7. Ziel dieser Arbeit

Im Rahmen dieser Arbeit sollen computergestützte und experimentelle Methoden zur Untersuchung regulatorischer Motive in untranslatierten Regionen von mRNAs entwickelt werden. Es sollen Strategien zur Vorhersage von RNA Strukturmotiven entwickelt und getestet werden. Die bioinformatische Validierung vorhergesagter Struktur motive erfolgt durch Musterdefinitionen und Datenbanksuchen. Bei konkreten mRNAs sollen neue potentiell regulatorische Struktur motive vorhergesagt werden. Die Rab1A 3'UTR wird im Rahmen dieser Arbeit experimentell auf einen möglichen posttranskriptionellen Regulationsmechanismus untersucht.



# 3. Material und Methoden

## 3.1. Material

### 3.1.1. Mausstämme

Für die Präparation von Proteinextrakten wurden verschiedene Gewebe des Mausstammes C57BL/6J verwendet. Dieser wurde ursprünglich am Jackson Laboratory, Bar Harbor, USA etabliert.

### 3.1.2. Zelllinien

Die *in vivo* Lokalisationsuntersuchungen mittels FRET wurden in der humanen Myelomzelllinie K562 durchgeführt.

### 3.1.3. Verwendete Chemikalien

Die Bezugsquellen der verwendeten Chemikalien sind in Tabelle 3.1 aufgeführt. Soweit nicht anders angegeben, wurden Chemikalien mit dem Reinheitsgrad p.a. verwendet.

### 3.1.4. Allgemeines zu Puffer und Lösungen

Sämtliche Lösungen und Puffer wurden mit destilliertem Wasser angesetzt und sofern möglich für 20 Minuten bei 120 °C autoklaviert.

#### Häufig verwendete Puffer und Lösungen:

TBE-Puffer, 10 x    890 mM Tris  
                          890 mM Borsäure  
                          25 mM EDTA

TE-Puffer, 10 x    100 mM Tris-HCl, pH 8,0  
                          10 mM EDTA

### 3. Material und Methoden

<b>Chemikalie</b>	<b>Bezugsquelle</b>
Acrylamid	Roth, Karlsruhe
Anti-Digoxigenin AP	Roche, Mannheim
APS (Ammoniumperoxodisulfat)	Merck, Darmstadt
Borsäure	AppliChem, Darmstadt
Bromphenolblau	Serva, Heidelberg
CDP-Star ready to use	Roche, Mannheim
DIG Wash and Block Buffer Set	Roche, Mannheim
DTT	Sigma, Steinheim
Dynabeads M-280 Streptavidin	Dynal, Hamburg
EDTA	Sigma, Steinheim
Essigsäure	Roth, Karlsruhe
Glutaraldehyd	Sigma, Steinheim
Glycerin	Fluka, Seelze
Glycin	Roth, Karlsruhe
Isopropanol	Roth, Karlsruhe
Mark 12 Proteinstandard	Invitrogen, Karlsruhe
Methanol	Roth, Karlsruhe
Natriumchlorid	Roth, Karlsruhe
Röntgenfilm Entwickler	TETENAL, Norderstedt
Röntgenfilm-Fixierer	TETENAL, Norderstedt
Roti-Blue	Roth, Karlsruhe
Rotifect	Roth, Karlsruhe
RPNi 1640	BioWhittaker Europe, Belgien
Salzsäure	J.T. Baker, Holland
SDS	Roth, Karlsruhe
Streptavidin AP	Roche, Mannheim
TEMED	Serva, Heidelberg
Tris	Sigma, Steinheim
Xylencyanol	Sigma, Steinheim

Tabelle 3.1.: Bezugsquellen der Chemikalien

### 3.1.5. DNA/RNA Oligonukleotide

Die DNA-Oligonukleotide wurden von der Firma MWG Biotech, Ebersberg und die RNA-Oligonukleotide von der Firma Thermo Hybaid, Ulm bezogen. Die DNA-Oligonukleotide wurden im Synthesemaßstab 0,01  $\mu$ Mol und die RNA-Oligonukleotide im Synthesemaßstab 0,02  $\mu$ Mol bestellt. Die im Rahmen dieser Arbeit verwendeten Oligonukleotide sind in Tabelle 3.2 aufgeführt.

Typ	Name	Sequenz (5'-3')	Modifikation
DNA	RAB-Label1-Flu	GTTCTACGTTTCG	5' Fluorescein
DNA	RAB-Label2-Tex	GCCATCGTCATG	3' Texas Red
RNA	Rab_RNA	UACUGUAUGUAGCUGC- ACUACAACAGAUUCUU	
RNA	Rab_RNA_DIG	UACUGUAUGUAGCUGC- ACUACAACAGAUUCUU	5' Digoxigenin
RNA	Rab1A_long	UUGCCUGAAUUGUACUGUAUG- UAGCUGCACUACAACAGAUUC	5' Biotin

Tabelle 3.2.: Verwendete Oligonukleotide

## 3.2. Gelelektrophoretische Methoden

### 3.2.1. nicht-denaturierende PAA-Gelelektrophorese

Die nicht-denaturierende PAA-Gelelektrophorese wurde für die Gelverzögerungsanalyse eingesetzt. Wahlweise wurden kleine Gele (8 cm x 10 cm) oder große Gele (14 cm x 17 cm) verwendet. Die Gele wurden zwischen zwei Glasplatten gegossen und mindestens 1 h auspolymerisiert. Die Proben wurden mit 6-fach Auftragspuffer versetzt und aufgetragen.

Verwendete Lösungen:

Gellösung:                    10 % Polyacrylamid (30:0,8)  
                                   0,5x TBE  
                                   0,1 Vol% TEMED  
                                   0,5 Vol% APS (10 %)

Laufpuffer:                    0,5x TBE

6x Auftragspuffer:        0,25 % Bromphenolblau  
                                   0,25 % Xylencyanol  
                                   30 % Glycerin

### 3. Material und Methoden

#### 3.2.2. Proteingelelektrophorese nach Laemmli

Für die Auftrennung von Proteinen wurde die denaturierende, diskontinuierliche SDS-Gelelektrophorese nach Laemmli (Laemmli, 1970) durchgeführt.

Verwendete Lösungen:

Trenngel:	10 % Polyacrylamid (30:0,8) 0,375 M Tris/HCL ph 8,8 0,1 Vol% TEMED 0,5 Vol% APS (10 %)
Sammelgel:	4 % Polyacrylamid (30:0,8) 0,125 M Tris/HCl pH 6,8 0,1 Vol% TEMED 0,5 Vol% APS (10 %)
10 x Laemmli-Laufpuffer:	30,26 g Tris 144 g Glycin ad 1l H <sub>2</sub> O
1 x Laemmli-Laufpuffer:	10 Vol% 10 x Laemmli-Laufpuffer 1 Vol% SDS (10 %)
2 x Laemmli-Auftragspuffer:	62,5 mM Tris-Cl pH 6,8 20 mM DTT 10 % (v/v) Glycerin 3 % (w/v) SDS 0,01 % (w/v) Bromphenolblau

Das Trenngel wurde zwischen zwei Glasplatten (8 cm x 10 cm) gegossen, mit Isopropanol überschichtet und für mindestens 1 h auspolymerisiert. Anschließend wurde das Isopropanol abgezogen, mit H<sub>2</sub>O nachgewaschen und das Sammelgel gegossen, das für mindestens 1 h auspolymerisiert wurde. Die Proteinproben wurden mit Auftragspuffer versetzt, 15 min bei 95 °C denaturiert, kurz abzentrifugiert und anschließend auf das Gel aufgetragen. Die Gelelektrophorese wurde zunächst 15 min bei 80 V und anschließend 50 min bei 120 V durchgeführt. Danach wurde das Gel Coomassie gefärbt.

#### 3.2.3. Coomassie-Färbung von Proteingelen

Das Polyacrylamidgel wurde nach der Elektrophorese für 1 h in Coomassie-Blue Lösung inkubiert. Danach wurde das Gel für mindestens 12 h in Entfärber-Lösung unter leichtem Schütteln entfärbt. Anschließend wurde das Gel zur Dokumenta-

tion eingescannt und zur Lagerung in Polyethylenfolie eingeschweißt.

Verwendete Lösungen:

Coomassie-Blue 60 ml H<sub>2</sub>O  
20 ml Methanol  
20 ml Roti-Blue

Entfärber: 12,5 % Isopropanol  
10 % Essigsäure

### 3.3. Molekularbiologische und biochemische Methoden

#### 3.3.1. Herstellung von Proteinextrakten

Sowohl für die Gelverzögerungsexperimente als auch für die Proteinaufreinigung der RNA-Bindeproteine wurden Proteinextrakte aus verschiedenen Gewebetypen der Maus hergestellt. Hierzu wurde das Gewebe mit einer entsprechenden Menge (1  $\mu$ l pro mg Gewebe) Tris/HCl (ph 7,5) und gereinigtem Seesand versetzt und anschließend im Mörser zerkleinert. Der Ansatz wurde in ein Eppendorfgefäß überführt und für 10 min bei 10000 rpm in einer Eppendorfszentrifuge (Heraeus) zentrifugiert. Der Überstand wurde abgenommen und erneut für 10 min bei 10000 rpm zentrifugiert. Der Überstand wurde als Proteinextrakt direkt in den Experimenten eingesetzt. Nach Zugabe von 10 % (v/v) Glycerin wurden die Extrakte bei -20 °C gelagert.

#### 3.3.2. Verzögerungsgelelektrophorese

Die Gelverzögerung dient dem Nachweis spezifischer DNA-Protein oder RNA-Protein Wechselwirkungen. Dabei macht man sich zunutze, daß ein Nukleinsäuremolekül, an das ein oder mehrere Proteine binden, langsamer durch eine Gelmatrix während der Elektrophorese läuft. Das Prinzip dieser Methode ist in Abbildung 3.1 dargestellt. Das mit Digoxigenin oder Biotin markierte RNA-Oligonukleotid wird mit einem Überschuß unspezifischer Hefe-RNA und Proteinextrakt inkubiert. Unspezifische RNA-Bindeproteine binden hauptsächlich an die Hefe-RNA, während spezifische Bindeproteine an das markierte RNA-Oligonukleotid binden, das dadurch in der Gelelektrophorese verlangsamt wird. Dadurch verschwindet die Bande auf Höhe der unverzögerten Oligonukleotide und eine verzögerte Bande tritt auf. Die Spezifität der Bindung wird durch die Zugabe von unmarkierten RNA-Oligonukleotiden überprüft. Durch Zugabe zunehmender

### 3. Material und Methoden

Mengen dieser unmarkierten Oligos binden vermehrt spezifische Proteine an diese unmarkierten Oligos, so daß die markierten Oligos wieder ungestört durch die Gelmatrix wandern können. Die Bande der unverzögerten Oligonukleotide wird dadurch wieder sichtbar. Das umgekehrte Verhalten tritt bei der verzögerten Bande auf (s. Abbildung 3.1).

#### **Durchführung:**

5  $\mu$ l Proteinextrakt wurden zunächst für 10 min mit einem Überschuß (100-1000fach) Hefe-RNA und gegebenenfalls unmarkierten RNA-Oligonukleotiden bei Raumtemperatur inkubiert.

Danach wurden mit Digoxigenin oder Biotin markierte RNA-Oligonukleotide dazugegeben und für weitere 10 min bei Raumtemperatur inkubiert.

Nach Zugabe des Auftragspuffers wurden die Proben auf ein 10 %iges PAA-Gel geladen und die Verzögerungsgelelektrophorese für ca. 1 h bei 70 V durchgeführt.

Anschließend wurden die RNA-Oligonukleotide mittels Northern-Transfer auf eine Membran überführt, um sie mit Chemolumineszenz nachzuweisen.

#### **3.3.3. Northern-Transfer**

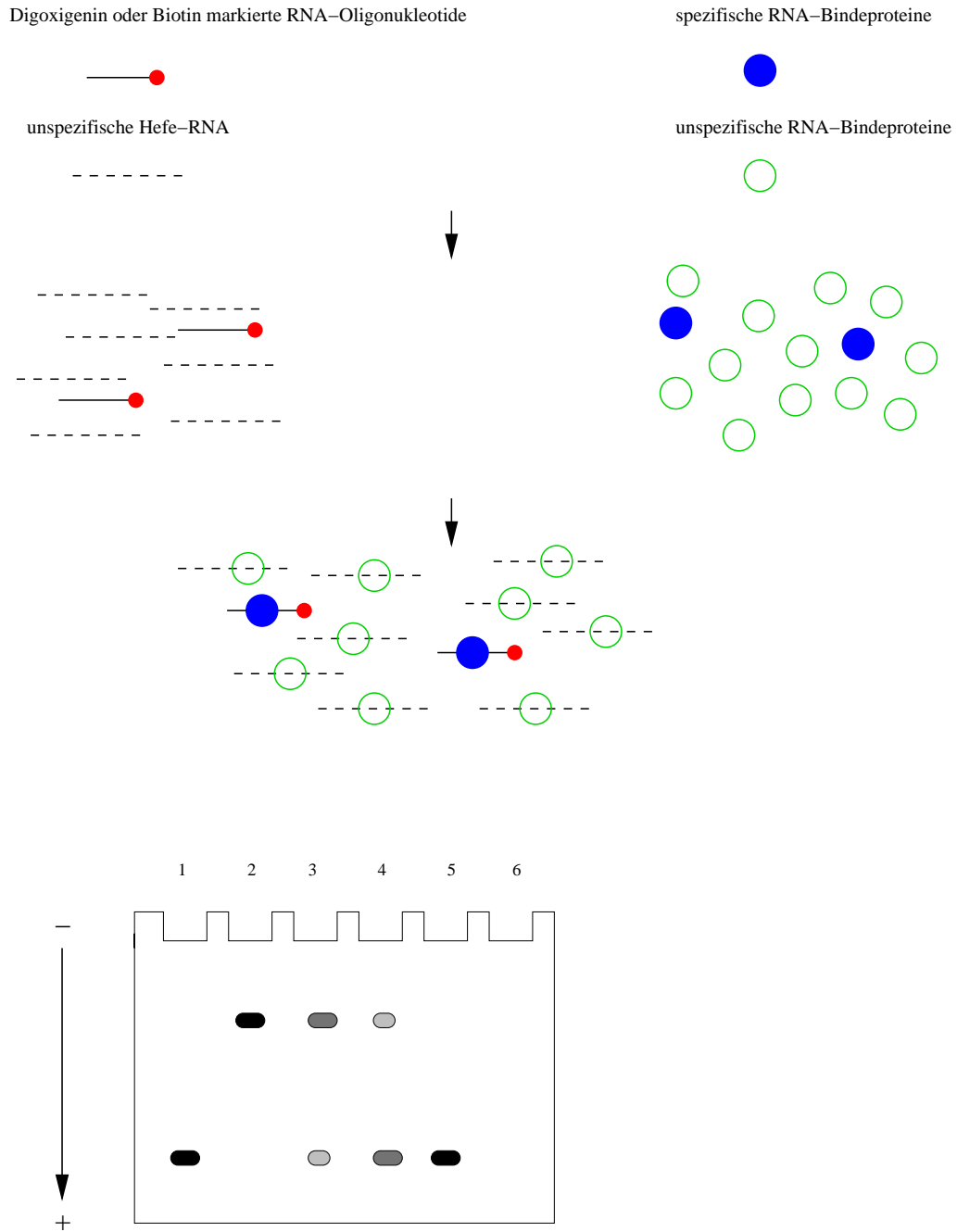
Der Northern-Transfer dient der Übertragung von RNAs, die gelelektrophoretisch aufgetrennt worden sind, auf eine Nitrocellulose- oder Nylonmembran (Sambrook et al., 1989).

Nach der Verzögerungsgelelektrophorese wurde die RNA mittels eines Northern-Transfers auf eine positiv geladene Nylonmembran (Roche, Mannheim) übertragen. Hierfür wurde eine Semi-dry Blottingapparatur aus zwei Graphitplatten verwendet. Das Gel wurde mit der Membran zwischen jeweils 5 Lagen Whatmann-Papier (in 0,5 x TBE getränkt) gelegt und für eine Stunde bei 30 Volt geblottet, wobei 0,5 x TBE als Transferpuffer benutzt wurde. Nach dem Transfer wurde die RNA durch 2 stündiges Backen bei 80 °C auf der Membran fixiert.

#### **3.3.4. Chemolumineszenz-Nachweis Digoxigenin- oder Biotinmarkierter Oligonukleotide**

Durch eine Chemolumineszenzreaktion können mit Digoxigenin oder Biotin markierte Oligonukleotide auf einer Membran durch Autoradiographie nachgewiesen werden. Das Verfahren beruht darauf, daß an spezifische Moleküle, die Digoxigenin (Antikörper) oder Biotin (Streptavidin) erkennen, das Enzym alkali-

### 3.3. Molekularbiologische und biochemische Methoden



1: markierte RNA-Oligonukleotide

2: markierte RNA-Oligonukleotide + Protein

3-5: markierte RNA-Oligonukleotide + RNA-Oligonukleotide (unmarkiert und steigende Konzentrationen) + Protein

Abbildung 3.1.: Prinzip der Gelverzögerung.

### 3. Material und Methoden

sche Phosphatase gebunden ist. Durch Zugabe des Chemolumineszenz-Substrates CDP-Star kommt es zu einer enzymatischen Reaktion mit der alkalischen Phosphatase, welche zu einer Lichtemission führt.

Sämtliche Reagenzien und Lösungen für diesen Nachweis wurden von der Firma Roche Molecular Biochemicals, Mannheim bezogen.

Verwendete Lösungen:

Washing Buffer:	10 x Washing Buffer 1:10 verdünnen
Blocking Solution:	10 x Blocking Solution 1:10 mit Maleic Acid Buffer verdünnen
Maleic Acid Buffer:	10 x Maleic Acid Buffer 1:10 verdünnen
Antibody Solution:	Anti-Digoxigenin-AP (oder Streptavidin-AP) 1:10000 (bzw. 1:5000) in Blocking Solution verdünnen
Detection Buffer:	10 x Detection Buffer 1:10 verdünnen
CDP-Star	ready to use

#### **Durchführung:**

Alle Inkubationsschritte wurden bei Raumtemperatur auf einem Schüttler durchgeführt.

Membran zwei Minuten in 100 ml Washing Buffer inkubieren.

Washing Buffer abgießen und 100 ml Blocking Solution zugeben; 30 min inkubieren.

Blocking Solution abgießen und 100 ml Antibody Solution zugeben; 30 min inkubieren.

Antibody Solution abgießen und 2 mal 15 min in Washing Buffer waschen.

Membran 2-5 min in Detection Buffer äquilibrieren.

Membran auf eine Polyethylenfolie legen und einige Tropfen CDP-Star zugeben: Membran direkt mit Folie bedecken, so daß sich das CDP-Star gleichmäßig ver-



teilt und 5 min inkubieren.

Überschüssige Flüssigkeit entfernen und Folie verschweißen.

Membran für 1-20 min mit Röntgenfilm exponieren.

#### **3.3.5. Autoradiographie**

Mit Hilfe der Autoradiographie lassen sich Chemolumineszenz-Signale sichtbar machen. Dazu wurden die Membranen in einer Expositionskassette entsprechender Größe durch Auflegen eines Röntgenfilms (Kodak, XAR) exponiert. Die Expositionsdauer betrug je nach Signalstärke zwischen 1 und 20 Minuten.

#### **3.3.6. mRNA Detektion durch FRET**

Mit Hilfe des Fluoreszenz Resonanz Energie Transfers (FRET), auch Förster Transfer genannt, soll die räumliche Verteilung einer mRNA in der lebenden Zelle sichtbar gemacht werden. Das Prinzip dieser Methode beruht darauf, daß zwei DNA-Oligonukleotide als Hybridisierungsproben benutzt werden, wobei diese Oligos mit jeweils einer fluoreszierenden Gruppe modifiziert sind. Eine dieser Gruppen wirkt als Donor-Fluorophor und das andere als Akzeptor-Fluorophor. Diese beiden Fluorophore werden durch Hybridisierung an die Ziel-mRNA in direkter Nachbarschaft gebracht. Wenn dann das Donor-Fluorophor angeregt wird, gibt es seine Energie nicht als Fluoreszenz ab, sondern regt das Akzeptor-Fluorophor an. Dies bezeichnet man als Fluoreszenz Resonanz Energie Transfer. Dazu muß das Emissionsspektrum des Donors mit dem Anregungsspektrum des Akzeptors überlappen. Das Akzeptor-Fluorophor gibt diese Energie anschließend als Fluoreszenz wieder ab. Diese Fluoreszenz läßt sich selektiv detektieren, da der Fluoreszenz-Wellenlängenbereich des Akzeptors ein anderer ist als der Bereich, in dem die Fluoreszenz des Donor-Fluorophors liegen würde. Das Prinzip dieser Methode ist in Abbildung 3.2 dargestellt. Der große Vorteil liegt darin, daß die Ziel-mRNA spezifisch detektiert werden kann, da es nur dann zum FRET kommt, wenn die beiden Fluorophore durch korrekte Hybridisierung der Oligonukleotide richtig positioniert sind.

### 3. Material und Methoden

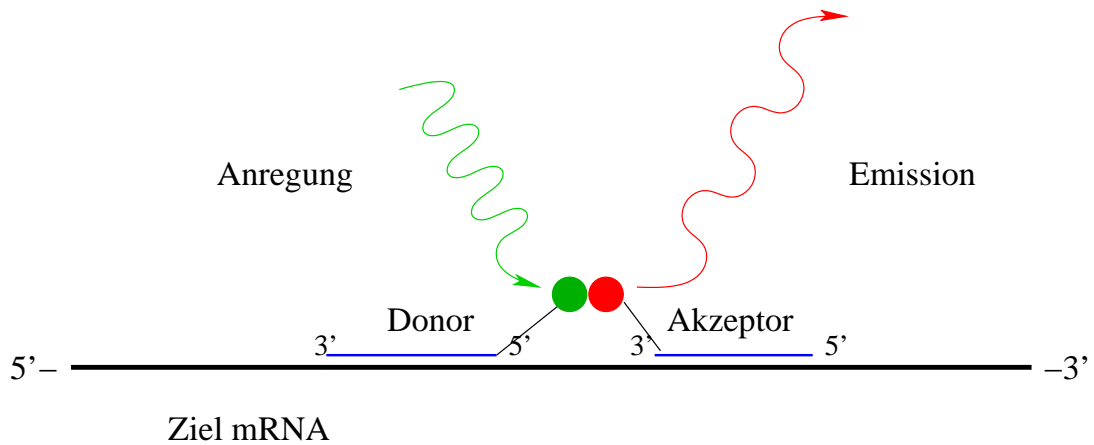


Abbildung 3.2.: Prinzip des Fluoreszenz Resonanz Energie Transfers. Das Donorfluorophor gibt seine Anregungsenergie an das Akzeptorfluorophor ab, dessen Fluoreszenz detektiert werden kann.

#### **in vivo Hybridisierung**

Die Methode des spezifischen mRNA Nachweises mittels FRET wird in der laufenden Doktorarbeit von Ralf Palmisano weiterentwickelt und optimiert. Die modifizierten DNA-Oligonukleotide wurden durch Transfektion mit Liposomen in die Zellen gebracht.

Verwendete Lösungen:

Transfektionsmittel:	Rotifect
Waschlösung:	PBS
Medium:	RPNi 1640
Fixierlösung:	Glutaraldehyd

**Durchführung:**

3 Ansätze:

- Kontrolle 1: nur Donor-Fluorophor (Fluorescein)
- Kontrolle 2: nur Akzeptor-Fluorophor (Texas-Red)
- FRET-Ansatz: beide Oligonukleotide

je 2 Eppendorfgefäße:

- Lösung A: DNA-Oligo (1  $\mu\text{g}$ ) und 60  $\mu\text{l}$  Medium
- Lösung B: 10  $\mu\text{l}$  Transfektionsmedium mit Liposomen

Lösungen A und B zusammenpipettieren und 15 min bei Raumtemperatur stehen lassen.

Zellen präparieren: Volumen so wählen, daß  $5 \times 10^5$  Zellen je Ansatz vorhanden sind. Zellen abzentrifugieren und 2 mal mit 5 ml PBS waschen. Zellen in 2,7 ml Medium resuspendieren und 900  $\mu\text{l}$  je Ansatz in eine Multiwellschale geben.

Die drei Ansätze (Kontrolle 1 und 2 und FRET-Ansatz) werden zu den vorbereiteten Zellen gegeben und 1 h bei 37 °C inkubiert.

Nach der Inkubation 3 mal mit PBS waschen; anschließend mit PBS und 20  $\mu\text{l}$  Glutaraldehyd waschen, 30 sec abzentrifugieren, 10  $\mu\text{l}$  des Überstandes werfen und den Rest auf einen Objektträger geben.

Nachweis des FRET im konfokalen Laser Scanning Mikroskop.

**konfokale Laser-Scanning-Mikroskopie**

Mit Hilfe eines Laser-Scanning-Mikroskops ist eine dreidimensionale Analyse der fluoreszierenden Oligonukleotide innerhalb der Zelle möglich. Die FRET Experimente lassen sich durchführen, indem zunächst mit einem Laser gezielt das Donor-Fluorophor angeregt wird und dann die Emission des Akzeptor-Fluorophors mit einer CCD-Kamera gemessen wird. Als Kontrolle wird auch die Emission des Donors, nach Anregung des Donors und die Emission des Akzeptors, nach Anregung des Akzeptors ermittelt. Im Rahmen dieser Arbeit wurde ein Laser-Scanning-Mikroskop vom Typ TCS SP2 der Firma Lyca benutzt. Die entsprechenden Wellenlängen, mit denen die Fluorophore angeregt und bei denen ihre Emission gemessen wurde, werden bei der Darstellung der Ergebnisse angegeben.

### Protein-Isolation mit Dynabeads

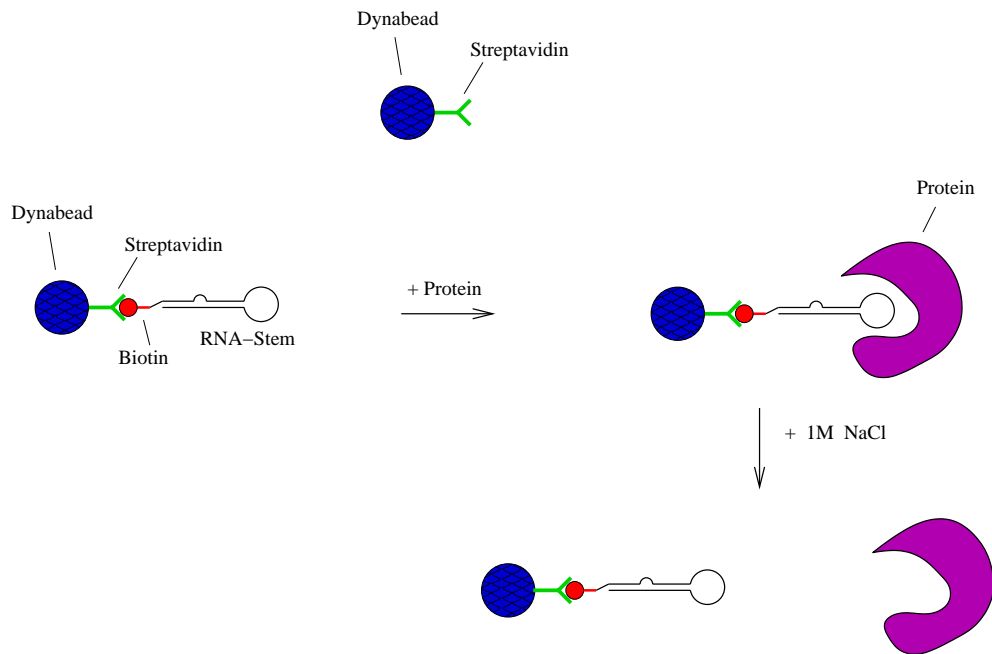


Abbildung 3.3.: Prinzip der Proteinisolation mit Dynabeads. Proteine aus einem Zellextrakt binden an das RNA-Oligonukleotid, das mit den Dynabeads über Streptavidin Moleküle verbunden ist. Unter Hochsalzbedingungen können die Proteine wieder von dem Oligonukleotid getrennt werden, nachdem sie durch einen Magneten vom Rest des Extraktes isoliert worden sind.

#### 3.3.7. Proteinaufreinigung

Um spezifische RNA-Bindeproteine zu isolieren, wurde im Rahmen dieser Arbeit eine moderne Aufreinigungsmethode, bei der sogenannte Dynabeads eingesetzt werden, verwendet. Das Prinzip der Methode ist in Abbildung 3.3 dargestellt. Bei den Dynabeads handelt es sich um magnetische, sphärische Partikel (Durchmesser  $< 100 \mu m$ ), an deren Oberfläche Streptavidin Moleküle gebunden sind. An diese Streptavidin Moleküle können mit Biotin markierte RNA-Oligonukleotide gebunden werden. Nach Zugabe von Proteinextrakt binden Proteine an die RNA-Oligonukleotide. Mit einem Magneten werden anschließend die Dynabeads mit den RNA-Bindeproteinen isoliert. Durch Hochsalzbedingungen können die Proteine dann von dem RNA-Oligo gelöst und anschließend in einem Proteingel aufgetrennt werden.

### RNA-Oligonukleotidbindung an Dynabeads

Für die Proteinisolation wurden RNA-Oligonukleotide verwendet, die an ihrem 5' Ende mit Biotin modifiziert waren. Diese konnten an die Dynabeads gebunden werden. Dabei kann ein mg Dynabeads bis zu 200 pmol biotinylierte RNA-Oligonukleotide binden.

Verwendete Lösungen:

Dynabeads M-280 Streptavidin (10 mg/ml)

2 x B+W Binding-Buffer:                      10 mM Tris/HCl pH 7,5  
                                                            1 mM EDTA  
                                                            2 M NaCl

1 x B+W Binding-Buffer:                      1:1 Verdünnung des  
                                                            2 x B+W Binding-Buffers

Buffer C:                                              20 mM Tris/HCl pH 7,5  
                                                            1 mM EDTA  
                                                            50 mM NaCl

1M NaCl

#### Durchführung:

Ein entsprechendes Volumen Dynabeads (meistens 200  $\mu$ l) in ein Eppendorfgesäß geben und anschließend 2 min in Magnetständer (DynaL MPC-Q) stellen.

Überstand vorsichtig abnehmen und Dynabeads 2 mal in gleichem Volumen 2x B+W Binding-Buffer waschen. Überstand abnehmen, nachdem Eppendorfgesäß mindestens 1 min im Magnetständer gestanden hat.

Dynabeads in gleichem Volumen 1 x B+W Binding-Buffer resuspendieren und die Hälfte als unmarkierte Dynabeads (Kontrolle) abnehmen.

Zugabe von entsprechender Menge RNA-Oligonukleotiden (200 pmol pro mg Beads) zu den Dynabeads und 15 min bei Raumtemperatur unter vorsichtigem Schütteln inkubieren. Anschließend für 2 min in Magnetständer stellen.

Überstand abnehmen und 2-3 mal in 2 x B+W Binding-Buffer und anschließend 2 mal in Buffer C waschen. Dynabeads in Buffer C resuspendieren (Endkonzentration: 10  $\mu$ g/ $\mu$ l).

### 3. Material und Methoden

Mit den unmarkierten Dynabeads (Kontrolle) wird auf die gleiche Art verfahren.

#### **Proteinaufreinigung mit Dynabeads**

Die mit RNA-Oligonukleotiden markierten Dynabeads wurden eingesetzt, um Proteine aus einem Zellextrakt zu isolieren, die an das RNA-Motiv binden können. Um unspezifische Bindungen so weit wie möglich zu verhindern, wurde den Ansätzen Hefe-Gesamt-RNA als Kompetitor zugegeben. Neben den mit RNA-Oligos markierten Dynabeads wurden auch unmarkierte Dynabeads als Negativkontrolle eingesetzt, um unspezifische Proteinbindungen an die Beads nachzuweisen. Nachfolgend wird das Standard-Protokoll zur Proteinaufreinigung mit Dynabeads angegeben.

#### **Durchführung:**

250-500  $\mu\text{g}$  Dynabeads (das entspricht ca. 0,5-1  $\mu\text{g}$  RNA-Oligonukleotiden bei den markierten Beads) wurden je nach Experiment mit 10-30  $\mu\text{l}$  Proteinextrakt und einem 50-100fachen Überschuß an Hefe RNA für 15 Minuten bei Raumtemperatur unter vorsichtigem Schütteln inkubiert.

Anschließend wurden die Eppendorfgefäße für mindestens 1 Minute in den Magnetständer gestellt und der Überstand vorsichtig abgenommen.

Danach wurden die Beads 3 mal für 5 min in 300  $\mu\text{l}$  Buffer C gewaschen, anschließend noch einmal in 50  $\mu\text{l}$  Buffer C.

Nach dem letzten Waschen wurden die Beads in 10-15  $\mu\text{l}$  1 M NaCl Lösung inkubiert, um die Proteine von den RNA-Oligonukleotiden zu lösen.

Je nach Experiment wurden die Proteine mit den Beads oder ohne die Beads in einem eindimensionalen SDS-Gel aufgetrennt.

#### **3.3.8. Proteinidentifikation**

Nach der Aufreinigung der RNA-bindenden Proteine mit Hilfe der Dynabead-Technologie können die isolierten Proteine durch massenspektrometrische Methoden identifiziert werden. Dazu werden die Proteinbanden aus dem jeweiligen Gel ausgeschnitten und einem proteolytischen Verdau (in der Regel mit dem Enzym Trypsin) unterzogen. Die Massen der so erhaltenen Peptide können dann mittels MALDI-TOF Massenspektrometrie (s.u.) ermittelt werden. Durch den

Vergleich der so erhaltenen Peptidmassen mit simulierten Proteinverdaus bereits bekannter Proteine kann das Protein identifiziert werden. Voraussetzung ist jedoch, daß das Protein bereits in einer Datenbank enthalten ist. Der Datenbankvergleich kann mit dem *Mascot Search* Programm durchgeführt werden. URL: <http://www.matrixscience.com>

#### **MALDI-TOF Spektrometrie**

MALDI steht für Matrix-assisted-laser-desorption-ionisation Spektrometrie und TOF für time-of-flight, was sich auf den Massenanalysator bezieht. Die zu analysierende Probe wird mit einem Überschuß eines geeigneten Matrixmaterials gemischt. Durch einen intensiven Impuls eines Lasers wird sowohl das Matrixmaterial als auch Probenmoleküle ionisiert. Die Ionen werden in einem elektrischen Feld beschleunigt und durchlaufen anschließend eine feldfreie Strecke. Hier werden sie nach ihrem Masse/Ladungsverhältnis aufgetrennt, so daß die Masse der Probenmoleküle aus der in einer bestimmten Zeit zurückgelegten Strecke berechnet werden kann.

Die MALDI-Analysen im Rahmen dieser Arbeit wurden von der Arbeitsgruppe von Dr. Carsten Niehaus am Lehrstuhl für Genetik der Universität Bielefeld oder alternativ von der Firma Seqlab (Göttingen) durchgeführt.

## **3.4. Informatikmethoden**

### **3.4.1. Datenbanken**

Im Zeitalter der großen Genom- und Sequenzierprojekte spielen Datenbanken eine entscheidende Rolle. Die Menge an vorhandenen Daten (in der Regel Nukleotidsequenzen) ist unüberschaubar groß geworden. So ist zum Beispiel die Größe der EMBL-Datenbank (s.u.) von 13,8 Gigabasen im Jahr 2001 auf über 23 Gigabasen im Jahr 2002 angewachsen. Darüber hinaus gibt es eine Vielzahl von Datenbanken, die nur ganz spezielle Einträge enthalten, wie zum Beispiel Transkriptionsfaktorbindestellen. Im Rahmen dieser Arbeit wurden im wesentlichen zwei Datenbanken verwendet, die EMBL Nukleotid Datenbank und die UTRdb, die ausschließlich UTR Sequenzen enthält.

#### **EMBL**

Die EMBL Nukleotid Datenbank enthält praktisch alle öffentlich zugänglichen Sequenzen (Stoesser et al., 2003). Sie wird am European Bioinformatics Institute (EBI) in Hinxton verwaltet. Durch eine internationale Zusammenarbeit mit

### 3. Material und Methoden

GenBank aus den USA und DDBJ aus Japan sind ständig weltweit aktuelle Sequenzinformationen vorhanden. Wie bereits erwähnt, nimmt die Datenmenge exponentiell zu, so daß im September 2002 bereits 18,3 Millionen Einträge mit über 23 Gigabasen in der EMBL-Datenbank enthalten waren. Diese Datenbank wurde im Rahmen meiner Arbeit vor allem dann eingesetzt, wenn andere als UTR Sequenzen benötigt wurden, oder bestimmte UTR Sequenzen nicht in der UTR Datenbank (s.u.) enthalten waren.

Auf Sequenzen der EMBL Nukleotid Datenbank kann über den SRS Server des EBI zugegriffen werden.

<http://srs.ebi.ac.uk/>

#### **UTRdb und UTRsite**

UTRdb ist eine spezielle, nicht redundante Datenbank für 5' und 3' untranslatierte Regionen von mRNAs (Pesole et al., 2002). Die Datenbank ist in mehrere Kollektionen untergliedert. UTRdb 15.0 enthält Kollektionen für die 5'UTRs von Mensch (Human), Nagern (Rodent), anderen Säugetieren (Other\_mammal), anderen Vertebraten (Other\_vertibrate), Invertebraten, Pflanzen (Plant), Pilzen (Fungi) und Viren (Virus). Diese Kollektionen haben insgesamt eine Größe von 16,8 Megabasen. Desweiteren gibt es Kollektionen für die 3'UTRs von Mensch, Nagern, anderen Säugetieren, anderen Vertebraten, Invertebraten, Pflanzen, Pilzen und Viren, die insgesamt eine Größe von 47,2 Megabasen haben, wobei die Kollektion der 3'UTRs vom Menschen mit 18,9 Megabasen mit Abstand am größten ist. Die UTRdb 12.0 aus dem Jahr 2000 enthielt dagegen nur 6,5 Megabasen für die 5'UTRs und 23,5 Megabasen für die 3'UTRs. Die Sequenzinformationen über UTRs haben sich somit in zwei Jahren verdoppelt. Die einzelnen Einträge enthalten neben den reinen Sequenzen auch Informationen über die Vollständigkeit einer UTR Sequenz und über das Vorhandensein bereits bekannter regulatorischer Motive, wie zum Beispiel das *Iron Responsive Element*. Solche bereits bekannten regulatorischen Motive sind in einer weiteren Datenbank, die als UTRsite bezeichnet wird, zusammengefaßt. In der aktuellen Version aus dem Jahr 2002 enthält diese Datenbank 18 verschiedene regulatorische Motive (im Vergleich zu 13 aus dem Jahr 2000).

Sowohl UTRdb als auch UTRsite sind unter <http://bighost.area.ba.cnr.it/BIG/UTRHome/> zugänglich.



### 3.4.2. Sequenzalignment

Im Rahmen dieser Arbeit wurden die Programme `DIALIGN2` und `ClustalW` zur Berechnung von Sequenzalignments verwendet.

#### **DIALIGN2**

Das Programm `DIALIGN2` wurde von Burkhard Morgenstern entwickelt (Morgenstern, 1999).

`DIALIGN2` eignet sich sowohl zum paarweisen als auch multiplen Alignment von Nukleotid- oder Proteinsequenzen. Im Vergleich zu vielen anderen Sequenzalignmentprogrammen (s. zum Beispiel `ClustalW`) werden bei `DIALIGN2` nicht einzelne Basen (oder Aminosäuren), sondern größere Segmente der Sequenzen verglichen, wobei diese Segmente keine Lücken (gaps) aufweisen. Obwohl `DIALIGN2` multiple globale Alignments berechnet, liegt der große Vorteil für die vorliegende Arbeit darin, daß unter bestimmten Bedingungen auch lokale Ähnlichkeiten in einem multiplen globalen Alignment gefunden werden können. Dabei müssen diese lokal ähnlichen Bereiche nicht in allen Sequenzen, die zur Berechnung des Alignments verwendet wurden, vorhanden sein.

Das Programm ist auf dem Bielefelder Bioinformatik Server (BiBiServ) unter <http://bibiserv.TechFak.Uni-Bielefeld.DE/dialign/> verfügbar.

#### **ClustalW**

Zur Berechnung globaler Sequenzalignments wurde das Programm `ClustalW` verwendet (Thompson et al., 1994). Multiple Alignments werden hierbei progressiv berechnet. Das heißt, es werden zunächst die besten paarweisen Alignments berechnet, um anschließend immer größere Gruppen von Sequenzen zu alignieren.

Das Programm kann über das EBI unter <http://www.ebi.ac.uk/clustalw/> benutzt werden.

### 3.4.3. RNA-Sekundärstrukturvorhersage

Die Vorhersage von RNA-Sekundärstrukturen wurde im Rahmen dieser Arbeit mit den Programmen `mfold` und `RNAfold` durchgeführt. Beide Programme arbeiten mit Energie-Minimierungs-Algorithmen.

#### **mfold**

Das Programm `mfold` wurde von Michael Zuker entwickelt (Zuker, 1989). In dieser Arbeit wurde die Version `mfold 3.1` benutzt, die die neuesten Energieparameter verwendet (Mathews et al., 1999). Neben der energetisch günstigsten Struktur

### 3. Material und Methoden

können auch energetisch suboptimale Strukturen ausgegeben werden. Dabei ist der Energiebereich, in dem suboptimale Strukturen ausgegeben werden sollen, frei wählbar. Desweiteren wird ein Dotplot angezeigt, der einen Überblick über einzelne Basenpaarwahrscheinlichkeiten liefert. Die Temperatur ist auf 37 °C eingestellt und läßt sich in der benutzten Version nicht variieren.

`mfold` 3.1 ist auf Michael Zukers Homepage unter <http://www.bioinfo.rpi.edu/~zukerm/> verfügbar.

#### **RNAfold**

Das Programm `RNAfold` ist Teil des Vienna RNA-Pakets (Hofacker et al., 1994). Im Rahmen dieser Arbeit wurde die lokal vorhandene Version `RNAfold` 1.1.0.42 verwendet. Dieses Programm berechnet (genau wie `mfold`) von einer RNA Sequenz die sogenannte mfe-Struktur (minimal free energy) und die Basenpaarwahrscheinlichkeiten als Dotplot. Energetisch suboptimale Strukturen können mit dem Programm `RNAsubopt` berechnet werden (s.u.). `RNAfold` rechnet standardmäßig mit einem Temperaturwert von 37 °C, die Temperatur kann jedoch als Parameter angegeben werden. Desweiteren können verschiedene andere Parameter angegeben werden, wie zum Beispiel das Erlauben von standardmäßig nicht vorkommenden Basenpaaren (z.B. AG Paare). Wenn solche von der Standard-einstellung abweichende Parameter benutzt wurden, wird im Ergebnisteil an den entsprechenden Stellen darauf hingewiesen.

Das Vienna RNA-Paket kann unter <http://www.tbi.univie.ac.at/~ivo/RNA/> benutzt werden.

#### **RNAsubopt**

`RNAsubopt` ist ebenfalls Teil des Vienna RNA-Pakets und wurde von Stefan Wuchty und Kollegen entwickelt (Wuchty et al., 1999). Es berechnet sämtliche suboptimale RNA-Sekundärstrukturen innerhalb eines Energiebereichs, der vom Benutzer gewählt werden kann. Damit unterscheidet es sich von dem Programm `mfold`, das nur eine Auswahl von suboptimalen Strukturen ausgibt. Wie bei `RNAfold` lassen sich auch hier verschiedene Parameter, wie zum Beispiel die Faltungstemperatur, frei wählen.

#### **3.4.4. RNA-Strukturalignment**

Im Rahmen dieser Arbeit wurde erstmals das Programm `RNAforester` eingesetzt, das zeitgleich zu meiner Arbeit in der Doktorarbeit von Matthias Höchsmann entwickelt worden ist. `RNAforester` ermöglicht die effiziente Berechnung

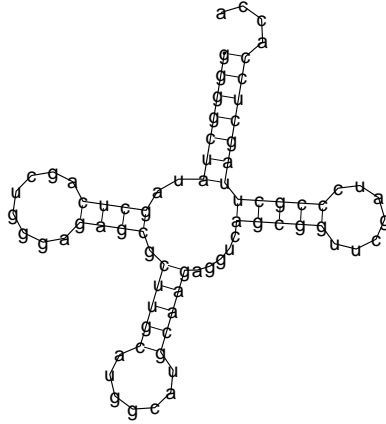
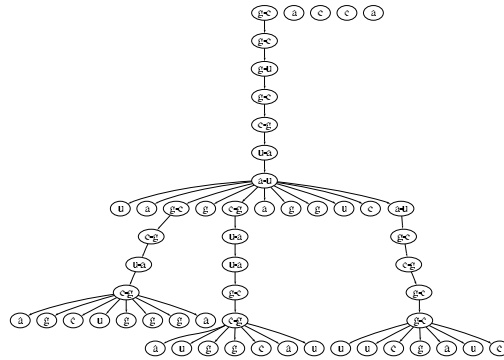
Abbildung 3.4.: Sekundärstruktur der *E.coli* tRNA für Alanin.

Abbildung 3.5.: Baum-Repräsentation der tRNA aus Abbildung 3.4. Basenpaarungen entsprechen internen Knoten und ungepaarte Basen entsprechen den Blättern des Baums.

von RNA-Sekundärstrukturalignments, sowie eine anwendungsorientierte Visualisierung dieser Alignments. Dabei erlaubt das Programm sowohl die Berechnung von lokalen paarweisen Strukturalignments als auch von globalen multiplen Strukturalignments.

Das Programm beruht darauf, daß RNA-Sekundärstrukturen als spezielle Datenstrukturen, sogenannte Bäume, repräsentiert werden und mit diesen ein Baumalignment berechnet wird. Abbildung 3.4 zeigt die Sekundärstruktur einer Transfer RNA als sogenannten "Squiggle-Plot" und Abbildung 3.5 zeigt die Repräsentation dieser Sekundärstruktur als Baum. Dabei gibt die 5'-3' Richtung der RNA-Sekundärstruktur die Reihenfolge der Knoten und Blätter im Baum vor. Durch die Anwendung bestimmter Editoperationen, wie zum Beispiel Basenpaareretzungen oder Basenpaareinfügungen bzw. Löschungen, kann für mehrere Bäume

### 3. Material und Methoden

ein Alignment berechnet werden. Da man nur an bestimmten Alignments, die ein gewisses Optimalitätskriterium erfüllen, interessiert ist, werden die einzelnen Editoperationen mit einem sogenannten Score gewertet. Im Rahmen meiner Arbeit wurden Ähnlichkeiten (und keine Distanzen, was jedoch als Parameter eingestellt werden kann) betrachtet, so daß der Gesamt-Score maximiert wird. Das Programm erlaubt die Definition eigener Score-Werte, welche dem Programm als Parameter übergeben werden können.

Das Programm bekommt als Eingabe zwei (bei paarweisen globalen oder lokalen Alignments) oder mehrere (bei multiplen globalen Alignments) RNA Sekundärstrukturen. Diese werden in der sogenannten Punkt-Klammer Notation eingegeben. In dieser Notation werden ungepaarte Basen als Punkt und die Basen, welche eine Basenpaarung bilden, als Klammerpaar angegeben. Das Programm erlaubt die Auswahl verschiedener Parameter. Im Rahmen meiner Arbeit wurden vor allem folgende Parameter verwendet:

<code>-ls</code>	berechnet ein lokales Alignment
<code>-multi</code>	berechnet ein multiples Alignment
<code>-bp_rep_score=num</code>	setzt den Score für eine Basenpaarersetzung auf num
<code>-bp_del_score=num</code>	setzt den Score für eine Basenpaarlöschung/einfügung auf num
<code>-b_match_score=num</code>	setzt den Score für einen Basentreffer auf num
<code>-b_rep_score=num</code>	setzt den Score für eine Basenersetzung auf num
<code>-b_del_score=num</code>	setzt den Score für eine Baseneinfügung/löschung auf num
<code>-sq</code>	generiert das Alignment als Squiggle-Plot in einer Postscriptdatei
<code>-sq_hidebasenum</code>	gibt keine Basennummerierung im Squiggle-Plot an
<code>-sq_basenuminterval=n</code>	gibt die Nummer jeder n-ten Base an
<code>-sq_gray</code>	benutzt nur Graustufen im Squiggle-Plot

Die zur Berechnung der jeweiligen Strukturalignments gewählten Score-Werte werden im Ergebnisteil an den entsprechenden Stellen angegeben.

Das Programm kann auf dem BiBiServ benutzt werden.

<http://bibiserv.TechFak.Uni-Bielefeld.DE/RNAforester>.

### 3.4.5. Musterdefinition durch Signifikanzberechnungen

Zur Definition eines Datenbank-Suchmusters für potentielle regulatorische RNA-Motive wurde ein, von Carsten Meyer und Robert Giegerich entwickelter, Ansatz verwendet (Meyer und Giegerich, 2002). Das auf dem Prinzip der Algebraischen Dynamischen Programmierung (ADP) (Giegerich und Meyer, 2002) beruhende Verfahren kann dazu eingesetzt werden, die zu erwartende Trefferanzahl eines bestimmten Suchmusters in einer Zufallssequenz bestimmter Länge und Basenzusammensetzung nach oben abzuschätzen. Somit kann man einem erzeugten Suchmuster einen Signifikanzwert zuordnen. Dadurch läßt sich das Muster (im biologisch sinnvollen Rahmen) so lange variieren, bis es eine Signifikanz erreicht, mit der man verläßliche Datenbanksuchen durchführen kann.

Bei dem Verfahren wird das Muster zunächst mit einer Grammatik beschrieben. Auf diese Grammatik wird in einer zweiten Phase eine Auswertungsalgebra angewendet. Im Rahmen meiner Arbeit wurde ausschließlich eine Algebra zur Signifikanzberechnung verwendet. Mit dem Verfahren lassen sich durch Anwendung anderer Algebren jedoch auch andere Fragestellungen, wie zum Beispiel die Bestimmung der minimalen freien Energie eines RNA-Motivs, beantworten. Die Beschreibung von RNA-Motiven durch eine Grammatik wird im Ergebnisteil dargestellt.

### 3.4.6. Motiv-Beschreibung und Suche

Für die Beschreibung und Datenbanksuche von RNA-Motiven, die sowohl aus Sequenz- als auch aus Strukturanteilen bestehen, wurden spezielle Programme eingesetzt. Zum einen wurde das von Thomas Macke und Kollegen entwickelte Programm *RNAMotif* (Macke et al., 2001) und zum anderen das in der laufenden Doktorarbeit von Dirk Strothmann entwickelte Programm *HyPa* verwendet.

#### **RNAMotif**

Das Programm zeichnet sich durch eine flexible Sprache zur Strukturmusterbeschreibung sowie der Möglichkeit, eine benutzerdefinierte Score-Funktion zu verwenden, aus.

Die Musterbeschreibung erfolgt in der sogenannten Descriptor-Datei, die aus verschiedenen Teilen besteht. In dem "parameter" Abschnitt wird definiert, welche Basenpaarungen zugelassen werden, wobei in meiner Arbeit immer mit den Watson-Crick Basenpaaren und dem zusätzlichen GU-Basenpaar gearbeitet wurde. In dem folgenden "descriptor" Abschnitt erfolgt die eigentliche Beschreibung des RNA-Motivs als Suchmuster. Der Aufbau einer solchen Beschreibung wird im Ergebnisteil ausführlich beschrieben. Der "score" Abschnitt erlaubt es dem Benutzer, für bestimmte Teile des Suchmusters bestimmte Score Werte anzugeben (z.B. das Vorkommen eines Cytosin in einem Bulge höher zu bewerten, als das

### 3. *Material und Methoden*

Vorkommen eines Adenin). Dadurch wird eine nach Gesamtscore-Werten geordnete Ausgabe erreicht. Es ist genauso möglich, die thermodynamische Stabilität einer Gesamtstruktur oder einzelner Strukturabschnitte zu bewerten. Dazu verfügt das Programm über eine spezielle Funktion, die in dem "score" Abschnitt eingesetzt werden kann, um die Energie der Struktur zu bestimmen. Dazu werden die gleichen Energiewerte (Mathews et al., 1999) verwendet, die auch von den oben beschriebenen Strukturvorhersageprogrammen benutzt werden.

Das Programm muß lokal installiert werden und wird von der Kommandozeile aus aufgerufen. Als Parameter wird die Descriptor-Datei und die zu durchsuchende Sequenz im FASTA Format übergeben. Die Ausgabe erfolgt in einem bestimmten Format, das im Ergebnisteil dargestellt wird.

#### **HyPa**

HyPa ist die Abkürzung für "hybrid patterns". Es können (wie bei `RNAMotif`) Muster beschrieben werden, die aus Sequenz- und Strukturanteilen bestehen. Im Gegensatz zu `RNAMotif` ist es bei HyPa möglich, für Helixbereiche eine bestimmte Anzahl von Insertionen und Deletionen zu definieren. Darüber hinaus ähneln sich beide Programme in ihrer Funktionalität. Bei HyPa ist es ebenfalls möglich, bestimmten Musterabschnitten Score-Werte zuzuordnen und so die Ausgabe zu ordnen. Die Art der Musterbeschreibung wird im Ergebnisteil beschrieben. Im Ergebnis- und Diskussionsteil werden beide Programme hinsichtlich ihrer Funktionalität und Effizienz verglichen.

Das Programm läuft ebenfalls lokal. Im Gegensatz zu `RNAMotif` werden die zu durchsuchenden Sequenzen nicht im FASTA Format übergeben. Bei HyPa wird zunächst ein Index der zu durchsuchenden Sequenz generiert. Dies ist erforderlich, da die Motivsuche bei HyPa auf einer speziellen Datenstruktur, einem sogenannten Affix-Array, erfolgt.

## 4. Ergebnisse

In diesem Kapitel wird zunächst die Entwicklung computergestützter Methoden und Strategien zur Vorhersage potentiell regulatorischer RNA-Struktur motive beschrieben. Hierfür wurden Testsysteme ausgewählt, die unterschiedliche Anforderungen an die Vorhersagestrategien stellen. Anschließend werden Methoden dargestellt, die zur Definition und Beschreibung eines Musters für Datenbanksuchen eingesetzt wurden. Dies dient der Validierung des vorhergesagten regulatorischen Motivs. Danach erfolgt die Anwendung der Strategien auf verschiedene UTRs, um neue potentiell regulatorische Motive zu untersuchen. Im abschließenden Teil wird die experimentelle Überprüfung eines dieser neu vorhergesagten regulatorischen Motive beschrieben.

### 4.1. Strategieentwicklung

#### 4.1.1. Testsysteme

Für die Entwicklung von Methoden zur Vorhersage regulatorischer RNA Struktur motive ist die Auswahl geeigneter Testsysteme entscheidend. Ein solches Testsystem muß die Kombination von Sequenz- und Strukturelementen innerhalb regulatorischer Motive in UTRs berücksichtigen. Im folgenden werden mehrere Testsysteme beschrieben.

Die UTR Sequenzen für die Testsysteme wurden aus der UTR-Datenbank (Pesole et al., 2002) entnommen. Die Zugangsnummern (Accessionnumber) sowie die Sequenzen sind im Anhang A aufgelistet.

#### **Testsystem 1: Iron Responsive Element**

Das in der Einleitung beschriebene *Iron Responsive Element* (IRE) wurde aus verschiedenen Gründen als ein Testsystem ausgewählt. Zum einen ist für das IRE eine Konsensusstruktur bekannt (s. Abbildung 4.1). Bei Eukaryonten besteht das IRE aus einem Sechserloop mit der Konsensussequenz 5'-CAGUGH-3', einem Helixbereich mit fünf Basenpaaren variabler Sequenz, an den sich auf der 5'-Seite ein Cytosin-Bulge anschließt, der jedoch zu einem Internalloop erweitert sein kann. Unterhalb des Bulgeloops folgt ein weiterer basengepaarter Bereich, dessen Länge und Sequenz variabel ist. Damit weist das IRE die für ein Testsystem notwendige Kombination aus spezifischen Sequenz- und Strukturmerkmalen auf. Ein weiterer Vorteil ist, daß es viele verschiedene UTRs gibt, in denen IREs

#### 4. Ergebnisse

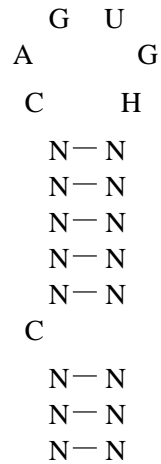


Abbildung 4.1.: *Iron Responsive Element*: eukaryotische Konsensusstruktur.

vorkommen, wobei diese UTRs deutliche Größenunterschiede aufweisen (von 150 bis 2500 Nukleotiden). Somit können die Vorhersagestrategien auch auf ihre Effizienz getestet werden.

#### Testsystem 2: modifizierte IREs und UTRs

Das in Abschnitt 2.6 beschriebene Programm FOLDALIGN wurde unter anderem mit modifizierten IREs getestet (Gorodkin et al., 2001). Hierzu wurden 14 UTR Sequenzen ausgewählt, aus denen das IRE zunächst entfernt wurde. Danach wurde jede UTR randomisiert, wobei die Dinukleotidverteilung erhalten blieb. Die IREs wurden in ihrer Sequenz geändert, wobei die Struktur beibehalten wurde. Der 6er-Loop des IRE wurde in einigen Fällen zu einem 7er-Loop erweitert und die IREs wurden anschließend an völlig verschiedenen Positionen in die veränderten UTRs eingefügt. Durch viermaliges Wiederholen dieser Prozedur wurden insgesamt 56 Sequenzen als Testsystem erhalten. Für das IRE ergibt sich danach folgende Konsensusstruktur:

```
NNNNNCNNNNNCAXGWHNNNNNNNNNN  
(((((((((.....))))))))))
```

mit W=A,U und H=A,C,U und X=A,C,G,U (wobei X auch fehlen darf)

Dieses Testsystem wird im Rahmen meiner Arbeit ebenfalls eingesetzt, um die Vergleichbarkeit mit dem FOLDALIGN Ansatz zu gewährleisten.



## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

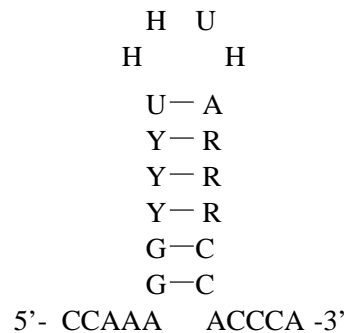


Abbildung 4.2.: Konsensusstruktur des Histon Stemloops, H=A,C,U; Y=C,U; R=A,G.

### Testsystem 3: Histon Stemloop

Das IRE zeigt vor allem in der Loop Region eine starke Sequenzkonservierung. Der Stemloop, der in der 3'UTR der Histon mRNAs von Metazoen anstelle des PolyA Stranges zu finden ist (s. Abbildung 4.2), weist dagegen vor allem in der basengepaarten Region sowie in den einzelsträngigen Bereichen vor und hinter dem Stemloop eine starke Sequenzkonservierung auf (Williams und Marzluff, 1995). Die Loopregion ist dagegen auf Sequenzebene deutlich variabler als beim IRE. Besonders die Sequenzkonservierung im basengepaarten Bereich ist ungewöhnlich für regulatorische Motive in UTRs und daher wurden 3'UTRs von Histon mRNAs als ein weiteres Testsystem ausgewählt, das die IRE Testsysteme ergänzen soll.

## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

### 4.2.1. Strategie 1: Sequenzalignment mit nachfolgender Strukturvorhersage

#### Sequenzalignment

Die Grundidee dieser Strategie basiert darauf, daß ein potentiell regulatorisches Motiv ein gewisses Maß an Sequenzkonservierung aufweisen kann. Durch ein Sequenzalignment mehrerer UTRs, in denen ein regulatorisches Motiv vermutet wird, soll dieses aufgrund der Sequenzähnlichkeit innerhalb des Motives entdeckt werden. Für das Sequenzalignment wurde das in Abschnitt 3.4.2 beschriebene Programm DIALIGN2 verwendet. Es eignet sich für die Berechnung multipler Alignments und das Auffinden lokaler Ähnlichkeiten innerhalb der global alignierten Sequenzen. Ein großer Vorteil von DIALIGN2 besteht darin, daß lokal ähnliche Be-

#### 4. Ergebnisse

reiche in einem multiplen Alignment selbst dann gefunden werden, wenn sie nur in zwei der alignierten Sequenzen auftreten. Das ist eine wichtige Eigenschaft, da man nicht sicher sein kann, daß jede der im Alignment eingesetzten Sequenzen das regulatorische Motiv enthält.

Die Strategie wird zunächst am Beispiel der Histon 3'UTRs dargestellt. Nachfolgend ist das Alignment für die UTR Sequenzen von Histon mRNAs von Mensch, Maus und Ratte abgebildet. Der Histon Stemloop ist in blau dargestellt.

```

Histon 3'UTRs: Alignment (DIALIGN format):

Maus H2A      1  ggccactGCT CGCGT----- -----CGGC CCGCGTCTCt
Maus H3       1  attaggta- ----- -----
Ratte H2B     1  g----- ----- -----
Mensch H2A    1  g-----GCT GACGTccggc ccaagtgggc ccagccCGGC CCGCGTCTCg

                *** *****                **** *****

Maus H2A      30  GTGAtataaC CCCCGCGAGC TCCCAA---- ---AAAGGCT CTTTCAGAG
Maus H3       10  GTGA----- ----GTGAAt ttggac---- CCCAAAGGCT CTTTCAGAG
Ratte H2B     2   ----- CCCACAGAGC TCACAAcaga CCCAAAGGCT CTTTCAGAG
Mensch H2A    45  aagggg---C ACCTGTGAAC TC---A---- ---AAAGGCT CTTTCAGAG

                **** * ***** ***** ***** *****
                                ***** *****
                                ***** *****

Maus H2A      73  CCACCCACtg aaTCAGAT-A AAGAGTTGTG TCACGGT--- -----
Maus H3       46  CCACCCACAT TTTct-ATAA AAG-GCTGTA TATCGATaag cttttataaa
Ratte H2B     42  CCAC----- ----- -----
Mensch H2A    82  CCACCCACGT TTTCAAATAA AAGAGTTGTt aatgctg--- -----

                ***** ***** ***** *****
                *****
                ****

Maus H2A      109 ----- -----
Maus H3       94  cccactcag caactcc
Ratte H2B     46  ----- -----
Mensch H2A    119 ----- -----

```

Die Großbuchstaben innerhalb des Alignments geben an, welche Sequenzabschnitte für die Berechnung des Alignments berücksichtigt wurden. Die Anzahl der Sterne (\*) unterhalb der Zeilen des Alignments ist ein Maß für die Ähnlichkeit der Sequenzen. Der Bereich, der mit je 3 Sternreihen markiert ist, ist eine Region, die ein potentielles regulatorisches Motiv enthalten könnte. In diesem Beispiel sind die Sequenzen innerhalb dieses Bereiches absolut identisch.

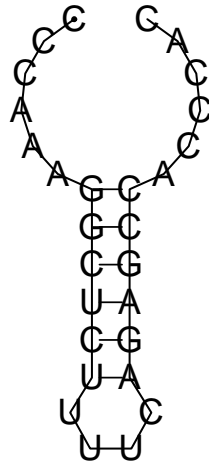


Abbildung 4.3.: Strukturvorhersage (RNAfold) für den durch DIALIGN2 identifizierten sequenzähnlichen Bereich der Histon 3'UTRs (hier gezeigt: Histon 3 der Maus).

### Strukturvorhersage

Der zweite Schritt dieser Strategie ist die Berechnung einer Strukturvorhersage für diese sequenzähnlichen Bereiche, um zu überprüfen, ob diese Bereiche ähnliche Strukturen ausbilden. Für die Strukturvorhersage wurden die in Abschnitt 3.4.3 beschriebenen Programme `mfold` oder `RNAfold` eingesetzt. Da die Sequenzen in dem interessanten Bereich identisch sind, wurde nur für eine Sequenz eine Strukturvorhersage berechnet. Die in der Abbildung 4.3 gezeigte Strukturvorhersage (berechnet mit `RNAfold`) entspricht der Konsensusstruktur des Histon Stemloops, der somit durch diese Strategie identifiziert werden konnte.

### Anwendung der 1. Vorhersagestrategie auf IREs

Aufgrund der sehr hohen Sequenzähnlichkeit der Histon 3'UTRs im Bereich des regulatorischen Stemloops konnte dieser durch Anwendung der ersten Vorhersagestrategie leicht identifiziert werden. Die Testsysteme mit den *Iron Responsive Elements* (vor allem die modifizierten IREs) zeigen deutlich stärkere Sequenzvariationen innerhalb des vorherzusagenden regulatorischen Elementes. Im folgenden wird die Anwendung der ersten Vorhersagestrategie auf 5'UTRs von Ferritin mRNAs aus verschiedenen Organismen beschrieben. Zunächst wurde das Sequenzalignment mit `DIALIGN2` berechnet. Die Positionen des IRE ab dem Cytosin-Bulge sind in blau dargestellt.

## 4. Ergebnisse

Ferritin 5'UTRs: Alignment (DIALIGN format):

```

Mensch      1  c----CAGAC -GUUCUuCGC CGAGAGUCGU CGGGGUUUC UGCUUCAACA
Maus        1  ----CAGAC -GUUCU-CGC CCAGAGUCGC CGCGGUUUC UGCUUCAACA
Schmetterling 1  gagaugguuu uGUUAU-CGC GCUGCGUAGu uuauuuauu u-----
Stier       1  c-----
Schwein     1  ggcacgag--
Mosquito    1  cauuuCAGAC uagua----
          *****
          *****
          *****
          *****
          *****
          *****

Mensch      46  GUGCUUGGAC GGAACCCGGC GCUCGUU--- CCCACCCCG GCCGGCCGCC
Maus        44  GUGCUUGAAC GGAACCCGGU GCUCGACCCC UCCGACCCC GCCGGCCGU
Schmetterling 41  -----
Stier       2  -----U GCUCGUCCCC CCCUACCCCG GCCGGCCACU
Schwein     28  GUGCUUGGAC GGAACCCGGC GCUCGUC--- ACCCGCCCG GCCGGCCGU
Mosquito    16  -----
          *****
          *****
          *****
          *****
          *****
          *****

Mensch      93  CAUAGCC--A GCCCU----
Maus        94  UCGAGCCUGA GCCCU----
Schmetterling 45  UAAAAUCUGA GCCAUcguug cgaguaaccc ugUUUUAGUU UcgUUGCAAg
Stier       33  CAGAGCC--A GCCCU----
Schwein     75  CAGAGCC--A GCCCU----
Mosquito    25  GAUAGaggug gaagcauac gac----- --UGUGAGUU U-----AAC
          *****
          *****
          *****
          *****
          *****
          *****

Mensch      113 CUCUUCACCG CACCCUCGGA CUGCCCAAG GCCCcgCGG CCGCUC CAGC
Maus        116 UUCGUCguu-
Schmetterling 95  aua-----
Stier       52  CACUUGACAG CGCCUCGGA CCGGCCAAG GUCC---CCG CCGCUCAGU
Schwein     89  -----
Mosquito    60  UUCcuaaua u-----
          *****
          *****
          *****
          *****
          *****
          *****

Mensch      163 GCCGC---- -GCAGCCACC GCCGCCCGG CCGCC--UC
Maus        138 GUCGCCACCG CGCCUCgccc CGCGCCACC -----
Schmetterling 98  -----G- ----CCUU- CUGGCCAGU GUGUGUAAAG Gcuagac---
Stier       99  BCCGCucgG- ----CCGU- CGCGCCACC ACCACCACCA CCGCC--CG
Schwein     89  -----
Mosquito    77  GCAGCCAUCG AGUCACCUU- CUGGCCAGU GUGUAUAAAG GuugacaaCG
          *****
          *****
          *****
          *****
          *****
          *****

Mensch      194 UCCUAGUCG CCGcc----
Maus        168 -----
Schmetterling 130 -UUGAAGUCG UCGuacggga -----CUG UCauauaua
Stier       139 UUUUCAGCCG CCCACC----
Schwein     97  UCCUCAGCCU CCCGCC----
Mosquito    126 UUUGAAaacu agugaaagcu gcuaggaau ucuaguuCUG UCuagu----
          *****
          *****
          *****
          *****
          *****
          *****

```

## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

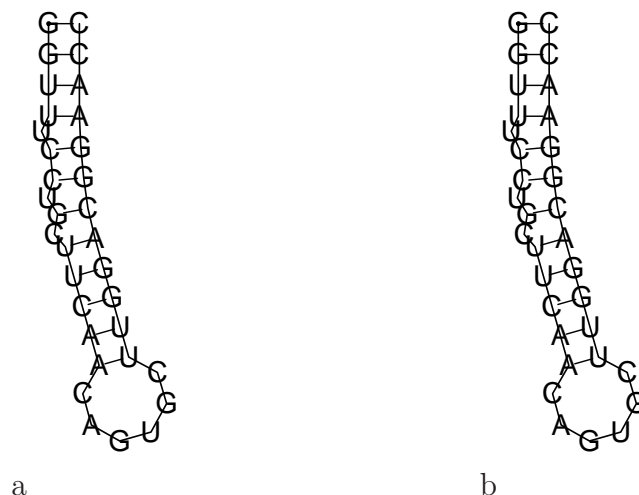


Abbildung 4.4.: Strukturvorhersage für den durch DIALIGN2 identifizierten sequenzähnlichen Bereich der Ferritin 5'UTRs (a: Mensch, b: Schwein).

Mensch	209	-----	-----	-----	-----	-----
Maus	168	-----	-----	-----	-----	-----
Schmetterling	162	aauaaaUAU	UUUuuucucg	uccaaguauc	ugugauuuua	uuguagaaaa
Stier	155	-----	-----	-----	-----	-----
Schwein	113	-----	-----	-----	-----	-----
Mosquito	172	-----UAU	UUUgaaagag	cauucggaaa	ccaguguaac	aagaa-----
**** **						

In diesem Alignment konnte das IRE in der Ferritin 5'UTR von Mensch, Maus und Schwein als ein Bereich mit einer besonders hohen Sequenzähnlichkeit (markiert durch 3 Sternreihen) gefunden werden. Obwohl das IRE in der 5'UTR von Schmetterling und Mosquito korrekt aligniert wurde, ist die Bewertung dieses Bereiches (nur eine Sternreihe) zu schwach, um als eindeutig definiertes Element erkannt zu werden. In der Ferritin 5'UTR des Stiers ist zwar die Loopsequenz des IRE vorhanden, die angrenzenden Basen können allerdings die Helix nicht korrekt ausbilden (möglicherweise Fehler in der Datenbanksequenz, da das IRE eigentlich vorhanden sein müsste). Die Loopsequenz alleine reicht nicht aus, um mit der Loopsequenz des IREs einer anderen UTR aligniert zu werden. Ausgehend von dem Alignment würde man die IREs in den Ferritin 5'UTRs der beiden Insekten zunächst übersehen. Durch die in Abschnitt 4.3.2 beschriebene Mustersuche können sie aber dennoch gefunden werden.

Die Abbildung 4.4 zeigt eine Strukturvorhersage für den sequenzähnlichen Bereich der Ferritin 5'UTRs von Mensch und Schwein. Somit findet man neben der Sequenz auch die korrekte Struktur des IREs.

#### 4. Ergebnisse

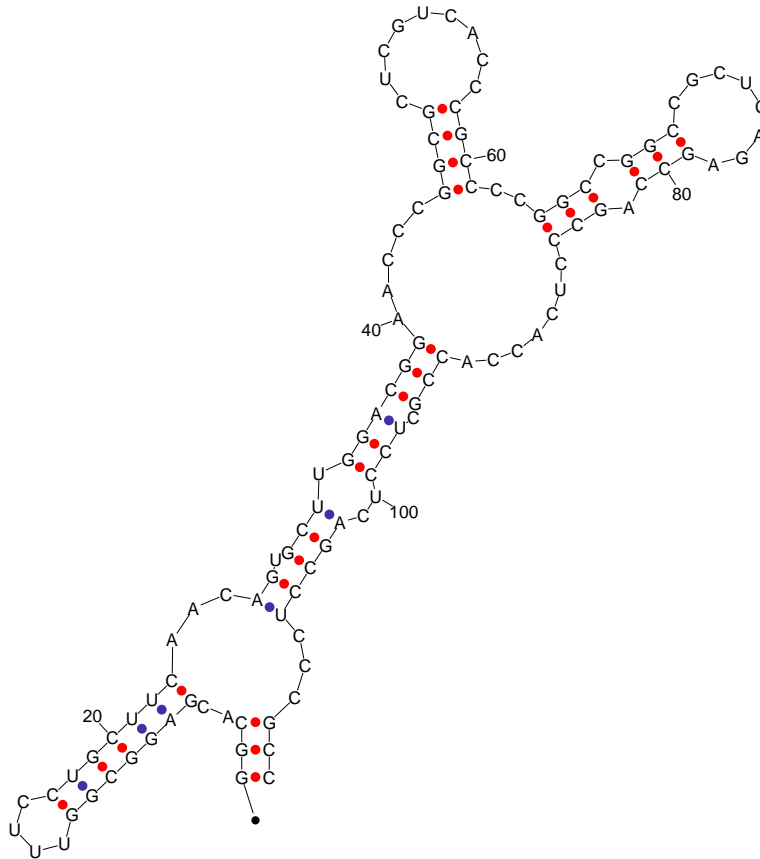


Abbildung 4.5.: Strukturvorhersage für die Ferritin 5'UTR des Schweins. Energetisch optimale Struktur mit  $dG = -38.8$  kcal/mol.

Um zu überprüfen, ob ein auf diese Weise vorhergesagtes Motiv überhaupt in einer UTR vorkommen kann, muß getestet werden, wie ein solches Motiv die Energie der Struktur der gesamten UTR ändert. Dazu wird zunächst die Struktur der UTR vorhergesagt. Wenn das vorhergesagte Motiv nicht in der energetisch günstigsten Struktur vorkommt, wird es bei einer erneuten Strukturvorhersage erzwungen und der so berechnete Energiewert mit dem der energetisch günstigsten Struktur verglichen.

In der Abbildung 4.5 ist die mit `mfold` berechnete energetisch günstigste Struktur (-38.8 kcal/mol) der Ferritin 5'UTR des Schweins dargestellt, wobei das IRE in dieser Struktur nicht als Motiv vorkommt. Zwingt man das IRE bei der Strukturvorhersage in die Struktur, ergibt sich die in Abbildung 4.6 gezeigte Struktur (-32.9 kcal/mol). Somit führt das IRE zu einer Struktur, deren Energiewert ca. 15 % schlechter ist als der Energiewert der energetisch günstigsten Struktur. Dies ist ein gutes Beispiel dafür, daß die energetisch günstigste Struktur nicht die

## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

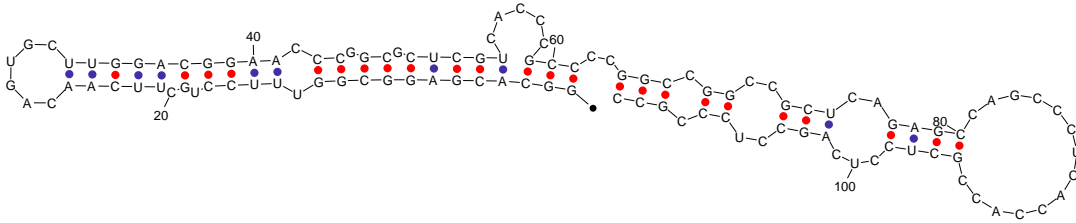


Abbildung 4.6.: Strukturvorhersage für die Ferritin 5'UTR des Schweins, mit IRE (dG=-32.9 kcal/mol). Der IRE Stemloop bildet den linken Teil der Struktur.

biologisch relevante Struktur sein muß. Dieses Phänomen wird vor allem bei der zweiten Vorhersagestrategie eine entscheidende Rolle spielen.

Die Grenzen der ersten Vorhersagestrategie werden bei der Anwendung des Testsystems mit den modifizierten IREs deutlich. Bei diesen Sequenzen ist nur noch eine sehr geringe Sequenzähnlichkeit innerhalb des IRE vorhanden. Nachfolgend ist das DIALIGN2 Alignment für einige dieser Sequenzen gezeigt, wobei die Positionen des IRE wieder blau hervorgehoben sind.

```

modifizierte 5'UTRs: Alignment (DIALIGN format):

seq_AF068224      1  -----  -----  -----  -----  -----
seq_AF161708     1  guuauccggu gu-----  -----  -----  -----
seq_D15071.1     1  gccc-----  -----  -----  -----  -----
seq_D86626.1    1  aa-----  -----UCCU  AAUGCCGAAA  CUUACACAUU
seq_M12120.1    1  acccccucuu  cgacucuuga  cgagcgUCUU  AUUACUCAAC  AACAGACAUU

                        ****  *****  *****

seq_AF068224     1  -----U  UAGCgc----  -----  -----  -----
seq_AF161708    13  ----AAAAGG  UAGUAUCGUC  A-----  -----  -----
seq_D15071.1     5  -----GG  CCGCGACGCC  A-----  -----  -----
seq_D86626.1    28  AUAAAAAAGU  UAGCuccgaa  ccgccugcgu  ucuaaaggaa  cccuaacaua
seq_M12120.1    51  CCAAAccuug  uc-----  -----  -----  -----

                        *****  *****  *

seq_AF068224     8  -----  -----  -----  -----  -----
seq_AF161708    30  -----  -----  AGCGAGAUGU  UCGUUAGAUC  UUUUUGUUCG
seq_D15071.1    18  -----  -----  CCCGA-----  -----  -----CG
seq_D86626.1    78  auuucuccug  cuuucuaauc  ACCAAGAUGU  AUGAGACAUG  AGGAUAUUCG
seq_M12120.1    63  -----  -----  -----  -----  -----

                        *****  *****  *****
**

```

#### 4. Ergebnisse

```

seq_AF068224      8  -----  -----  -----  -----ACA AAg-----
seq_AF161708     60  GCUCCaagag  cgaguugaua  uuuaguguug  aauuagaACA  AAuguuaaug
seq_D15071.1     25  GCUCCcuccc  g-----  -----  -----  -----
seq_D86626.1    128  Gacg-----  -----  -----  -----  -----
seq_M12120.1     63  -----  -----  -----  -----  -----

          *****                               *** **

seq_AF068224     14  -----  -----  -----  -----  -----
seq_AF161708    110  aaagaauuaa  UUUUGAGGUC  AAAACUGCGA  AUUAAUUUUA  UCUUuccgcc
seq_D15071.1    36  -----  -----  -----  -----  -----
seq_D86626.1   132  -----  UGUUCAGUGA  AACACUUCGG  AGAAUUUUA  ACAU-----
seq_M12120.1    63  -----  -----  -----  -----  -----

          ***** ***** ***** *****

seq_AF068224     14  -----  -CGCUGCUCU  CUCAUGCAGA  A-UUGAACUC  UGAAAGCUAG
seq_AF161708    160  guUACUUGUU  UCGCUCAUUG  CUAACGUUUU  A-UCGUAUGC  UGUAAGCUCG
seq_D15071.1    36  -----  -----  ---GCAGUGC  GgUCCUGGCC  AGUGAGCUGG
seq_D86626.1   166  -----  -----  -----  -----  -----
seq_M12120.1    63  --UACCCUU  UUGCUGAUUG  CAAGCAGUGC  G-----  -----

          ***** ***** ***** * ***** *****
          ***** ***** ***** * ***** *****
          ***** *****

seq_AF068224     52  u--CUCAAG  Auaaug---  UUCGCUCGGU  Caggagagcu  gacggacuug
seq_AF161708    209  a-----  -----  -----  -----  -----
seq_D15071.1    63  gccgCUCAGG  AgcccccAAC  UUCGCUCGCU  C-----  -----
seq_D86626.1   166  -----  -----AAC  UUa-----  -----  -----
seq_M12120.1    92  -----  -----  -----  -----  -----

          ***** *          *** ***** *
          ***** *****

seq_AF068224     95  gacacccucc  uuuucccgcg  guUCCGCagc  g-----  -----
seq_AF161708    210  -----  -----  -----  -----  -----
seq_D15071.1    94  -----  -----  --UCCGCCCA  GUCGCGGCAC  AGGCCc---
seq_D86626.1   172  -----  -----  -----  -----  -----
seq_M12120.1    92  -----  -----  --CGGCGCA  UGCACGUGAC  AUGCCugcgc

          ***** ***** *****

seq_AF068224     126  -----  -----  -----
seq_AF161708    210  -----  -----  -----
seq_D15071.1    118  -----  -----  -----
seq_D86626.1    172  -----  -----  -----
seq_M12120.1    119  uacuaaaau  uacuuccuu  gaauuuuga

```

Hier ist nur noch ein sehr kleiner Bereich durch 3 Sternreihen als besonders ähnlich markiert. Dieser Bereich entspricht jedoch nicht dem IRE, das in der Sequenz AF068224 um Position 75, in der Sequenz AF161708 um Position 60, in der Sequenz D15071.1 um Position 50, in der Sequenz D86626.1 um Position 135 und in der Sequenz M12120.1 um Position 100 zu finden ist.



### Strategiebewertung

Die Anwendung der Testsysteme hat gezeigt, daß diese Strategie zur Vorhersage potentiell regulatorischer Struktur motive dann funktioniert, wenn das regulatorische Motiv auf Sequenzebene konserviert ist. So ließen sich die hoch konservierten Stemloop Strukturen in den Histon 3'UTRs finden, während bei den Ferritin 5'UTRs nur noch drei von fünf durch das Sequenzalignment mit DIALIGN2 vorhergesagt werden konnten. Die Ursache hierfür liegt darin, daß das IRE in den Ferritin 5'UTRs der beiden Insekten sehr viel näher Richtung 3'-Ende lokalisiert ist. Unter diesen Bedingungen lassen sich diese IREs mit DIALIGN2 jedoch nicht detektieren.

Der Vorteil dieser Vorhersagestrategie liegt darin, daß sie einfach und effizient ist, da auch große Sequenzen verarbeitet werden können, wobei ein potentielles regulatorisches Motiv jedoch an ähnlichen Positionen innerhalb der UTR lokalisiert sein muß. Wie die Anwendung des Testsystems mit den modifizierten IREs zeigt, lassen sich Motive mit sehr geringer Sequenzkonservierung mit dieser Strategie nicht mehr finden. Dieser Mangel wird bei der nächsten Vorhersagestrategie umgangen.

### 4.2.2. Strategie 2: Strukturvorhersage mit nachfolgendem Strukturalignment

Die Anwendung dieser Strategie soll es ermöglichen, potentielle regulatorische Motive in UTRs vollkommen unabhängig von der Sequenz zu finden. Das Prinzip dieser Strategie basiert auf der Berechnung eines paarweisen lokalen Strukturalignments oder eines multiplen globalen Strukturalignments von vorher bestimmten UTR Strukturen. Für die Berechnung und Visualisierung der Strukturalignments wurde das in Abschnitt 3.4.4 beschriebene Programm *RNAforester* verwendet.

#### Paarweise lokale Strukturalignments mit *RNAforester*

Zunächst wird die Visualisierung von paarweisen RNA Strukturalignments mit *RNAforester* anhand eines Beispielles dargestellt.

Für ein Alignment A zweier Strukturen  $S_1$  und  $S_2$  wird eine RNA Sekundärstruktur  $S_2$  an  $S_1$  gezeichnet, die die Unterschiede als Abweichungen der Struktur  $S_2$  von  $S_1$  hervorhebt. Umgekehrt kann auch  $S_1$  an  $S_2$  gezeichnet werden, wobei beides alternative Visualisierungen desselben Alignments A sind. Dabei können in den Zeichnungen sämtliche Informationen des berechneten Alignments, wie zum Beispiel alternative Basenpaarungen, dargestellt werden.

Die Visualisierung von Strukturalignments mit *RNAforester* wird im folgenden anhand zweier Transfer-RNAs (tRNAs) aus *E. coli* demonstriert. In Abbildung 4.7 sind die Sekundärstrukturen für die Alanin-tRNA (a) und die Leucin-tRNA (b) dargestellt. Die Abbildung 4.8 zeigt das Alignment der beiden tRNA Strukturen

#### 4. Ergebnisse

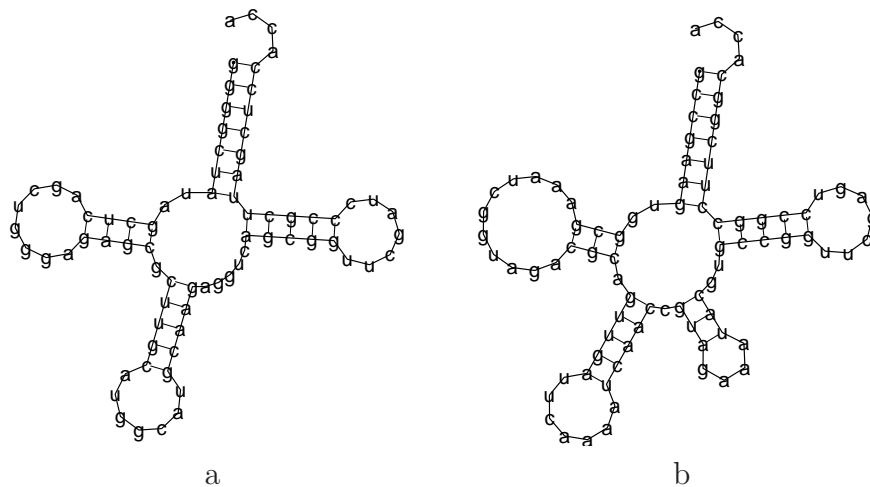


Abbildung 4.7.: Sekundärstrukturen der *E. coli* tRNAs für Alanin (a) und Leucin (b) (Strukturen wurden aus der Genomischen tRNA-Datenbank entnommen (Lowe und Eddy, 1997)).

als sogenannten Squiggle-plot in der Form  $S_{Ala}$  an  $S_{Leu}$ . Die schwarz gezeichneten Bereiche markieren hierbei Strukturelemente, die in beiden Strukturen mit derselben Sequenz vorkommen, wobei rot gezeichnete Bereiche solche Elemente darstellen, bei denen die Sequenz in beiden Strukturen unterschiedlich ist. Das durch die gestrichelte blaue Linie gezeigte Basenpaar kommt nur in der Alanin tRNA vor, während der grün gezeichnete Stemloop nur in der Leucin tRNA zu finden ist.

Mit dieser Vorhersagestrategie soll zunächst das IRE als gemeinsames regulatorisches Element in der Ferritin 5'UTR des Menschen und der Succinatdehydrogenase 5'UTR von *Drosophila melanogaster* nachgewiesen werden. Das folgende DIALIGN2 Sequenzalignment zeigt, daß nur die Loopsequenz des IRE korrekt aligniert worden ist, während die angrenzenden Bereiche, die die Helix ausbilden, aufgrund starker Sequenzunterschiede nicht aligniert wurden (IRE wieder in blau).

Ferritin 5'UTR (Mensch), Succinatdehydrogenase 5'UTR (Drosophila)  
Alignment (DIALIGN format):

Mensch	1	ccagacguuc uucgccgaga gucgucggG UUUCUGCuu caa-----
Drosophila	1	ca----- -----G UUUCGUGCaa cuuuuuguac
		* *****
		* *****
		* *****
Mensch	44	----- ----- -----
Drosophila	24	gcaaaauaga aaaacauuaa auuugcucuc agcaaaucga uaaauugcaaa

## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

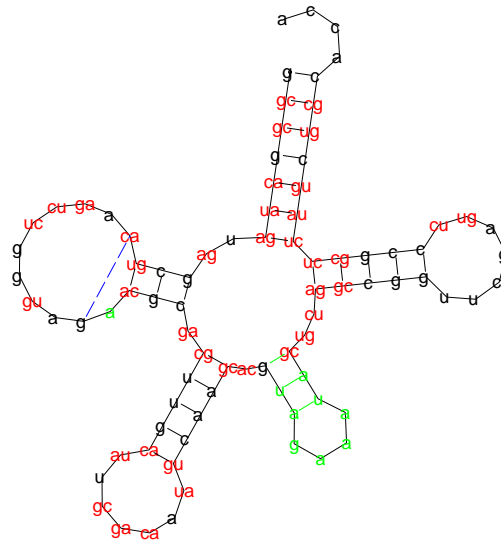


Abbildung 4.8.: Strukturalignment der Alanin und Leucin tRNAs. Farbkennung: schwarz: Strukturelement in beiden Strukturen mit derselben Sequenz; rot: Strukturelement in beiden Strukturen mit unterschiedlicher Sequenz; blau: Strukturelement nur in  $S_{Ala}$ ; grün: Strukturelement nur in  $S_{Leu}$ .

```

Mensch          44  --CAGUGCuu ggacggaacc cggcgucug uccccacccc ggccggccgc
Drosophila      74  cgCAGUGCcg uuucaauugc agcaca--- -----
                *****

Mensch          92  ccauagccag ccucuccguca ccucucACC GCAcccucgg acugcccac
Drosophila     101  ----- -----ACC GCAacgaaa-----
                ***  ***

Mensch         142  ggccccccgc gccgcuccag cggcgcgag ccaccgccgc cgccggccgc
Drosophila    113  ----- ----- -----

Mensch         192  ucuccuagu cgccgcc
Drosophila    113  ----- -----

```

Für die zweite Strategie mußten zunächst die Strukturen der beiden UTRs mit `mfold` berechnet werden. In Abbildung 4.9 ist eine Struktur der Ferritin 5'UTR des Menschen und in Abbildung 4.10 eine Struktur der Succinatdehydrogenase

#### 4. Ergebnisse

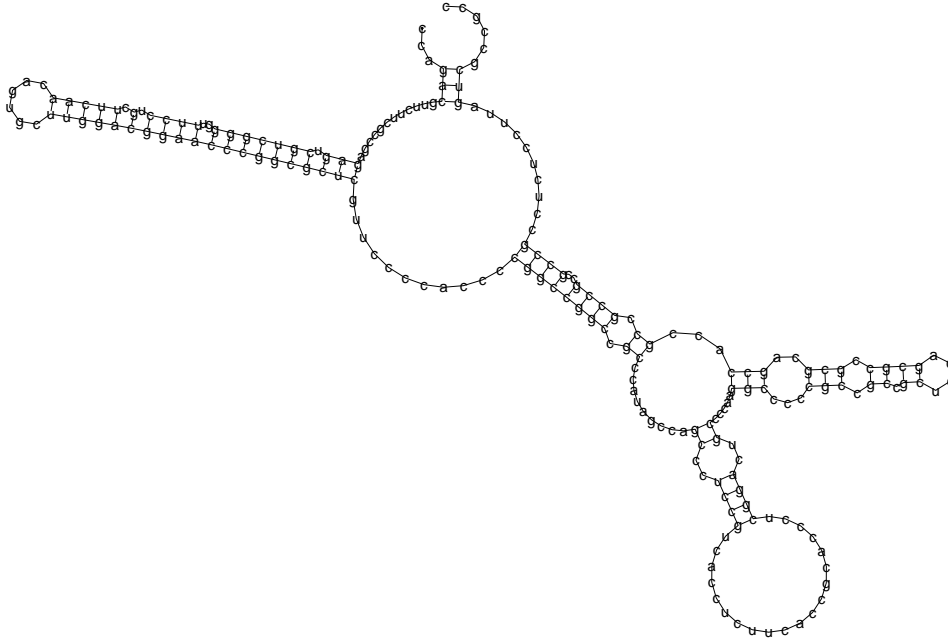


Abbildung 4.9.: Strukturvorhersage für die Ferritin 5'UTR des Menschen. Energetisch optimale Struktur mit  $dG=-59,9$  kcal/mol.

5'UTR von *Drosophila* gezeigt, wobei es sich bei beiden Strukturen um die energetisch günstigsten handelt. Da jedoch, wie bereits erwähnt, die energetisch günstigste Struktur nicht zwangsläufig die biologisch relevante ist, die das regulatorische Motiv enthält, wurden bei dieser Strategie immer mehrere suboptimale Strukturen aligniert.

## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

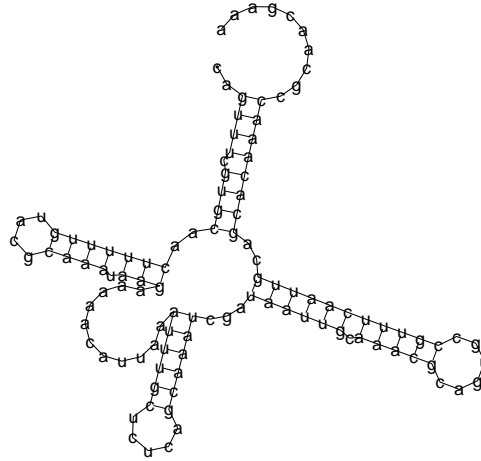


Abbildung 4.10.: Strukturvorhersage für die Succinatdehydrogenase 5'UTR von *Drosophila melanogaster*. Energetisch optimale Struktur mit  $dG = -21,1$  kcal/mol.

Für die Berechnung paarweiser Strukturalignments wurden im Rahmen dieser Arbeit folgende Score-Werte als Parameter verwendet:

Basenpaarerhaltung oder Ersetzung	Score = 4
Basenpaarlöschung oder Einfügung	Score = -3
Basentreffer	Score = 0
Basenersetzung	Score = 0
Basenlöschung oder Einfügung	Score = -5

Durch die Wahl dieser Parameter wurden ausschließlich Strukturmerkmale für die Berechnung des Score berücksichtigt. Da für die beiden Parameter Basentreffer und Basenersetzung der Wert 0 angegeben wurde, spielt die Primärsequenz bei der Berechnung des Alignments keine Rolle. Sowohl das Einfügen bzw. Löschen eines Basenpaares oder einzelner Basen wurde stark negativ gewichtet.

Zunächst wurde ein globales Strukturalignment mit `RNAforester` berechnet, das in Abbildung 4.11 in der Form  $S_{Mensch}$  an  $S_{Drosophila}$  dargestellt ist. Das IRE konnte in diesem globalen Alignment nicht als gemeinsames Strukturmotiv bestimmt werden, da es in der Ferritin 5'UTR an einer anderen Position lokalisiert ist als in der Succinatdehydrogenase 5'UTR, und die Kosten für die Löschung sämtlicher Basen vor dem IRE viel zu hoch wären, um ein optimales globales Alignment zu ergeben.

Statt eines globalen Alignments wurde daher ein lokales Strukturalignment mit

#### 4. Ergebnisse

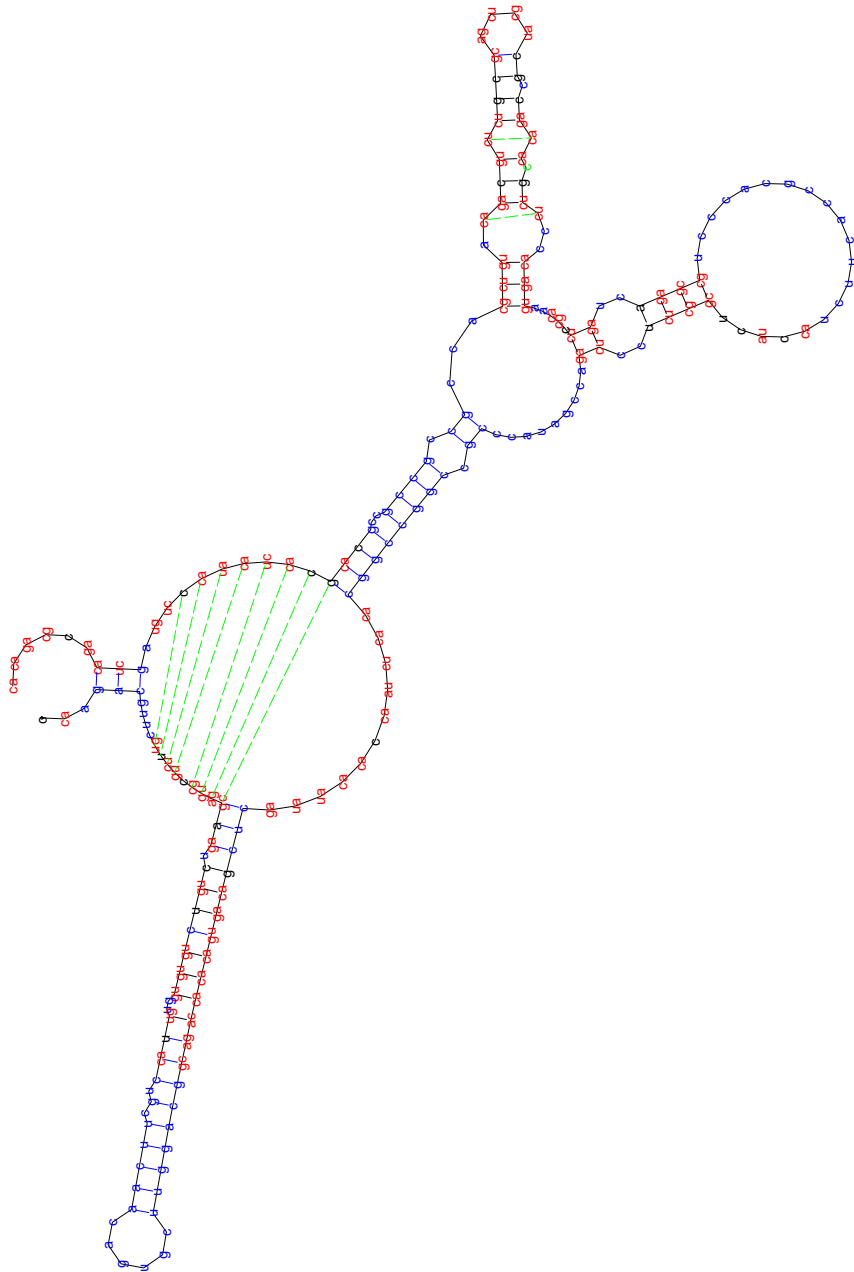


Abbildung 4.11.: Globales Strukturalignment für die Ferritin 5'UTR des Menschen und die Succinatdehydrogenase 5'UTR von *Drosophila melanogaster*.

## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

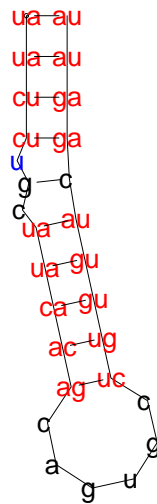


Abbildung 4.12.: Lokales Strukturalignment für die Ferritin 5'UTR des Menschen und die Succinatdehydrogenase 5'UTR von *Drosophila melanogaster*.

RNAforester berechnet. Die Score-Parameter wurden dabei nicht geändert. In der Abbildung 4.12 ist das lokale Strukturalignment für die beiden UTR-Strukturen gezeigt. Dieses liefert das *Iron Responsive Element* als ähnlichstes Strukturelement in den beiden Strukturen. Während das Motiv mit der ersten Vorhersagestrategie aufgrund der geringen Sequenzkonservierung (nur der Loop und der Cytosin-Bulge sind konserviert) nicht entdeckt wurde, konnte es mit der zweiten Strategie, die unabhängig von der Sequenz funktioniert, vorhergesagt werden.

Um die Flexibilität dieser Strategie auszutesten, wurde ein Strukturalignment der Ferritin 5'UTR und Transferrinrezeptor 3'UTR des Menschen berechnet. Während die Ferritin 5'UTR nur 208 Nukleotide lang ist, ist die Transferrinrezeptor 3'UTR mit 2464 Nukleotiden sehr viel länger. Diese UTR enthält insgesamt 5 *Iron Responsive Elements*. In Abbildung 4.13 ist eine Strukturvorhersage für die Transferrinrezeptor 3'UTR gezeigt. Da Vorhersagen für die Gesamtstruktur solcher großer Sequenzen nicht mehr verlässlich sind, kann es bei diesem Alignment nur darum gehen, kleinere Strukturelemente zu betrachten. Daher wurde auch nur ein lokales Alignment für die beiden Strukturen berechnet. In Abbildung 4.14 ist das Ergebnis dieses lokalen Strukturalignments gezeigt. Obwohl das IRE auch hier nur sehr wenig Sequenzkonservierung aufweist und daher mit der ersten Strategie nicht entdeckt werden konnte, ließ es sich mit der zweiten Strategie als ähnlichstes Strukturelement in den beiden UTRs vorherhersagen.

#### 4. Ergebnisse

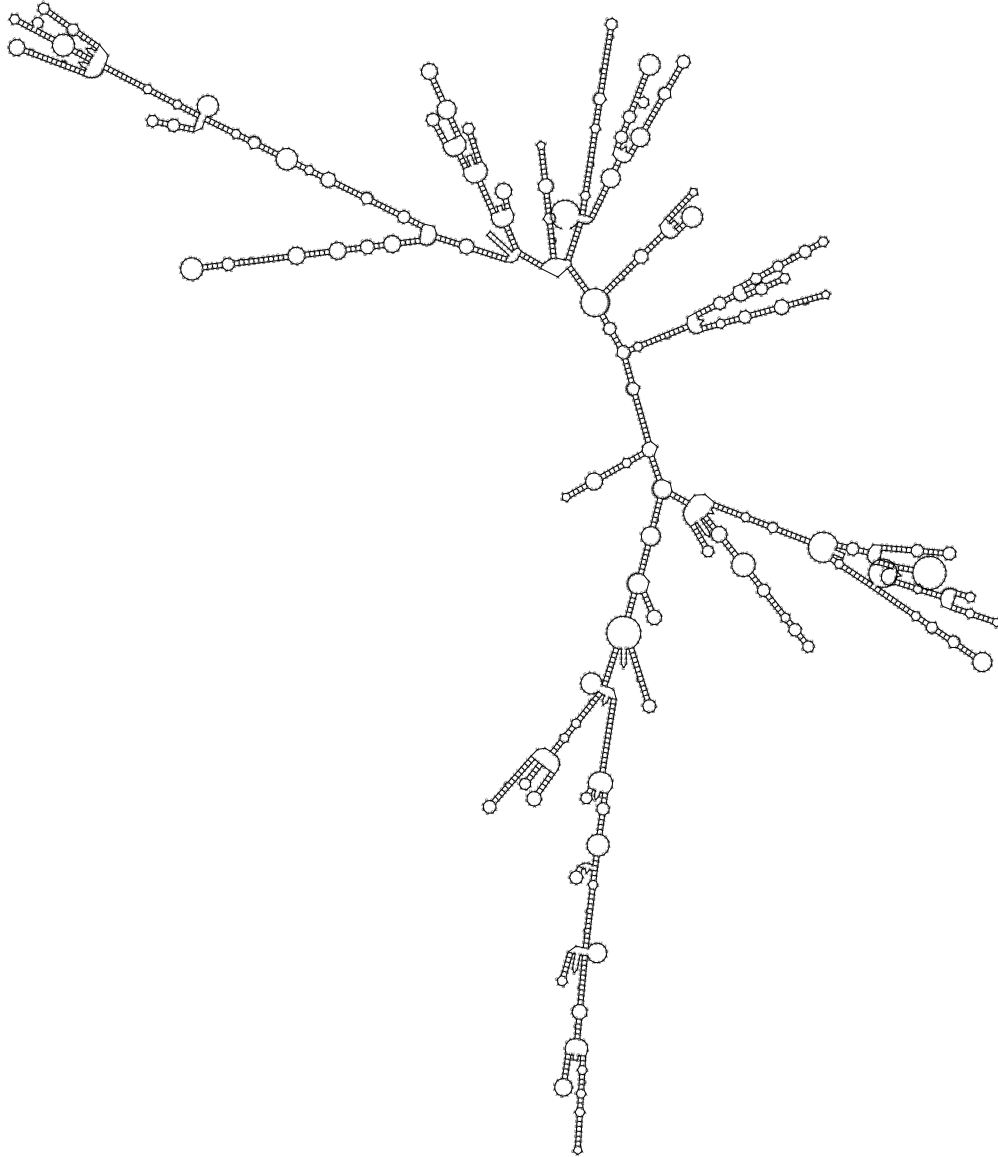


Abbildung 4.13.: Strukturvorhersage für die Transferrinrezeptor 3'UTR des Menschen.



## 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

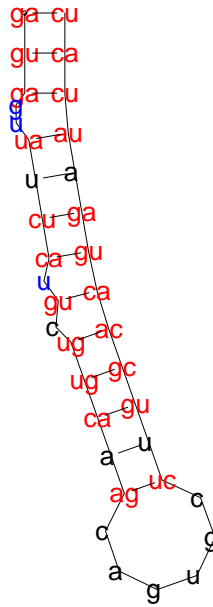


Abbildung 4.14.: Lokales Strukturalignment der Ferritin 5'UTR (208 Nucleotide) und der Transferrinrezeptor 3'UTR (2464 Nucleotide) des Menschen.

Im folgenden wird die Anwendung der Strategie auf das Testsystem der modifizierten IREs exemplarisch an zwei Sequenzen dargestellt, die unterschiedliche Längen in der Loopsequenz haben. In der Abbildung 4.15 sind die beiden Strukturen gezeigt, die für die Berechnung des Strukturalignments verwendet wurden. Das in der Abbildung 4.16 gezeigte lokale Alignment lieferte das IRE, wobei die unterschiedliche Looplänge durch das zusätzliche, blau gezeichnete Nucleotid dargestellt wird. Für dieses Ergebnis mußten mehrere suboptimale Strukturen aligniert werden.

Die Strategie der Strukturmotivvorhersage durch die Berechnung paarweiser Strukturalignments mit *RNAforester* wurde in (Höchsmann et al., 2003) veröffentlicht.

#### 4. Ergebnisse

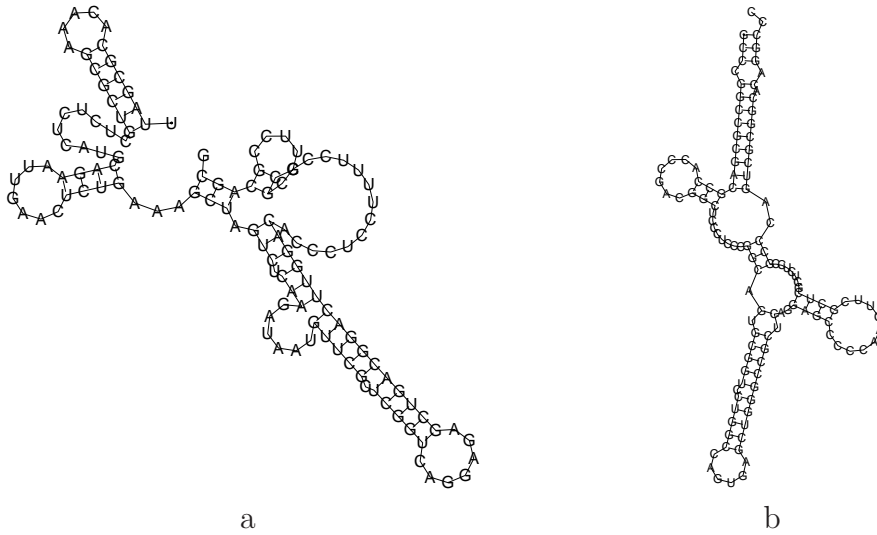


Abbildung 4.15.: Sekundärstrukturen der modifizierten UTR Sequenzen AF068224.1 (a) und D15071.1 (b).

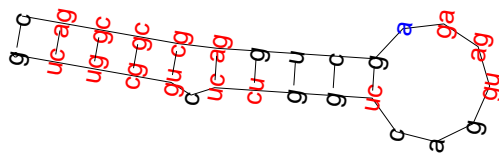


Abbildung 4.16.: Lokales Strukturalignment der modifizierten UTR Sequenzen AF068224.1 und D15071.1. Das IRE wurde trotz der unterschiedlichen Looplänge (blau: zusätzliches Nukleotid) gefunden.

### Multiple globale Strukturalignments mit RNAforester

In diesem Abschnitt wird die Motivvorhersage durch die Berechnung multipler globaler Strukturalignments beschrieben. Der Vorteil im Vergleich zur Verwendung paarweiser lokaler Strukturalignments liegt darin, daß man zum einen Strukturen vieler verschiedener UTRs (oder der gleichen UTR aus verschiedenen Organismen), und gleichzeitig viele suboptimale Strukturen vergleichen kann. Für die Darstellung der Ergebnisse eines multiplen RNA-Sekundärstrukturalignments wurde eine spezielle Visualisierung entwickelt. Diese ermöglicht, wie bei den paarweisen Strukturalignments, eine komfortable Auswertung der Ergebnisse.

Die Visualisierung wird anhand eines Strukturalignments der 20 tRNAs aus *E.coli* demonstriert. Dabei gibt es verschiedene Möglichkeiten der Darstellung. In Abbildung 4.17 zeigen rot dargestellte Basen bzw. Bindungen ein Vorkommen der betreffenden Elemente (Basen bzw. Bindungen) in allen alignierten Strukturen an. Elemente, die nicht in allen Strukturen zu finden sind, werden in Grau dargestellt, wobei die Abstufungen von dunkel nach hell ein Maß für die Anzahl der Strukturen sind, in denen das jeweilige Element vorkommt. Bei den angegebenen Basen handelt es sich jeweils um die, die am häufigsten an der jeweiligen Position vorkommt. Bei der in Abbildung 4.18 gezeigten Darstellung wird an jeder Position die Häufigkeit des Vorkommens jeder einzelnen Base durch den Durchmesser eines Kreises symbolisiert. Die einzelnen Basen sind durch Kreise unterschiedlicher Farbe gekennzeichnet und bilden jeweils die vier Ecken eines Quadrates. Schwarze Kreise innerhalb dieser Quadrate symbolisieren Lücken (gaps). Diese Art der Darstellung erlaubt es, für jede einzelne Position die vorkommenden Basen zu erkennen. Dies ist in der ersten Darstellungsart nicht möglich, da hier nur die häufigste Base an jeder Position angegeben wird.

Die Strategie zur Motivvorhersage mit einem multiplen Strukturalignment wurde ebenfalls mit Hilfe des *Iron Responsive Elements* entwickelt. In der Abbildung 4.19 ist das multiple Strukturalignment der Ferritin 5'UTRs und der SLC11A3 Eisentransporter 5'UTRs von Mensch und Maus dargestellt. Hierbei wurden die mit `mfold` berechneten optimalen Strukturen der vier UTRs aligniert. Die vielen hellgrauen Bereiche zeigen, daß die alignierten Strukturen global recht unterschiedlich sind. Die einzige Ausnahme ist das IRE, das in allen Strukturen vorkommt und deshalb als rotes Strukturelement in dem Alignment erkennbar ist. Somit stellt das multiple Strukturalignment eine gute Alternative zu dem paarweisen lokalen Strukturalignment dar, da mit dem multiplen globalen Alignment ebenfalls lokale Struktur motive gefunden werden können.

#### 4. Ergebnisse

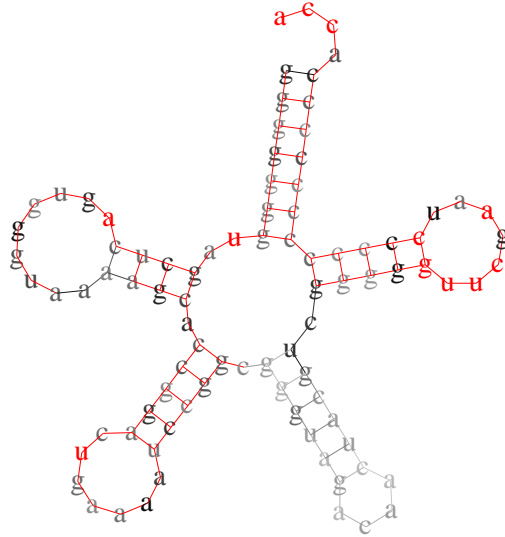


Abbildung 4.17.: Multiples Strukturalignment der 20 tRNAs aus *E.coli*. Rot dargestellte Basen und Bindungen kommen in allen alignierten Strukturen vor.

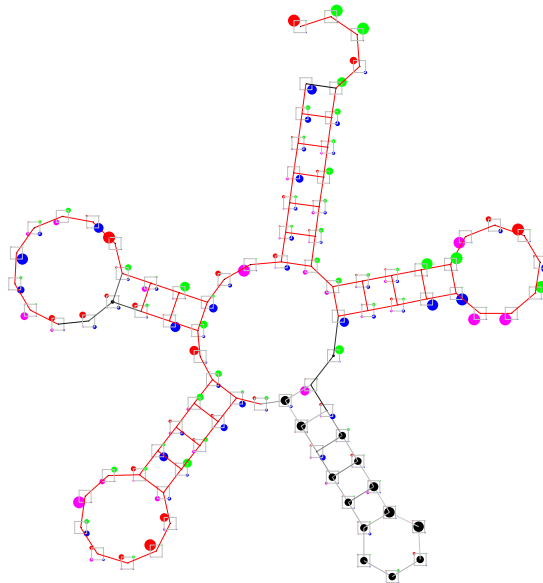


Abbildung 4.18.: Alternative Darstellung des multiplen Strukturalignments der 20 tRNAs aus *E.coli*. Die Häufigkeit des Vorkommens jeder einzelnen Base wird an jeder Position durch die verschiedenfarbigen Kreise symbolisiert: Adenin: rot, Cytosin: grün, Guanin: blau und Uracil: magenta.

#### 4.2. Strategien zur Vorhersage von RNA-Strukturmotiven

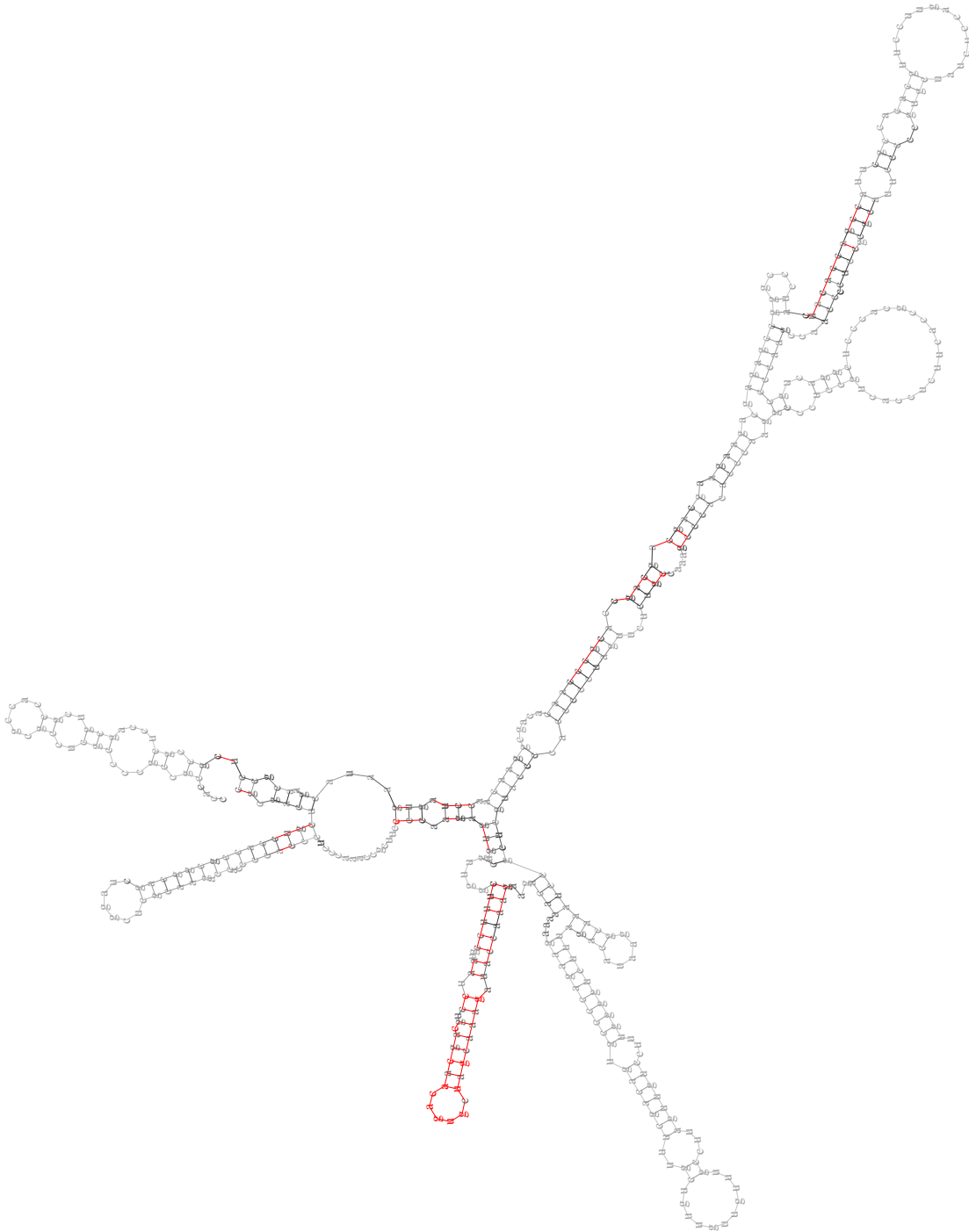


Abbildung 4.19.: Multiples globales Strukturalignment der Ferritin (5HSA015337, 5MMU002159) und SLC11A3 Eisentransporter (5HSA023193, 5MMU011005) 5'UTRs von Mensch und Maus. Das IRE ist in rot dargestellt, da es in allen alignierten Strukturen vorkommt.

##### **Strategiebewertung**

Durch die Anwendung verschiedener Testsysteme konnte die Flexibilität dieser Strategien zur Motivvorhersage in UTRs nachgewiesen werden. Die Beschränkung auf Strukturelemente ermöglicht es, Motive vorherzusagen, die nur sehr gering oder gar nicht auf Sequenzebene konserviert sind. Dennoch lassen sich durch die Art der Visualisierung der Strukturalignments auch kompensatorische Basenaustausche erkennen, die wichtige Hinweise auf das tatsächliche Vorhandensein solcher Motive liefern. Der Erfolg der Strategie hängt stark von den im ersten Schritt durchzuführenden Strukturvorhersagen ab. Wie in Abschnitt 5.1.2 ausführlich diskutiert wird, kann der Erfolg der Strategie durch die Betrachtung suboptimaler Strukturen gesteigert werden.

Durch das Strukturalignment der Ferritin 5'UTR mit der Transferrinrezeptor 3'UTR konnte gezeigt werden, daß mit dieser Strategie auch in sehr großen Strukturen lokale Struktur motive gefunden werden können. Damit ist diese Strategie sämtlichen anderen Strategien zur Vorhersage potentieller regulatorischer Motive in diesem Punkt klar überlegen (s. Diskussion).

## 4.3. Validierung durch Musterdefinition

Die Vorhersage von Strukturmotiven in UTRs ist der erste Schritt zur Entdeckung neuer regulatorischer Motive für die posttranskriptionelle Genregulation. Durch die in den nächsten Abschnitten beschriebenen Methoden zur Musterdefinition, Musterbeschreibung und Datenbanksuche erfolgt eine Validierung der Vorhersagen. Dabei sollen ausreichende Informationen gesammelt werden, um in einer anschließenden Laborphase die tatsächliche biologische Bedeutung eines neu entdeckten Strukturmotives zu bestimmen.

Die Beschreibung eines Musters erfolgt, um Datenbanken zu durchsuchen. Damit soll überprüft werden, ob ein vorhergesagtes oder bereits bekanntes Motiv auf bestimmte Gene beschränkt ist und somit eine exklusive Regulationsfunktion für dieses Gen übernehmen kann. Eine Datenbanksuche kann jedoch auch Hinweise auf die tatsächliche Funktion eines Motivs liefern, für den Fall, daß das Motiv in anderen Genen gefunden wird, für die bereits ein posttranskriptioneller Regulationsmechanismus bekannt ist. Wird ein bereits bekanntes Motiv in mRNAs gefunden, für die bis dahin keine durch das Motiv vermittelte Regulation bekannt war, kann experimentell überprüft werden, ob diese neu gefundenen mRNAs durch das Motiv reguliert werden. Die Datenbanksuche nach kombinierten Sequenz- und Strukturmotiven stellt somit eine vielfältige und mächtige Methode zur Untersuchung der posttranskriptionellen Genregulation dar.

Musterdefinition und Datenbanksuche werden in den folgenden Abschnitten am Beispiel des *Iron Responsive Elements* beschrieben.

### 4.3.1. Musterdefinition durch Signifikanzberechnungen

In der vorliegenden Arbeit werden die Muster, ausgehend von den vorhergesagten Motiven, mit Hilfe von Signifikanzberechnungen definiert (Abschnitt 3.4.5). Dazu muß das Muster zunächst durch eine Grammatik beschrieben werden. Im folgenden ist eine, im Vergleich zu der in (Meyer und Giegerich, 2002) veröffentlichten, vereinfachte Grammatik zur Beschreibung der in Abbildung 4.1 gezeigten Konsensusstruktur des *Iron Responsive Element* angegeben.

```
IRE_cons inp = axiom lcomps where
  lcomps      = sadd <<< base -~~ lcomps   |||   rcomps      ... h
  rcomps      = adds <<< rcomps  ~~- base   |||   p stack1   ... h
  stackscheme r = (sr <<< base -~~ r  ~~- base)  'with' basepairing
  stack1      = tabulated (rep 3 stackscheme cbulge)
  cbulge      = bulge <<< iupac "c" -~~ p stack2
  stack2      = tabulated (rep 4 stackscheme hairpin)
  hairpin     = hl <<< base -~~ iupac "cagugh" ~~- base)
                                     'with' basepairing
```

#### 4. Ergebnisse

Die verwendete ADP-Notation wird ausführlich in (Meyer und Giegerich, 2002) erklärt. Die einzelnen Elemente der Konsensusstruktur sind hier in den sogenannten Parsern "stack1", "cbulge", "stack2" und "hairpin" definiert. "stack1" und "stack2" sind die basengepaarten Bereiche, wobei die untere Helix drei und die obere fünf Basenpaare enthält. In der Grammatik wird das durch die wiederholte Anwendung des Konstruktes "stackscheme" beschrieben. Die Anzahl der Wiederholungen, und damit die Länge der Helix wird durch Anwendung des Konstrukts "rep" festgelegt. Das letzte Basenpaar der oberen Helix ist dabei in dem Parser "hairpin" enthalten, weshalb im Parser "stack2" das Konstrukt "stackscheme" nur viermal wiederholt wird.

Der Vorteil der Algebraischen Dynamischen Programmierung liegt darin, daß durch die Anwendung verschiedener Auswertungsalgebren auf die das Muster beschreibende Grammatik verschiedene Fragestellungen beantwortet werden können. Die in (Meyer und Giegerich, 2002) beschriebene "counting algebra" und "expectation algebra" sind zwei Beispiele für solche Auswertungsalgebren. Durch die Anwendung der "counting algebra" kann das Vorkommen des Musters in einer Sequenz gezählt werden. Aus Effizienzgründen werden in meiner Arbeit die Programme RNAMotif und HyPa für diese Aufgabe verwendet. Die "expectation algebra" ermöglicht die Abschätzung des Erwartungswertes eines Musters in einer Zufallssequenz. Die Basenzusammensetzung und die Sequenzlänge sind die einzigen anzugebenden Parameter für die Abschätzung. Diese Auswertungsalgebra wurde im Rahmen meiner Arbeit eingesetzt, um signifikante Suchmuster für die Datenbanksuche zu definieren.

Für diese Datenbanksuche wurden in meiner Arbeit entweder die 5'UTR Kollektionen, oder die 3'UTR Kollektionen der UTR Datenbank Version 15.0 (s. Abschnitt 3.4.1) eingesetzt. Die Größe und Basenzusammensetzung dieser Kollektionen ist in Tabelle 4.1 angegeben.

UTR	Größe	Basenzusammensetzung
5'UTR	16,8 Millionen Basen	Adenin: 0.2572 Cytosin 0.2522 Guanin 0.2495 Thymin 0.2411
3'UTR	47,2 Millionen Basen	Adenin: 0.2858 Cytosin: 0.2094 Guanin: 0.2109 Thymin: 0.2939

Tabelle 4.1.: Größe und Basenzusammensetzung der UTRdb 15.0.

Mit Hilfe der "expectation algebra" lassen sich für beliebige Muster Erwartungswerte für das Auftreten des Musters in einer Zufallssequenz, deren Länge und



Basenzusammensetzung den UTRdb Kollektionen entspricht, nach oben abschätzen. Für die 5'UTR Kollektionen ergibt sich für das, durch die oben angegebene Grammatik beschriebene, IRE-Muster ein Erwartungswert von 1,09 Treffern. Durch eine durchgeführte Datenbanksuche mit diesem Muster kann man dann entscheiden, ob die tatsächlich gefundene Trefferanzahl signifikant ist. Im folgenden Abschnitt wird die Beschreibung und Datenbanksuche des IRE-Musters aus Abbildung 4.1 mit den Programmen `RNAMotif` und `HyPa` präsentiert.

### 4.3.2. Musterbeschreibung und Datenbanksuche

Mit der im vorherigen Abschnitt beschriebenen Methode lassen sich signifikante Suchmuster für ein vorhergesagtes RNA-Motiv definieren. In diesem Abschnitt wird die Beschreibung solcher Muster mit den Programmen `RNAMotif` und `HyPa` dargestellt. Diese Programme wurden zur effizienten Suche in Datenbanken eingesetzt. Am Beispiel des *Iron Responsive Elements* wird die Funktionalität der Programme erläutert.

#### Musterbeschreibung und Datenbanksuche mit `RNAMotif`

Das Programm `RNAMotif` wurde in Abschnitt 3.4.6 vorgestellt. Wie dort beschrieben worden ist, wird das Suchmuster in der sogenannten "descriptor"-Datei definiert. Diese Datei sieht für das *Iron Responsive Element* aus Abbildung 4.1 wie folgt aus:

```
parms
    wc += gu;

descr
    h5( tag='stem1', len=3 )
        ss( len=1, seq="c" )
        h5( tag='stem2', len=5)
            ss( len=6, seq="cagugh" )
            h3(tag='stem2')
    h3(tag='stem1')
```

score

Diese "descriptor"-Datei gliedert sich in die Abschnitte "parms", "descr" und "score". In dem ersten Abschnitt werden zunächst die Basenpaarregeln als Parameter festgelegt. In diesem Fall wurden Watson-Crick und GU Basenpaare zugelassen. Der "descr" Abschnitt beschreibt das eigentliche Motiv. Helices werden mit

#### 4. Ergebnisse

"h5/h3" und einzelsträngige Bereiche mit "ss" definiert. Beiden Elementen lassen sich vorgegebene Längen (durch "len") und Sequenzen (durch "seq") zuordnen. Die Längenangaben können durch die Verwendung der Konstrukte "minlen" und "maxlen" auch variabel gehalten werden. Die Angabe eines "tag" für ein bestimmtes Strukturelement erhöht die Übersichtlichkeit der Beschreibung und ist notwendig, falls das Element in dem abschließenden "score" Abschnitt besonders bewertet werden soll. Im Rahmen meiner Arbeit waren für die Musterdefinitionen jedoch keine Score-Werte erforderlich. Die Verwendung des "score" Abschnittes ist in der Veröffentlichung von T. Macke und Kollegen (Macke et al., 2001) ausführlich beschrieben worden.

Mit dem Muster für das IRE wurden die 5'UTR Kollektionen der UTRdb durchsucht. Das Programm `RNAMotif` bekommt die "descriptor"-Datei sowie die zu durchsuchende Sequenz im FASTA Format als Eingabe. Es wurden insgesamt 31 Treffer in den 5'UTR Kollektionen gefunden. Die Tabelle 4.2 gibt einen Überblick über die Verteilung der Treffer auf die jeweilige Kollektion. Ein Treffer wird von `RNAMotif` dabei wie folgt dargestellt:

```
#RM descr h5(tag='stem1') ss h5(tag='stem2') ss h3(tag='stem2')
h3(tag='stem1')

>5HSA003828 5'UTR in Human ferritin heavy chain mRNA,
complete cds. 208 bp

5HSA003828      0.000 0      34   23 ctg c ttcaa cagtgc ttgga cgg
```

In der obersten Zeile werden die einzelnen Elemente des Musters, wie es in der "descriptor"-Datei beschrieben wurde, angegeben. Von dem jeweiligen Treffer wird zunächst der sogenannte "Header" des FASTA Eintrages angegeben. In der nächsten Zeile wird die Sequenz des Treffers, aufgeteilt in die einzelnen Musterelemente, angegeben. Die Zahl "23" gibt die Länge des gefundenen Musters und die Zahl "34" die Startposition des Musters in der gefundenen Sequenz an. Die Zahl "0" gibt an, daß das Muster auf dem sogenannten (+)Strang gefunden wurde. Eine "1" würde einen Treffer auf dem Gegenstrang anzeigen. Bei Datenbanksuchen in der UTRdb wurden nur Treffer auf dem (+)Strang berücksichtigt.

Die für das IRE gefundene Trefferanzahl von 31 liegt deutlich über dem Erwartungswert von 1,09. Bei der Suche wurde das IRE in vielen Ferritin 5'UTRs gefunden. Dort ist eine biologische Funktion für das IRE beschrieben, so daß diese Treffer als sogenannte "True positivs" gewertet werden können. Daneben wurde mit dem verwendeten Suchmuster aber zum Beispiel auch ein Treffer in der 5'UTR der "cell adhesion regulator" mRNA (CAR1) der Ratte gefunden. Für

Kollektion	Anzahl der Treffer
5'UTR Mensch	5
5'UTR Nager	4
5'UTR andere Säuger	2
5'UTR andere Vertebraten	9
5'UTR Invertebraten	11
5'UTR Viren	0
5'UTR Pflanzen	0
5'UTR Pilze	0
<b>gesamt</b>	<b>31</b>

Tabelle 4.2.: Ergebnis der Datenbanksuche für das IRE Muster mit `RNAMotif` und `HyPa`.

diese mRNA konnte in der Literatur kein Hinweis auf eine mögliche posttranskriptionelle Genregulation durch Eisen gefunden werden, so daß eine biologische Funktion experimentell nachgewiesen werden müßte.

Dieses Ergebnis verdeutlicht die Stärke der Datenbanksuchen. Da man die biologische Funktion von IREs in 5'UTRs, zum Beispiel der Ferritin 5'UTR, bereits kennt, kann man nun gezielt Experimente durchführen, um solch eine Funktion auch für die CAR1-mRNA nachzuweisen.

### Musterbeschreibung und Datenbanksuche mit `HyPa`

Das Programm `HyPa` wurde zeitgleich zu meiner Arbeit in der Doktorarbeit von Dirk Strothmann entwickelt. Es bietet eine ähnliche Funktionalität wie `RNAMotif` und wurde in meiner Arbeit vergleichend zu diesem Programm eingesetzt. Die Notation von kombinierten Sequenz- und Strukturmustern, wie zum Beispiel dem IRE aus Abbildung 4.1 sieht in `HyPa` wie folgt aus:

```
IRE_Muster = (stem1:=.{3}) C (stem2:=.{5}) (loop:=CAGUGH)
              ^(stem2) ^(stem1);
```

Die einzelnen Strukturelemente werden in separaten Klammern nacheinander aufgelistet. Dabei ist zum Beispiel die Länge der einzelnen Helices in geschweiften Klammern mit einem vorangestellten Punkt definiert. Wie bei `RNAMotif` bestehen die Muster auch hier aus gepaarten und ungepaarten Elementen, wobei der Cytosin Bulge einfach als "C" repräsentiert wird. Die 5'-Seiten von gepaarten Bereichen werden hier als "stem1", "stem2", u.s.w. durchnummeriert, wobei die 3'-Seiten mit einem vorangestellten "" beginnen. GU-Basenpaare werden automatisch berücksichtigt. Die Definition anderer Basenpaare, wie zum Beispiel AG-Paaren,

#### 4. Ergebnisse

ist ebenfalls möglich. HyPa durchsucht standardmäßig nur den (+)Strang der zu durchsuchenden Sequenz (Suche auf dem Gegenstrang als Parameter einstellbar). Bei der Suche in der UTR Datenbank ist das ein Vorteil gegenüber RNAMotif, da man hier nur an den Treffern im (+)Strang interessiert ist. Bei RNAMotif müssen alle Treffer daraufhin überprüft werden, auf welchem Strang sie liegen.

Das oben angegebene Muster wurde benutzt, um die 5'UTR Kollektionen der UTR Datenbank mit HyPa zu durchsuchen. Dazu mußte zunächst ein sogenannter Index der Datenbank erstellt werden, da HyPa die Suche basierend auf diesem Index durchführt. Das Resultat der Suche stimmte mit dem von RNAMotif überein (s. Tabelle 4.2). Um die Sucheeffizienz von RNAMotif und HyPa zu vergleichen, wurden die Zeiten ermittelt, die für die Suche nach dem IRE-Muster in der jeweiligen 5'UTR Kollektion der UTRdb benötigt wurden. Die Index-Erstellung dauert dabei sehr viel länger als die eigentliche Suche. Da sie jedoch für eine gegebene Datenbank nur ein einziges mal durchgeführt werden muß, wurde sie für die Bestimmung der eigentlichen Suchzeiten für dieses und die folgenden Muster nicht berücksichtigt. Die Suche wurde auf einem Rechner vom Typ "Sun Ultra-Sparc III" (900 MHz) durchgeführt. In der Tabelle 4.3 sind die ermittelten Zeiten aufgelistet. Die ermittelten Zeiten sind für HyPa zu klein, um noch gemessen

<b>Kollektion</b>	<b>Datenbankgröße in Millionen Basen</b>	<b>RNAMotif Zeit in sec</b>	<b>HyPa Zeit in sec</b>
5'UTR Mensch	4,52	0,7	< 0,1
5'UTR Nager	2,52	0,4	< 0,1
5'UTR andere Säuger	0,85	0,1	< 0,1
5'UTR andere Vertebraten	0,79	0,1	< 0,1
5'UTR Invertebraten	2,99	0,4	< 0,1
5'UTR Viren	3,40	0,4	< 0,1
5'UTR Pflanzen	1,49	0,1	< 0,1
5'UTR Pilze	0,26	< 0,1	< 0,1

Tabelle 4.3.: Vergleich von RNAMotif und HyPa für die Mustersuche nach dem IRE Muster aus Abbildung 4.1

werden zu können. Für RNAMotif liegen die Suchzeiten bei allen Kollektionen unter einer Sekunde. Der Unterschied wird jedoch deutlicher, wenn man größere Datenmengen durchsucht, wie zum Beispiel die 3'UTR Kollektion des Menschen. Diese hat eine Größe von 18,9 Millionen Basen. Für die Suche nach dem IRE-Muster ergibt sich für RNAMotif eine Zeit von 3,6 Sekunden, während für HyPa die Zeit auch hier zu klein ist, um gemessen zu werden.

In Abschnitt 4.4.2 wird ein größeres Muster mit mehr Variabilität in den Längen einzelner Strukturelemente definiert. Anhand dieses Musters wird erneut ein Vergleich der Programme RNAMotif und HyPa in Bezug auf die benötigte Zeit zur Durchsuchung der UTR Datenbank durchgeführt.

## 4.4. Strategieranwendung

In den folgenden Abschnitten werden die entwickelten Strategien zur Vorhersage und Validierung potentieller regulatorischer RNA-Struktur motive auf verschiedene mRNAs angewendet. Dadurch sollen neue, bisher nicht bekannte Motive und posttranskriptionelle Regulationsmechanismen untersucht werden.

### 4.4.1. Die Rab1A 3'UTR

Das Rab1A Gen kodiert für ein Protein aus der Gruppe der Rab Proteine. Bei den Rab Proteinen handelt es sich um GTPasen der p21 Ras Superfamilie, deren Funktion im zellulären Vesikeltransport liegt (Martinez und Goud, 1998). Die Abbildung 4.20 gibt einen Überblick über die Wirkungsorte der verschiedenen Rab Proteine.

Rab1A ist ein ubiquitär exprimiertes Protein, das eine wichtige Rolle beim Vesikeltransport vom Endoplasmatischen Retikulum zum Golgi-Apparat spielt (Martinez und Goud, 1998). Darüber hinaus hat es eine Funktion im frühen Transport der Golgivesikel innerhalb des Golgi-Apparates (Nuoffer et al., 1994). In einer Arbeit von Wedemeyer et al. wurde eine extreme Sequenzkonservierung der Rab1A 3'UTR bei verschiedenen Vertebraten beschrieben (Wedemeyer et al., 2000). Die durchschnittliche Sequenzkonservierung des untersuchten Abschnittes der Rab1A 3'UTR beträgt 92%. Die Abbildung 4.21 zeigt ein mit ClustalW berechnetes Alignment der in der Arbeit von Wedemeyer untersuchten Rab1A 3'UTRs. Es handelt sich bei den Sequenzen nicht um die kompletten 3'UTRs, sondern um einen bestimmten Abschnitt, der sich durch PCR amplifizieren ließ (Wedemeyer et al., 2000).

Im folgenden wird beschrieben, wie die Rab1A 3'UTR mit den entwickelten Strategien auf ein potentielles regulatorisches Element untersucht wird. In Abschnitt 4.5 werden die Ergebnisse von experimentellen Analysen des so vorhergesagten Motivs präsentiert.

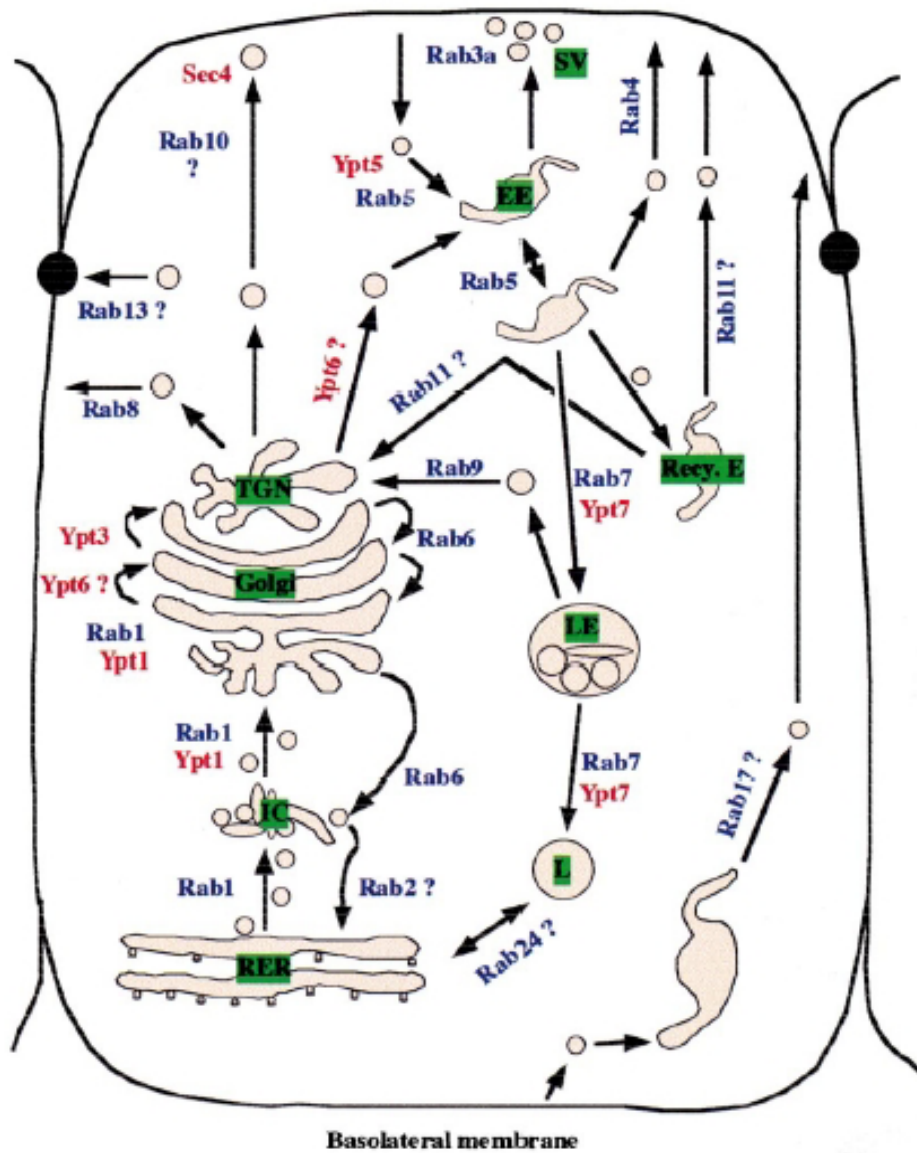


Abbildung 4.20.: Lokalisation und Funktion der Rab Proteine. RER: rauhes Endoplasmatisches Retikulum; TG: Trans-Golgi-Netzwerk; SV: synaptische Vesikel; EE: frühe Endosomen; LE: späte Endosomen; L: Lysosomen (aus (Martinez und Goud, 1998)).

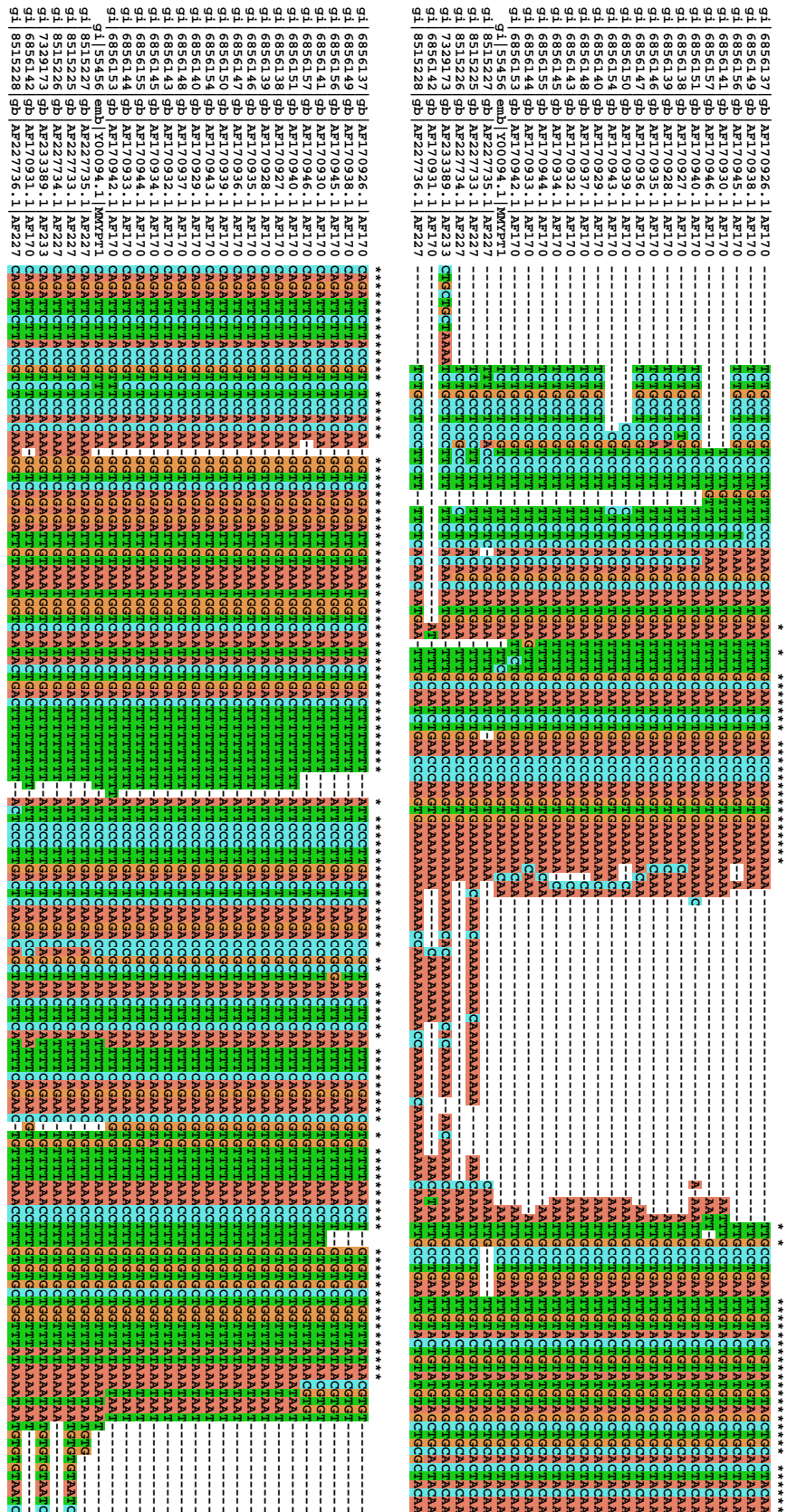


Abbildung 4.21.: Sequenzalignent (berechnet mit Clusta1W) der in (Wedemeyer et al., 2000) untersuchten Rab1A 3'UTRs.

#### 4. Ergebnisse

##### Strukturmotiv-Vorhersage

Die hohe Sequenzähnlichkeit der in Abbildung 4.21 dargestellten Rab1A 3'UTRs führte dazu, daß mit den in dieser Arbeit entwickelten Vorhersagestrategien zunächst kein spezielles Motiv entdeckt werden konnte. Das DIALIGN2 Alignment der Rab1A 3'UTRs von Mensch und Maus verdeutlicht das für die erste Strategie, da hier kein Bereich hervorgehoben wird.

Rab1A 3'UTRs (Mensch, Maus):

Alignment (DIALIGN format):

```
Mensch      1   TCTGCCTCCA TCCTTTTCTC ACAGCAATGA ATTttGCAAT CTGAACCCAA
Maus        1   TCTGCCTCCG TCCTTTTCTC ACAGCAATGA ATTc-GCAAT CTGAACCCAA

*****
*****
*****

Mensch      51  GTGAAAAA-C AAAATTCCTT GAATTGTACT GTATGTAGCT GCACTACAAC
Maus       50  GTGAAAAAaC AAAATTCCTT GAATTGTACT GTATGTAGCT GCACTACAAC

***** *
***** *
***** *

Mensch     100  AGATTCTTAC CGTCTCCACA AGGTCAGAGA TTGTAAATGG TCAATACTGA
Maus     100  AGATTCTTAC CGTTTCCACA AGGTCAGAGA TTGTAAATGG TCAATACTGA

*****
*****
*****

Mensch     150  CTTTTTTTTT TATTCCTTGG ACTCAAGACC GCTAACTTCA ATTTCAGAAC
Maus     150  CTTTTTTTTT TATTCCTTGG ACTCAAGACC GCTAACTTCA TTTTCAGAAC

*****
*****
*****

Mensch     200  gTGTITTTAAA CCTTTGTGTG CTGGTTTATA AAtaat-
Maus     200  -TGTITTTAAA CCTTTGTGTG CTGGTTTATA AAataat

*****
*****
*****
```

Aufgrund des DIALIGN2 Ergebnisses müßte man die gesamte alignierte Region als potentielles regulatorisches Element ansehen. Auch die Anwendung der zweiten



Strategie zur Strukturmotivvorhersage würde zu diesem Resultat führen, da man aufgrund der geringen Sequenzunterschiede annehmen muß, daß die Strukturen der UTRs insgesamt sehr ähnlich sind. Daher liefert auch ein lokales Strukturalignment kein spezielles Strukturmotiv, das weiter untersucht werden könnte.

Um die Rab1A 3'UTR auf ein kleineres Strukturmotiv zu untersuchen, müßten weniger stark konservierte Rab1A Sequenzen analysiert werden. Um solche Sequenzen zu finden, wurden Datenbanksuchen durchgeführt. Dabei wurde die Sequenz der Rab1A 3'UTR des Zitterrochen gefunden, die eine deutlich geringere Sequenzkonservierung aufweist als die in der Arbeit von Wedemeyer beschriebenen Sequenzen. Das mit DIALIGN2 berechnete Alignment der Rab1A 3'UTRs von Mensch und Zitterrochen zeigt einen begrenzten Bereich hoher Sequenzähnlichkeit.

Rab1A 3'UTRs (Mensch, Zitterrochen):  
Alignment (DIALIGN format):

```
Zitterrochen 1  gaatt----- ---TTCACTC TCCGCATTTG ATTTaaaagg cccaa---AA
Mensch        1  tctgcctcca tccTTTTCTC ACAGCAATGA ATTTtgcaat ctgaaccAA
```

```
***** ***** ***
```

```
**
```

```
Zitterrochen 40  GTAAGAAAtg ---TTGCCTG AATTGACTG TAAGTAGCCA CACTACTGCA
Mensch        51  GTGAAAAAca aaaTTGCCTG AATTGACTG TATGTAGCTG CACTACAACA
```

```
*****
***** ***** ***** *****
***** ***** ***** *****
***** ***** ***** *****
```

```
Zitterrochen 87  GActttcccc t-----
Mensch        101  GAttcttacc gtctccacaa ggtagagat tgtaaatggt caatactgac
```

```
**
**
**
```

```
Zitterrochen 98  ----- -TCCCTgcc ccc----- ----- TTTCAGAcac
Mensch        151  ttttttttt aTCCCTtga ctcaagaccg ctaacttcaa TTTCAGAacg
```

```
*****
```

```
*****
```

```
Zitterrochen120  aatcaggcac tgagaagcca ggattgGCTG Ggatt----- ---
Mensch          201  tgttttaaac ctttgtgt-- -----GCTG Gttataaat aat
```

```
**** *
```

#### 4. Ergebnisse

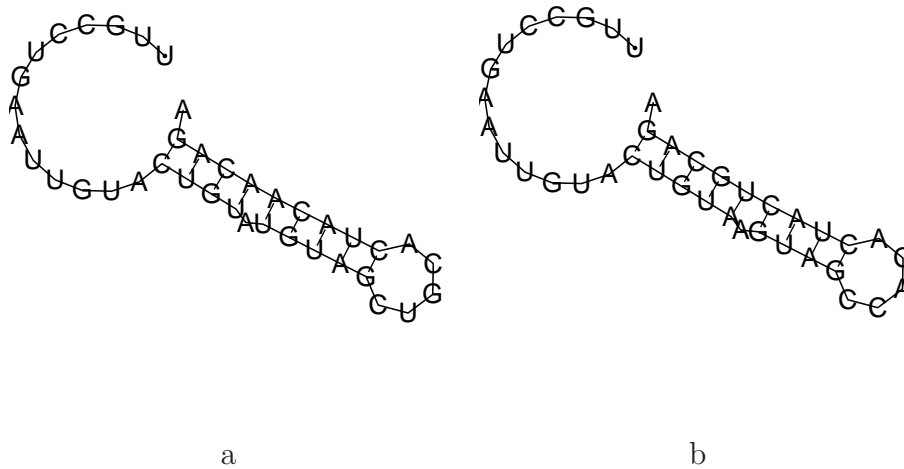


Abbildung 4.22.: Strukturvorhersagen für den durch das Alignment hervorgehobenen Bereich: a) Mensch, b) Zitterrochen.

Für diesen Bereich wurde die Struktur mit dem Programm `RNAfold` berechnet (s. Abbildung 4.22). Dieser bereits auf Sequenzebene hoch konservierte Bereich bildet sowohl bei Mensch als auch beim Zitterrochen einen Stemloop.

Dieser vorhergesagte Stemloop unterscheidet sich in der Länge der basengepaarten Bereiche zwischen Mensch und Zitterrochen. Während beim Menschen die obere Helix fünf Basenpaare und die untere vier Basenpaare lang ist, ist es beim Zitterrochen genau umgekehrt. Wenn man beim Zitterrochen den Bulgeloop um eine Position verschiebt, erhält man jedoch das gleiche Strukturmotiv wie beim Menschen.

Der durch die erste Strategie vorhergesagte Stemloop konnte durch Anwendung der zweiten Vorhersagestrategie bestätigt werden. Die Abbildung 4.23 zeigt eine Strukturvorhersage für einen Abschnitt der Rab1A 3'UTR des Menschen und Abbildung 4.24 eine Strukturvorhersage für die Rab1A 3'UTR des Zitterrochen.

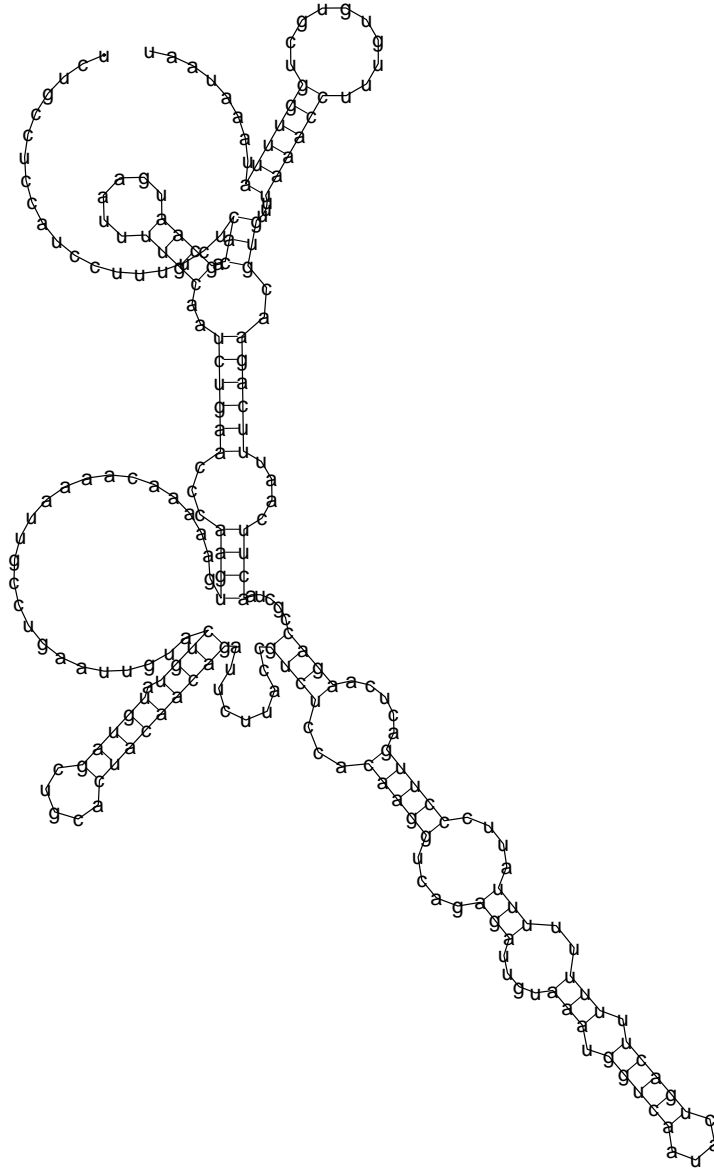


Abbildung 4.23.: Strukturvorhersage für die Rab1A 3'UTR des Menschen.

4. Ergebnisse

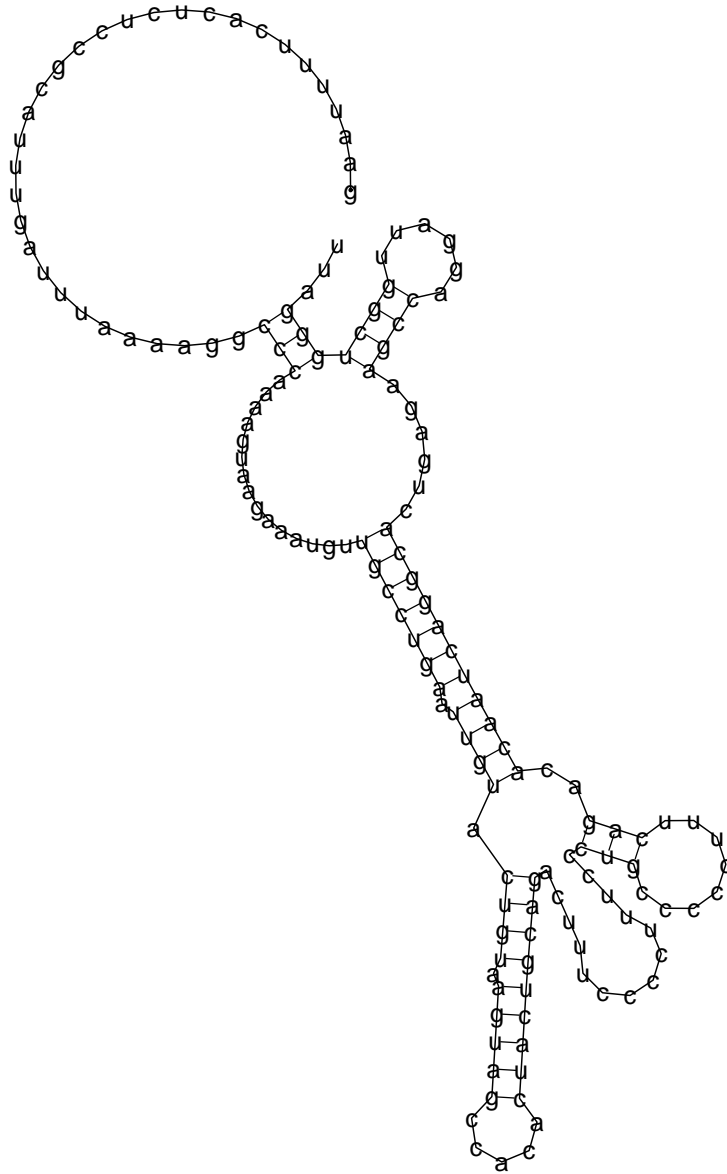


Abbildung 4.24.: Strukturvorhersage für die Rab1A 3'UTR des Zitterrochen.

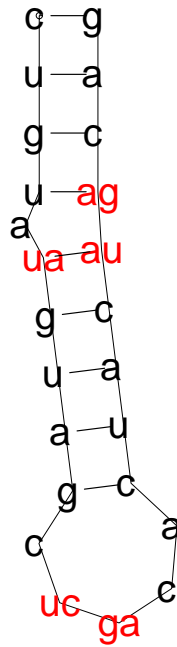


Abbildung 4.25.: Lokales Strukturalignment der Rab1A 3'UTR von Mensch und Zitterrochen.

Das lokale Strukturalignment (s. Abbildung 4.25) der beiden UTR Strukturen liefert den Stemloop, der auch durch die erste Strategie vorhergesagt wurde, als ähnlichstes Strukturelement in den beiden UTRs. Im Vergleich zur ersten Strategie fehlen hier die einzelsträngigen Bereiche vor dem Stemloop. Die hohe Sequenzkonservierung innerhalb des Stemloops ist in dem Strukturalignment deutlich erkennbar. Besonders interessant sind jedoch die Sequenzunterschiede innerhalb des basengepaarten Bereichs. Diese kompensatorischen Basenaustausche in der Struktur der Rab1A 3'UTR des Zitterrochen im Vergleich zu der des Menschen liefern entscheidende Hinweise darauf, daß dieses Strukturmotiv in den UTRs tatsächlich existiert.

Einen weiteren Nachweis für das Vorkommen dieses Stemloops in den nativen Strukturen ergeben die Wahrscheinlichkeitsverteilungen für Basenpaare in den beiden Strukturen. Diese Wahrscheinlichkeiten werden als Dotplots dargestellt (Abbildungen 4.26 und 4.27), wobei die Fläche der Rauten ein Maß für die Wahrscheinlichkeit des durch die Raute repräsentierten Basenpaares ist.

In beiden Strukturen ist die Wahrscheinlichkeit für die Bildung dieses Stemloops sehr hoch. Das läßt sich ebenfalls in den Dotplots der Rab1A 3'UTRs anderer Organismen beobachten (Daten hier nicht gezeigt).

#### 4. Ergebnisse

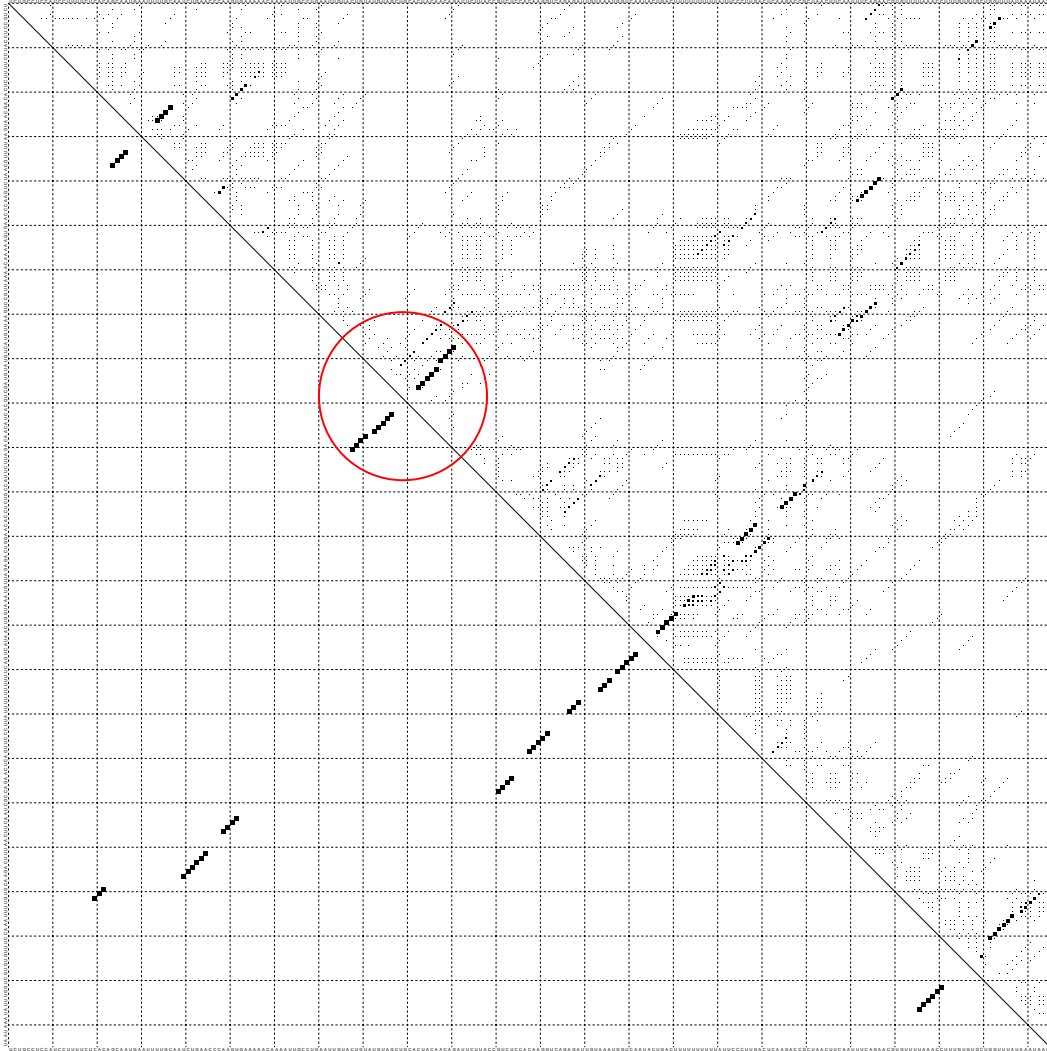


Abbildung 4.26.: Dotplot: Basenpaarwahrscheinlichkeiten (oben rechts) der Rab1A 3'UTR des Menschen (unten links: energetisch günstigste Struktur). Der rote Kreis markiert den Bereich des Stem-loops.

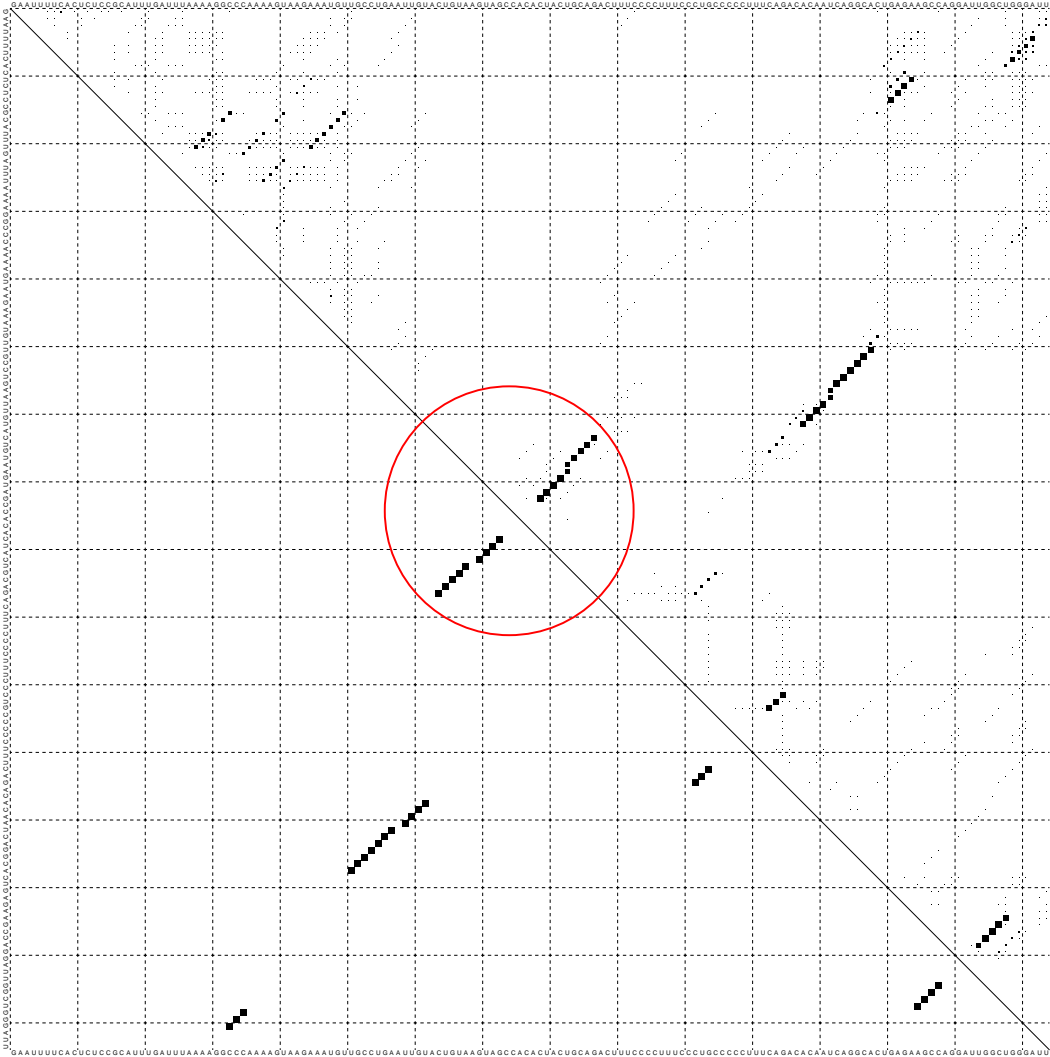


Abbildung 4.27.: Dotplot: Basenpaarwahrscheinlichkeiten (oben rechts) der Rab1A 3'UTR des Zitterrochen (unten links: energetisch günstigste Struktur). Der rote Kreis markiert den Bereich des Stem-loops.





Die Ergebnisse der tatsächlichen Suche in den 3'UTR Kollektionen der UTRdb werden im folgenden beschrieben.

### Datenbanksuche

Für die Durchsuchung der 3'UTR Kollektionen der UTR Datenbank (s. Abschnitt 3.4.1) mit dem durch die Signifikanzberechnung definierten Muster wurden die Programme `RNAMotif` und `HyPa` verwendet.

Das oben definierte Muster (s. Abbildung 4.28) läßt sich mit `RNAMotif` wie folgt beschreiben:

```
parms
    wc += gu;

descr

    h5( tag='stem1', len=4 )
        ss( len=1, seq="a" )
        h5( tag='stem2', len=5, seq="g$")
            ss( len=5, seq="cyrca" )
            h3(tag='stem2', seq="^c")
    h3(tag='stem1')
```

Auf die 5'-Seite der ersten Helix ("stem1"), die die Länge 4 haben muß, folgt der Adenin-Bulge. Danach wird die 5'-Seite der oberen Helix ("stem2") beschrieben, die die Länge 5 hat und mit einem Guanin endet (beschrieben durch `seq="g$"`). Der folgende 5er Loop hat zwei variable Positionen, wobei an der zweiten Position im Loop ein Uracil oder Cytosin und an der dritten Position ein Adenin oder Guanin auftreten kann. Die 3'-Seite der oberen Helix muß mit einem Cytosin beginnen (beschrieben durch `seq="^c"`).

Die Ergebnisse der Datenbanksuche sind in Tabelle 4.4 für die jeweilige Kollektion aufgelistet.

Die Datenbanksuche lieferte insgesamt 17 Treffer. Das Muster wurde in zehn Rab1A 3'UTRs verschiedener Organismen gefunden (bei der Maus gab es Treffer in zwei verschiedenen Datenbankeinträgen für Rab1), wobei diese Treffer auf Vertebraten beschränkt waren. Das in Abbildung 4.21 dargestellte `ClustalW` Alignment enthält 26 Sequenzen, von denen 25 den durch das Muster beschriebenen Stemloop enthalten (einzige Ausnahme ist die Rab1A 3'UTR des Tasmanischen Teufels, wo der Loop in der fünften Position ein Guanin statt eines Adenins hat und daher mit diesem Suchmuster nicht gefunden werden konnte). Von den zehn in der Datenbanksuche gefundenen Rab1A 3'UTRs finden sich neun in dem

#### 4. Ergebnisse

Kollektion	Acc.Nr UTRdb (EMBL)	Organismus	Gen
Mensch	3HSA026464 (AF083106)	Mensch	Sirtuin 1 mRNA
	3HSA033544 (BC000905)	Mensch	Rab1 mRNA
	3HSA038486 (AL136741)	Mensch	DKFZ Klon
	3HSA039949 (AF235040)	Mensch	Sir2alpha mRNA
Nager	3MMU010564 (AF214646)	Maus	Sir2alpha mRNA
	3MMU010608 (AF226873)	Maus	Rab1A mRNA
	3MMU013536 (BC002077)	Maus	Rab1 mRNA
	3MMU014957 (BC006584)	Maus	Klon, ähnlich zu Sir2alpha
andere Säuger	3MAG000002 (AF170931)	Känguruh	Rab1A mRNA
	3TVU000043 (AF227733)	Opossum	Rab1A mRNA
	3FCA000156 (AF227734)	Katze	Rab1A mRNA
	3BTA001889 (AF227735)	Stier	Rab1A mRNA
	3DBY000002 (AF233389)	Beutelmarder	Rab1A mRNA
andere Vertebraten	3AMI000013 (AF170926)	Alligator	Rab1A mRNA
	3GGA001772 (AF170930)	Huhn	Rab1A mRNA
	3DOM000021 (M38393)	Zitterrochen	Rab1 mRNA
Invertebraten	—	—	—
Virus	—	—	—
Pflanzen	3PPY000040 (AB014456)	Birne	O-Methyltransferase mRNA
Pilze	—	—	—

Tabelle 4.4.: Ergebnis der Datenbanksuche für das Rab1 Muster mit RNAMotif und HyPa.

ClustalW Alignment (der zehnte Treffer ist die Rab1A 3'UTR des Zitterrochen). Die fehlenden 16 Rab1A 3'UTRs sind nicht in der UTR Datenbank enthalten und konnten somit durch diese Datenbanksuche nicht gefunden werden. Neben den Rab1A 3'UTRs wurde das Muster in der 3'UTR der Methyltransferase einer Pflanze gefunden. Außerdem findet man diesen Stemloop in den 3'UTRs der Sirtuin Proteine von Mensch und Maus.

Aufgrund der Signifikanzberechnungen hätte man mit diesem Muster in einer Zufallssequenz höchstens 0.81 Treffer erwartet. Man findet mit 17 Treffern daher mehr, als in einer Zufallssequenz zu erwarten gewesen wären. Somit scheint dieser Stemloop sehr spezifisch für Rab1A 3'UTRs zu sein, was eine biologische Funktion des Motivs wahrscheinlich macht (die Rolle der zusätzlichen Treffer in der Methyltransferase und den Sirtuinen wird später diskutiert). Zur weiteren Überprüfung dieser starken Beschränkung auf nur sehr wenige verschiedene 3'UTRs wurden sämtliche 5'UTR Datensätze der UTR-Datenbank mit dem Muster durchsucht. Hier konnte kein einziger Treffer gefunden werden.

Wie bei dem IRE Muster in Abschnitt 4.3.2, wurde auch für das Rab1 Muster ein Vergleich der Mustersuche zwischen HyPa und RNAMotif durchgeführt. Das Rab1 Muster läßt sich in HyPa wie folgt beschreiben:

```
Rab1_Muster = (stem1:={4}) A (stem2:={4}) G (loop:=CYRCA) C
              ^(stem2) ^(stem1);
```

Die Ergebnisse der Suche mit HyPa stimmten mit den Ergebnissen der RNAMotif Suche überein (Tabelle 4.4). Auch hier wurden die Suchzeiten der beiden Programme miteinander verglichen. Die Ergebnisse dieses Vergleiches sind in Tabelle 4.5 aufgeführt. Genau wie bei dem IRE-Muster, das vom Aufbau her dem Rab-Muster sehr ähnlich ist, waren die Suchzeiten für HyPa zu klein, um gemessen werden zu können. RNAMotif benötigt dagegen für die Mustersuche in der 3'UTR Kollektion des Menschen 2,8 Sekunden. In Abschnitt 4.4.2 wird die Suche nach einem größeren und variableren Muster beschrieben. Dort wird der Unterschied der Suchzeiten zwischen HyPa und RNAMotif noch deutlicher.

#### 4. Ergebnisse

Kollektion	Datenbankgröße in Millionen Basen	RNAMotif Zeit in sec	HyPa Zeit in sec
3'UTR Mensch	18,91	2,8	< 0,1
3'UTR Nager	9,11	1,4	< 0,1
3'UTR andere Säuger	2,55	0,3	< 0,1
3'UTR andere Vertebraten	3,35	0,5	< 0,1
3'UTR Invertebraten	5,15	0,7	< 0,1
3'UTR Viren	3,36	0,5	< 0,1
3'UTR Pflanzen	4,33	0,6	< 0,1
3'UTR Pilze	0,47	< 0,1	< 0,1

Tabelle 4.5.: Vergleich von RNAMotif und HyPa für die Mustersuche nach dem Rab1 Muster aus Abbildung 4.28.

#### 4.4.2. Die BMP7 3'UTR

Die "Bone morphogenetic proteins" (BMPs) sind Signalmoleküle und gehören zu der TGF- $\beta$  (transforming-growth-factor) Superfamilie von Peptid Wachstumsfaktoren. Diese Signalmoleküle binden an spezifische Membranrezeptoren, wodurch letztendlich Proteine der sogenannten SMAD Familie phosphoryliert werden, die dann in den Kern wandern und dort als Transkriptionsfaktoren wirken (Massague, 2000). Dadurch werden verschiedenste Prozesse, zum Beispiel in der Embryonalentwicklung, gesteuert. Das BMP7 Protein induziert unter anderem die Bildung der Epidermis in dem Krallenfrosch *Xenopus laevis*.

In der laufenden Doktorarbeit von Alexander Sczyrba wurden ESTs (expressed sequence tags) von *Xenopus* geclustert und assembliert. Dabei sind zwei Varianten der BMP7 mRNA aufgefallen, bei denen es sich möglicherweise um alternative Spleißvarianten handelt. Die Variationen liegen dabei innerhalb der 3'UTR der BMP7 mRNA. Diese beiden Varianten der *Xenopus* BMP7 3'UTR werden in den folgenden Abschnitten untersucht.

#### Strukturmotivvorhersage

Zunächst wurde ein paarweises Alignment berechnet, um die Unterschiede der beiden Varianten der BMP7 3'UTR hervorzuheben.

BMP7 3'UTRs (Xenopus, Variante 1 und 2):  
Alignment (DIALIGN format):

```

Variante1      1  CAATAGCAGT TATTCTGTTT TTAACAGTCA TTTTAATGGT ATTGTCCTTA
Variante2      1  CAATAGCAGT TATTCTGTTT TTAACAGTCA TTTTAATGGT ATTGTCCTTA

*****
*****

Variante1     51  TGTTTATTTT AAagttttaa aacAGTAGAG ATACTTGACC ATCACACTTA
Variante2     51  TGTTTATTTT AA----- ---AGTAGAG ATACTTGACC ATCACACTTA

***** **
***** **

Variante1     101  AAAAATGCAT TGTACACCTT AACGGATGAA AAGATTTTGT TTTTGc---
Variante2     90  AAAAATGCAT TGTACACCTT AACGGATGAA AAGATTTTGT TTTTGcAtga

*****
*****

Variante1     148  -----A TGATTTTCAT GCAGTTAGGA AAATGTTATA TGTTcATTAT
Variante2     140  tttttttgcA TGATTTTCAT GCAGTTAGGA AAATGTTATA TGTTcATTAT

*
*

Variante1     189  TTCTGTAATG TATTGGCATG TTATTGTATC CAAGAAGAAC AATGACTGAT
Variante2     190  TTCTGTAATG TATTGGCATG TTATTGTATC CAAGAAGAAC AATGACTGAT

*****
*****

Variante1     239  GCAATCAAAA TATGGCTCTT CAACAGAAAA TGTACAACCTG CAATGGGCAA
Variante2     240  GCAATCAAAA TATGGCTCTT CAACAGAAAA TGTACAACCTG CAATGGGCAA

*****
*****

Variante1     289  ATAATTA AAC AATTTT
Variante2     290  ATAATTA AAC AATTTT

*****
*****

```

In der ersten Variante ist ab Position 63 ein 11 Nukleotide langes Stück eingefügt und ab der Position 148 ein 12 Nukleotide langes Stück gelöscht worden. Um zu überprüfen, ob eine der beiden Varianten bereits als Sequenz veröffentlicht worden ist, wurde in der UTR Datenbank nach der BMP7 3'UTR von *Xenopus laevis* gesucht. Die gefundene Sequenz (Acc.Nr. 3XLA000193) ist 347 Nukleotide lang und damit etwas länger als die in der Arbeit von Alexander Sczyrba durch EST-Clustering gefundenen Varianten eins (304 Nkl.) und zwei (305 Nkl.). Für diese drei Sequenzen wurde ebenfalls ein Alignment berechnet.

#### 4. Ergebnisse

BMP7 3'UTRs (Xenopus, UTRdb Eintrag, Variante 1 und 2):  
Alignment (DIALIGN format):

```

3XLA000193      1  caactcgagt tgtgtcagct gttgtcatgc aatgcagtat cgTTTACAGT
XL_BMP7_v1      1  CAATAGCAGT TATTCTGTTT ----- --TTAACAGT
XL_BMP7_v2      1  CAATAGCAGT TATTCTGTTT ----- --TTAACAGT

          *****

3XLA000193      51  CATT--ATG GTATTGCCT TATGTTTATT TTAAAGTTT --AACAGTAG
XL_BMP7_v1      29  CATTTAATG GTATTGCCT TATGTTTATT TTAAAGTTT taAACAGTAG
XL_BMP7_v2      29  CATTTAATG GTATTGCCT TATGTTTATT TTAA----- ----AGTAG

          *****
          *** *****

3XLA000193      97  AGATACT-GA CCATCACACT --AAAAATGC ATTGTACACC TTAACGGATG
XL_BMP7_v1      79  AGATACTTGA CCATCACACT TAAAAAATGC ATTGTACACC TTAACGGATG
XL_BMP7_v2      68  AGATACTTGA CCATCACACT TAAAAAATGC ATTGTACACC TTAACGGATG

          *****
          *****

3XLA000193      144  AAAAGATTTT GctTTTTTGC T----- ---ATGATT TCATGCAGTT
XL_BMP7_v1      129  AAAAGATTTT G--TTTTTGC T----- ---ATGATT TCATGCAGTT
XL_BMP7_v2      118  AAAAGATTTT G--TTTTTGC atgatttttt tgcATGATT TCATGCAGTT

          ***** * ***** *
          ***** * *****

3XLA000193      182  AGGAAAATGT TATATGTTCA TTATTTATGT AATGTATTGG CATGTTATTG
XL_BMP7_v1      165  AGGAAAATGT TATATGTTCA TTATTTCTGT AATGTATTGG CATGTTATTG
XL_BMP7_v2      166  AGGAAAATGT TATATGTTCA TTATTTCTGT AATGTATTGG CATGTTATTG

          *****
          *****

3XLA000193      232  TATCCAAGAA GAACAATGAC TGATGCAATC AAAATATGGC TCTTCAACAG
XL_BMP7_v1      215  TATCCAAGAA GAACAATGAC TGATGCAATC AAAATATGGC TCTTCAACAG
XL_BMP7_v2      216  TATCCAAGAA GAACAATGAC TGATGCAATC AAAATATGGC TCTTCAACAG
          *****
          *****

3XLA000193      282  AAAATGTACA ACTGCAATGG GCAAATAATT AAACAATTTT aagacacaaa
XL_BMP7_v1      265  AAAATGTACA ACTGCAATGG GCAAATAATT AAACAATTTT -----
XL_BMP7_v2      266  AAAATGTACA ACTGCAATGG GCAAATAATT AAACAATTTT -----

          *****
          *****

3XLA000193      332  aaaaaaaaaa aactca
XL_BMP7_v1      305  -----
XL_BMP7_v2      306  -----

```

Die in der UTR Datenbank gefundene BMP7 3'UTR zeigt am Anfang und Ende des Alignments Unterschiede zu den Varianten 1 und 2. Bei den in den Varianten 1 und 2 beobachteten Einfügungen bzw. Löschungen entspricht die 3'UTR aus der Datenbank der Variante 1, wobei die beobachtete Einfügung an Position 83

(entspricht der Einfügung an Position 63 in der Variante 1) zwei Nukleotide kürzer ist als in der Variante 1. Die Löschung ab Position 165 (entspricht Position 148 in Variante 1) stimmt exakt mit der Variante 1 überein.

Da die zusätzlichen Elemente in der Variante 1 und der Variante 2 zu klein sind, um ein potentiell regulatorisches Strukturmotiv zu bilden, wurde untersucht, wie sich diese Unterschiede auf die Gesamtstruktur der jeweiligen Variante auswirken.

Dazu wurden multiple Strukturalignments von 8 suboptimalen Strukturen der jeweiligen Variante berechnet. Zunächst wurden die Strukturen mit `mfold` berechnet. Die freie Energie der mfe-Struktur lag bei der ersten Variante bei -57.6 kcal/mol und bei der zweiten Variante bei -63.6 kcal/mol. Die zweite Variante kann somit etwas stabilere Sekundärstrukturen ausbilden. Abbildung 4.29 zeigt das Strukturalignment für die erste Variante und Abbildung 4.30 für die zweite Variante. Im Strukturalignment der zweiten Variante fällt eine verzweigte Stemloopstruktur auf, die komplett in rot dargestellt ist und somit in allen alignierten Strukturen vorkommt. Dieses Strukturmotiv ist im Alignment der ersten Variante nicht zu erkennen. Diese verzweigte Stemloop Struktur kann nur durch die Einfügung der 12 Nukleotide an Position 137 in Variante 2 gebildet werden. Diese eingefügten Nukleotide bilden die 3'-Seite des unteren Teils der zweiten Helix des Motivs (s. Abbildung 4.35).

Zur weiteren Überprüfung dieses Ergebnisses wurden die Basenpaarwahrscheinlichkeiten für die beiden Varianten 1 und 2 und die Sequenz aus der UTR Datenbank mit `RNAfold` berechnet und als Dotplot dargestellt. Die Analyse der Basenpaarwahrscheinlichkeiten bestätigt die Ergebnisse der Strukturalignments. Für die Variante 2 zeigt der Dotplot in Abbildung 4.32 in der Mitte der Diagonalen die verzweigte Stemloopstruktur. Die Abbildungen der Dotplots der ersten Variante (4.31) und der UTRdb Variante (4.33) verdeutlichen, daß dieses Strukturmotiv in diesen Varianten nicht gebildet werden kann. Das in Abbildung 4.34 gezeigte paarweise Strukturalignment der beiden Varianten 1 und 2 hebt diesen Unterschied noch einmal hervor. Der grün dargestellte Stemloop ist nur in der zweiten Variante möglich.

Geht man davon aus, daß es sich bei der zweiten Variante, die auf EST Daten basiert, tatsächlich um eine alternativ gespleißte BMP7 mRNA handelt, so würde der Unterschied nur in der 3'UTR liegen und hauptsächlich die verzweigte Stemloopstruktur betreffen. Somit könnte gerade das Auftreten des zusätzlichen Stemloops (s. Abbildung 4.34) die Bildung eines potentiell regulatorischen Motivs verursachen. Um zu überprüfen, ob dieses verzweigte Strukturmotiv in weiteren BMP7 3'UTRs oder in den UTRs anderer mRNAs vorkommt, wird im nächsten Abschnitt die Musterbeschreibung und Datenbanksuche für dieses Motiv beschrieben.

#### 4. Ergebnisse

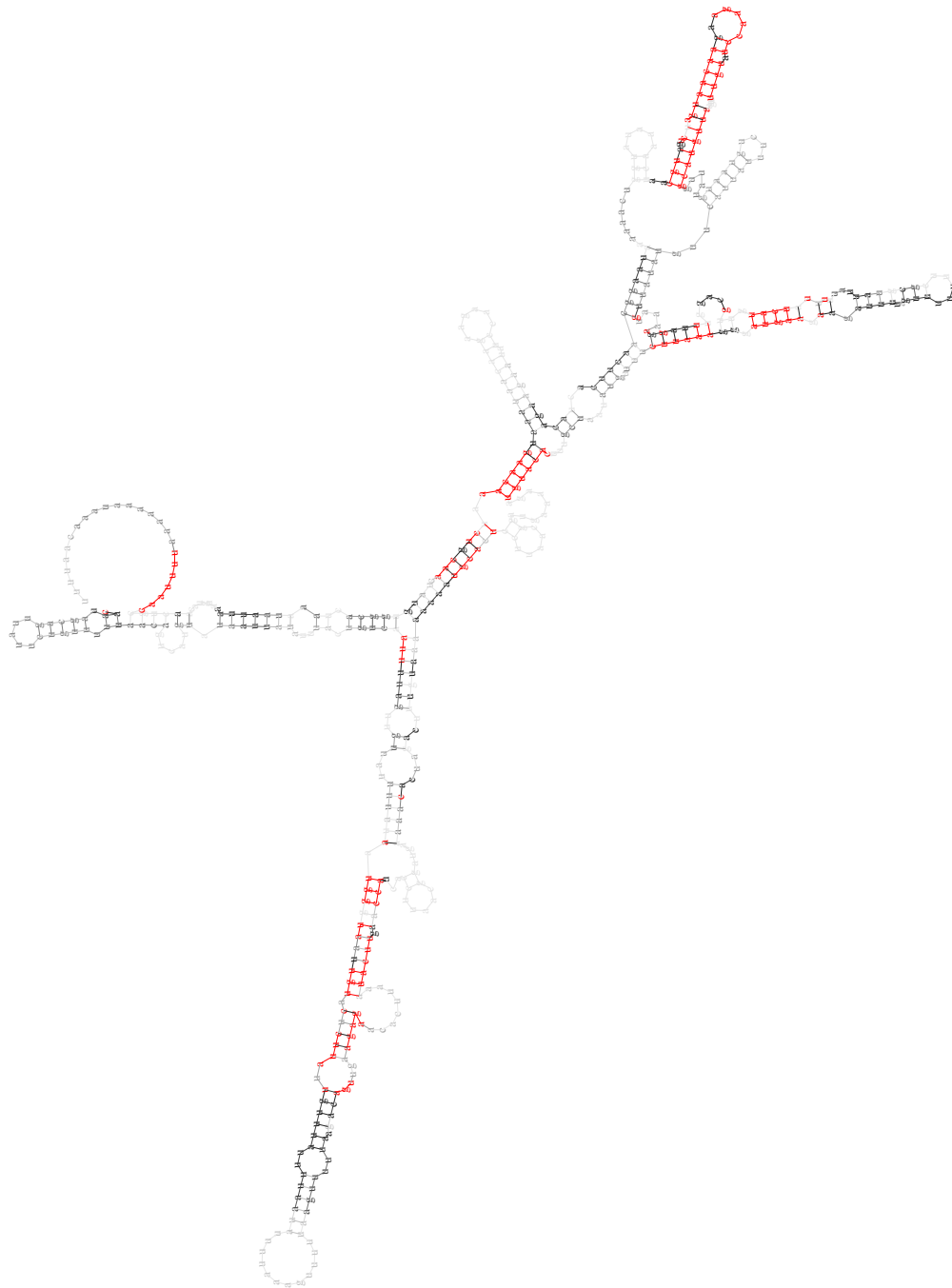


Abbildung 4.29.: Multiples globales Strukturalignment von 8 suboptimalen Strukturen der Variante 1 der BMP7 3'UTR.



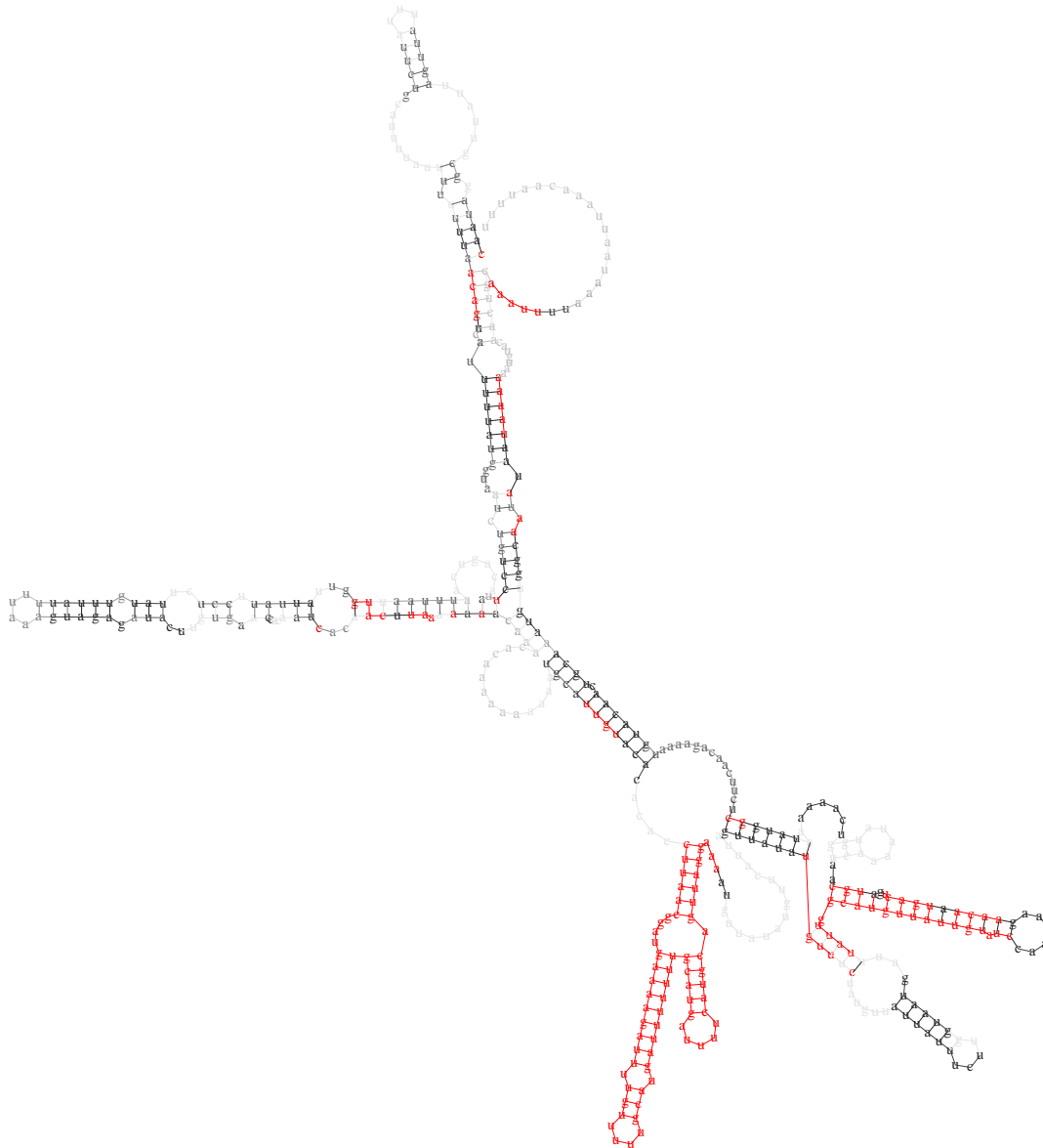


Abbildung 4.30.: Multiples globales Strukturalignment von 8 suboptimalen Strukturen der Variante 2 der BMP7 3'UTR. Die komplett in rot dargestellte verzweigte Stemloopstruktur kommt nur in dieser Variante vor.

#### 4. Ergebnisse

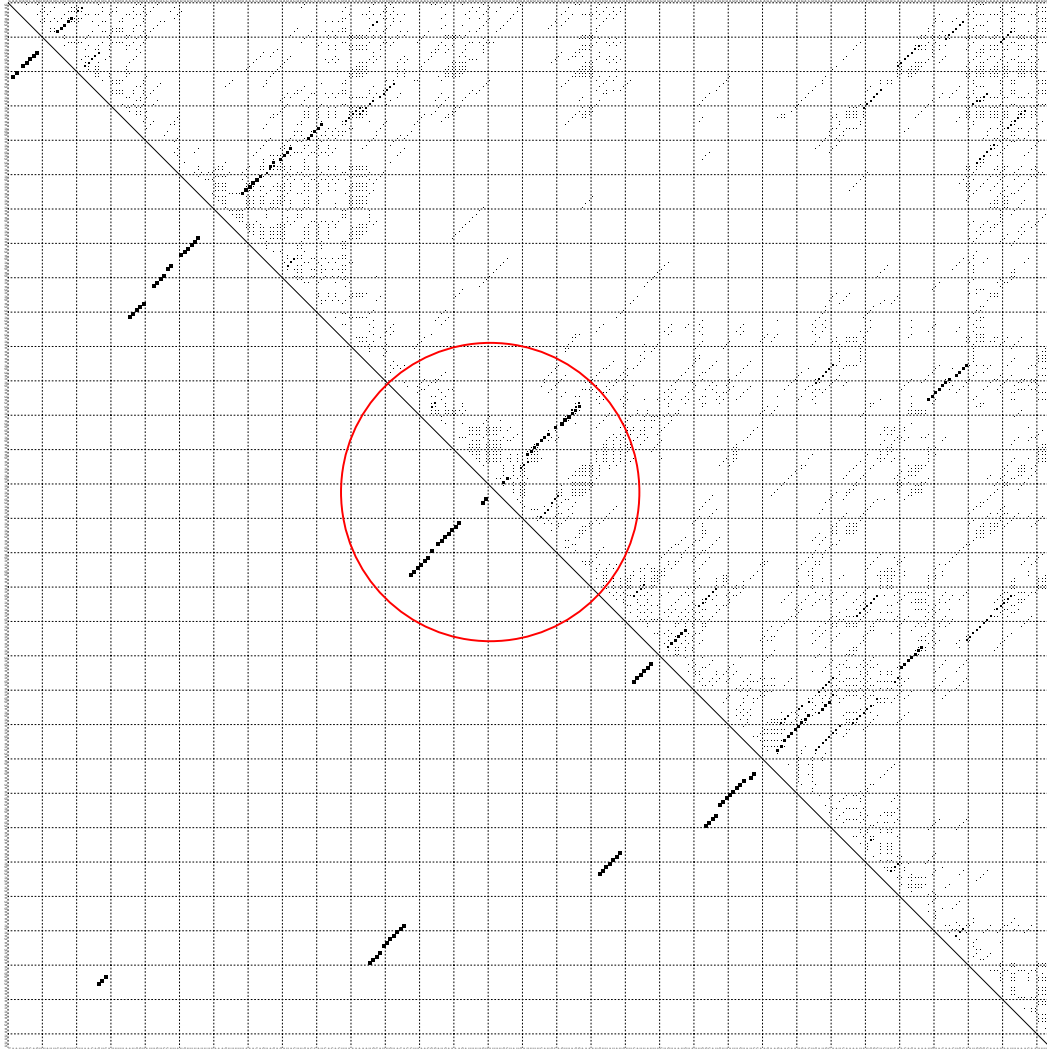


Abbildung 4.31.: Dotplot: Basenpaarwahrscheinlichkeiten (oben rechts) der Variante 1 der BMP7 3'UTR (unten links: energetisch günstigste Struktur). Der rote Kreis markiert den Bereich, in dem die verzweigte Stemloopstruktur liegen müßte.

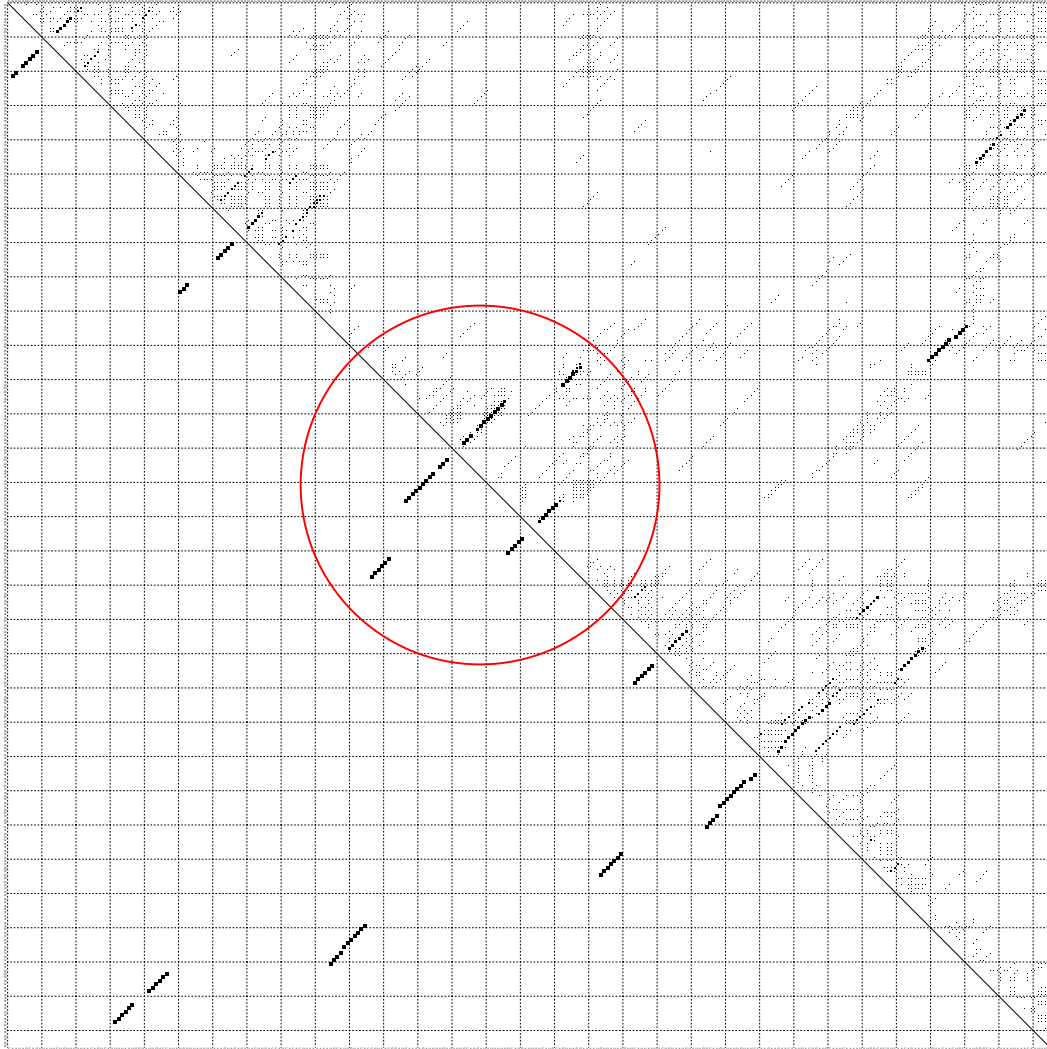


Abbildung 4.32.: Dotplot: Basenpaarwahrscheinlichkeiten (oben rechts) der Variante 2 der BMP7 3'UTR (unten links: energetisch günstigste Struktur). Der rote Kreis markiert den Bereich, in dem die verzweigte Stemloopstruktur liegt.

#### 4. Ergebnisse

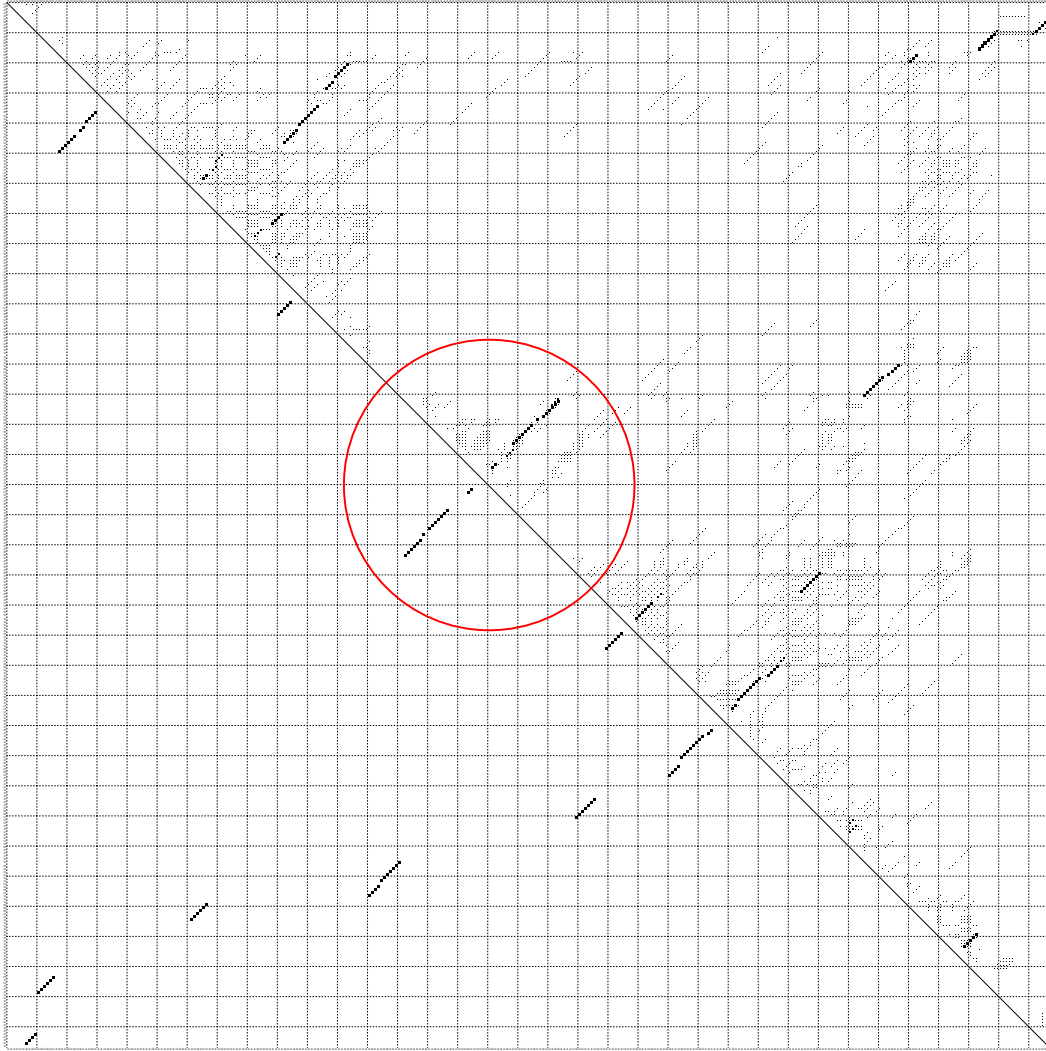


Abbildung 4.33.: Dotplot: Basenpaarwahrscheinlichkeiten (oben rechts) der BMP7 3'UTR aus der UTRdb (unten links: energetisch günstigste Struktur). Der rote Kreis markiert den Bereich, in dem die verzweigte Stemloopstruktur liegen müßte.

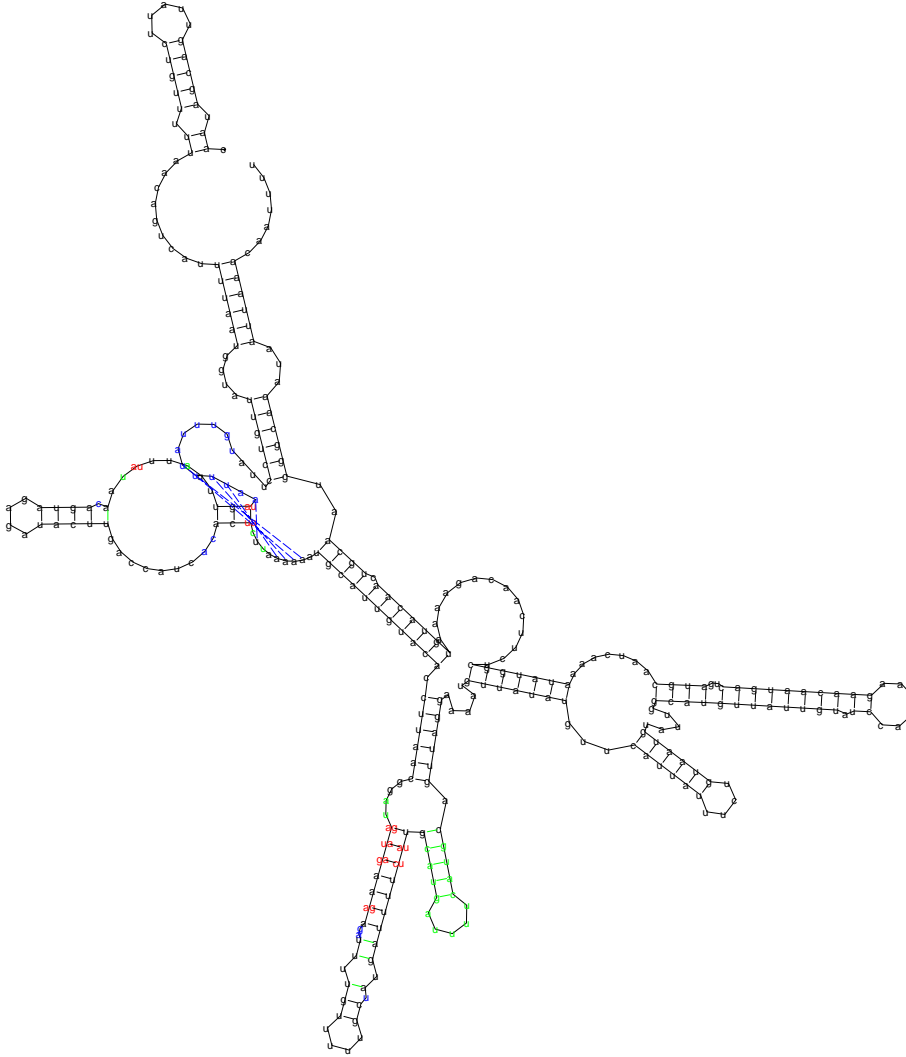


Abbildung 4.34.: Globales paarweises Strukturalignment der beiden BMP7 3'UTR Varianten. Der zusätzliche Stemloop in der Variante 2 ist in grün dargestellt.

#### 4. Ergebnisse

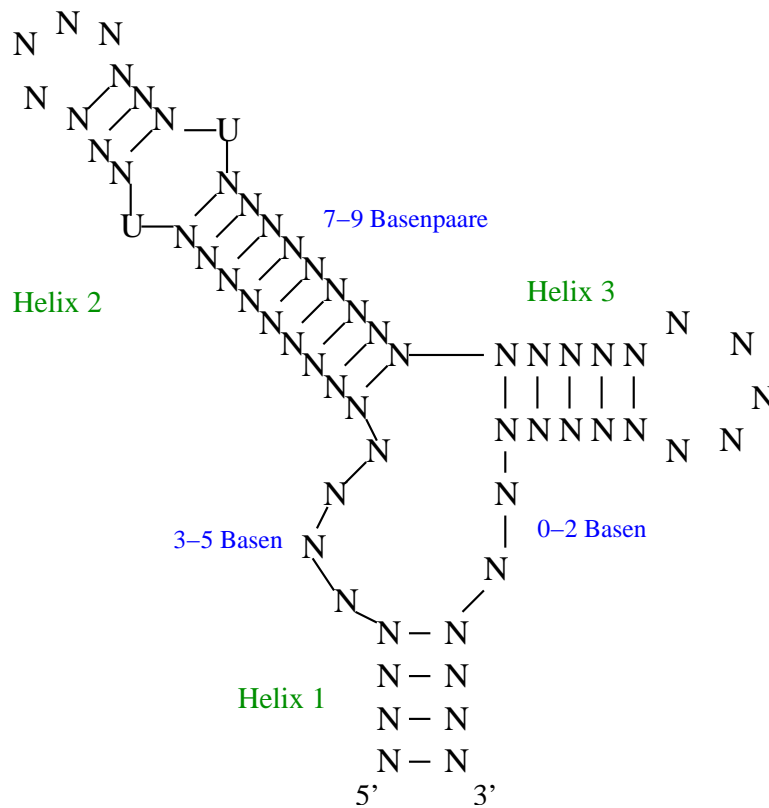


Abbildung 4.35.: Muster des verzweigten BMP7 Stemloops.

#### Musterbeschreibung und Datenbanksuche

Abweichend zum Vorgehen bei den Testsystemen und bei der Rab1A 3'UTR, wurde für die Beschreibung der verzweigten Stemloopstruktur der möglichen alternativen Spleißvariante der BMP7 3'UTR auf die Musterdefinition durch Signifikanzberechnungen verzichtet. Da das potentielle Strukturmotiv nur durch die Analyse einer einzigen Sequenz entdeckt wurde, wurde stattdessen direkt ein möglichst allgemeines Muster beschrieben. Dieses ist in der Abbildung 4.35 dargestellt. Die einzigen Sequenzvorgaben für das Muster waren die beiden ungepaarten Uracile in der zweiten Helix. Die Länge der ersten Helix wurde mit mindestens 4 festgelegt. Danach folgt ein 3-5 Nukleotide langer einzelsträngiger Bereich, an den sich der untere Teil der zweiten Helix mit einer variablen Länge von 7-9 anschließt. Auf die beidseitig ungepaarten Uracile folgt der obere Teil der zweiten Helix mit exakter Länge 3 und ein Loop der Länge 4. Die dritte Helix, die dem zusätzlichen Stemloop der zweiten Variante entspricht, hat eine vorgegebene Länge von 5, und der Loop hat eine Länge von ebenfalls 5. Die einzelsträngige Verbindung zwischen der ersten und dritten Helix kann zwischen 0 und 2 Basen lang sein. Die Beschreibung dieses Musters mit `RNAMotif` sieht wie folgt aus:

```

parms
    wc += gu;

descr
    h5( tag='stem1', len=4 )
        ss( minlen=3, maxlen=5 )
        h5( tag='stem2', minlen=7, maxlen=9)
            ss( len=1, seq="u")
                h5(tag='stem3', len=3)
                    ss( len=4 )
                        h3(tag='stem3')
                            ss( len=1, seq="u")
                                h3( tag='stem2')
                                    h5( tag='stem4', len=5)
                                        ss( len=5)
                                            h3(tag='stem4')
                                                ss(minlen=0, maxlen=2)
                                                    h3(tag='stem1')

```

Mit diesem Muster wurden die 3'UTR Kollektionen der UTR Datenbank durchsucht. Die Ergebnisse dieser Datenbanksuche sind in Tabelle 4.6 zusammengefaßt. Bei der Datenbanksuche wurden insgesamt 5 Treffer gefunden. Unter diesen Treffern war kein weiteres BMP Gen. Auch die BMP7 mRNA von *Xenopus* konnte nicht gefunden werden, da das Motiv in der Variante, wie sie in der UTRdb zu finden ist, nicht vorkommt. Zur Kontrolle, ob das Motiv auf 3'UTRs beschränkt ist, wurden auch die 5'UTR Kollektionen durchsucht. Hier konnten keine Treffer gefunden werden. Die Ergebnisse dieser Datenbanksuche und eine mögliche Bedeutung des vorhergesagten Strukturmotivs werden in Abschnitt 5.3 ausführlich diskutiert.

Da es sich bei diesem Muster um ein komplexeres als das IRE- bzw. Rab-Muster handelt, wurde auch hier ein Vergleich zwischen RNAMotif und HyPa für die benötigten Zeiten zur Mustersuche angestellt. In HyPa läßt sich das Muster aus Abbildung 4.35 wie folgt beschreiben:

```

bmp7_Muster = (stem1:=.{4}) .{3,5} (leftsidestem_a:=.{7,9}) U
(leftsidestem_b:=.{3}) (leftsidestemloop:=.{4}) ^(leftsidestem_b) U
^(leftsidestem_a) (rightsidestem:= .{5}) .{5} ^(rightsidestem)
.{0,2} ^(stem1);

```

In Tabelle 4.7 sind die ermittelten Zeiten für die Mustersuche in den einzelnen 3'UTR Kollektionen der UTRdb angegeben. Der Vergleich der Suchzeiten für

#### 4. Ergebnisse

Kollektion	Acc.Nr UTRdb (EMBL)	Organismus	Gen
Mensch	—	—	—
Nager	3MMU007732 (AB000499)	Maus	KQT2 mRNA
andere Säuger	—	—	—
andere Vertebraten	3XLA001652 (AB018694)	Xenopus	Xepsin mRNA
Invertebraten	3DME000470 (X54663)	Drosophila	Broad complex mRNA
Virus	—	—	—
Pflanzen	3TER000002 (AF034761) 3ZMA000245 (X67324)	Ringelblume  Mais	CoA-Reduktase mRNA  MFS18 mRNA
Pilze	—	—	—

Tabelle 4.6.: Ergebnis der Datenbanksuche für das BMP7 Muster

Kollektion	Datenbankgröße in Millionen Basen	RNAMotif Zeit in sec	HyPa Zeit in sec
3'UTR Mensch	18,91	88,4	11,2
3'UTR Nager	9,11	41,5	6,1
3'UTR andere Säuger	2,55	10,9	2,7
3'UTR andere Vertebraten	3,35	14,4	3,4
3'UTR Invertebraten	5,15	21,9	4,5
3'UTR Viren	3,36	15,1	3,1
3'UTR Pflanzen	4,33	16,8	4,5
3'UTR Pilze	0,47	1,7	1,0

Tabelle 4.7.: Vergleich von RNAMotif und HyPa für die Mustersuche nach dem BMP7 Muster aus Abbildung 4.35



das BMP7-Muster zeigt klare Vorteile bei HyPa. Während `RNAMotif` für die Suche nach dem Muster in der größten UTRdb 3'UTR Kollektion (Mensch: 18,9 Millionen Basen) 88,4 Sekunden benötigt, ist HyPa mit 11,2 Sekunden um fast den Faktor 8 schneller. Bei kleineren Kollektionen ist der Faktor noch ungefähr 5. Somit ist HyPa bei der Suche nach komplexeren Mustern gegenüber `RNAMotif` im Vorteil.

## 4.5. Experimente

In den folgenden Abschnitten werden Experimente beschrieben, mit denen eine biologische Funktion des vorhergesagten Stemloops in der Rab1A 3'UTR nachgewiesen werden sollte. Mit Hilfe von Verzögerungsgelen wurde zunächst überprüft, ob es eine spezifische Proteinbindung an den Stemloop gibt. Sollte das Strukturmotiv zum Beispiel eine Funktion für eine mögliche posttranskriptionelle Genregulation der Rab1A mRNA haben, ist solch eine spezifische Proteinbindung sehr wahrscheinlich. Anschließend wurde versucht, spezifische Bindeproteine zu isolieren und zu identifizieren. Für die Identifizierung wurden massenspektrometrische Verfahren eingesetzt. Durch *in vivo* Lokalisationsexperimente sollte die Verteilung der Rab1A mRNA in der Zelle untersucht werden.

### 4.5.1. Gelverzögerung

Durch Gelverzögerungsexperimente sollte eine spezifische Proteinbindung an den Stemloop in der Rab1A 3'UTR nachgewiesen werden. Das Prinzip dieser Methode ist in Abschnitt 3.3.2 beschrieben worden. Für die Untersuchungen wurden verschiedene RNA-Oligonukleotide verwendet. In Abbildung 4.36 sind zwei RNA-Oligonukleotide gezeigt, die die Sequenz des Menschen (und damit auch der Maus) haben und sich in der Länge der einzelsträngigen Bereiche vor und nach dem Stemloop unterscheiden. Das zweite Oligonukleotid wurde verwendet, um die Bedeutung des auf Sequenzebene hoch konservierten einzelsträngigen Bereiches auf der 5'-Seite zu untersuchen.

In dem ersten Gelverzögerungsexperiment wurde das RNA-Oligonukleotid mit dem kurzen einzelsträngigen 5'-Bereich eingesetzt. Die mit Digoxigenin markierten RNA-Oligonukleotide wurden durch Chemolumineszenz mit Hilfe eines Röntgenfilms sichtbar gemacht. Die Abbildung 4.37 zeigt die Autoradiographie dieses ersten Verzögerungsexperimentes. In Spur 1 wurde nur das mit Digoxigenin markierte RNA-Oligonukleotid aufgetragen. Durch die Zugabe von Proteinextrakt aus der Niere (Maus) in Spur 2 wurde eine starke Verzögerung erreicht, da die unverzögerte Bande fast vollständig verschwindet. Zusätzlich werden am oberen Rand zwei Banden sichtbar. Bei der oberen der beiden Banden, die in allen folgenden Spuren gleich stark ist, handelt es sich um Proteine, die mit dem Digoxigenin Antikörper kreuzreagieren. Das wird durch einen Vergleich mit Spur 6 deutlich, in der nur Proteinextrakt als Negativkontrolle aufgetragen wurde. Bei der darunter liegenden Bande könnte es sich um eine durch verzögerte RNA-Oligonukleotide verursachte Bande handeln. Der größte Teil der DIG-markierten Oligonukleotide ist jedoch nicht sichtbar, wie ein Vergleich mit der Stärke der Bande aus Spur 1 zeigt. Mögliche Gründe hierfür werden in Abschnitt 5.4.2 diskutiert. In den Spuren 3-5 wurde in ansteigenden Konzentrationen unmarkiertes RNA-Oligonukleotid zugegeben, wodurch die unverzögerte Bande des markierten Oligonukleotides wieder auftaucht. Dadurch konnte die Spezifität der Proteinbin-

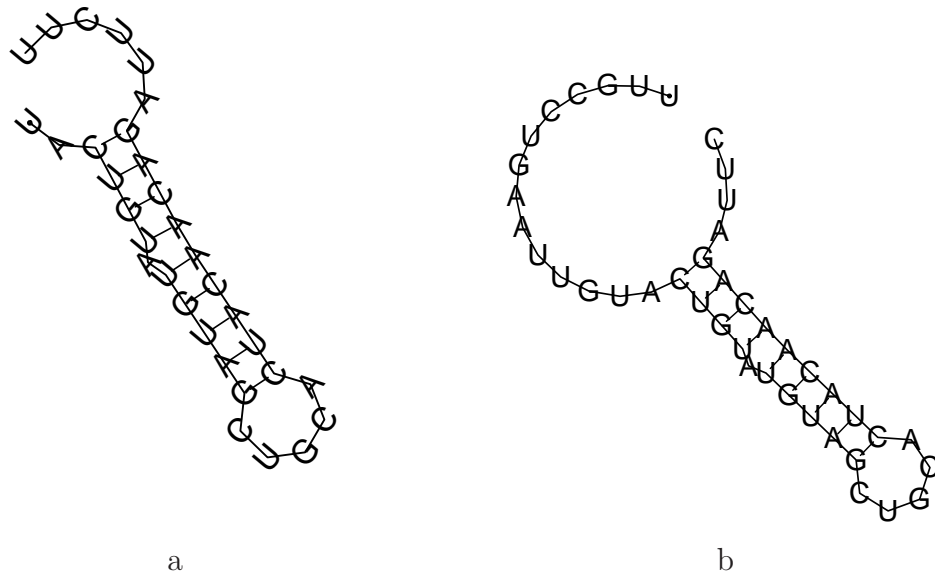


Abbildung 4.36.: RNA-Oligonukleotide, die für die Gelverzögerungsexperimente eingesetzt wurden. Das zweite Oligo (b) entspricht auf der 5' Seite der Vorhersage aus Abbildung 4.22.

dung an den Stemloop nachgewiesen werden. Unspezifische RNA-Bindeproteine wurden durch einen hohen Überschuß an Hefe Gesamt-RNA in den Spuren 2 bis 5 gebunden.

Diese Ergebnisse konnten auch für das zweite Oligonukleotid, das sich von dem ersten durch einen längeren einzelsträngigen Bereich auf der 5'-Seite unterscheidet, bestätigt werden. Dieses Oligonukleotid entspricht bezogen auf den 5' einzelsträngigen Bereich dem, welches durch die erste Vorhersagestrategie als potentielles regulatorisches Motiv gefunden werden konnte. Abbildung 4.38 zeigt die Autoradiographie der Gelverzögerung mit dem Biotin-markierten Oligonukleotid. Der Proteinextrakt wurde für dieses Experiment aus Herz hergestellt. Im Gegensatz zu den DIG-markierten Oligonukleotiden sieht man hier keine Kreuzreaktion mit dem Proteinextrakt (Spur 7). Bei diesem Experiment läßt sich die spezifische Proteinbindung nur indirekt zeigen, da keine verzögerten Banden sichtbar sind (s. Diskussion). Wie bei dem Experiment aus Abbildung 4.37 sieht man eine Zunahme der unverzögerten Bande nach Zugabe von nicht-markierten Oligonukleotiden (Spuren 3-6). Dies ist ein Hinweis auf eine spezifische RNA Bindung an das RNA-Oligonukleotid.

Nachdem eine spezifische Proteinbindung an den Stemloop in der Rab1A 3'UTR nachgewiesen werden konnte, wurde versucht, das Bindeprotein (oder mehrere) zu isolieren.

#### 4. Ergebnisse

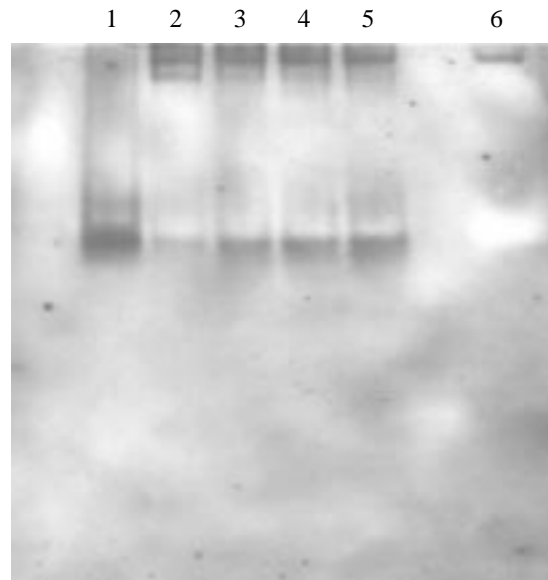


Abbildung 4.37.: Autoradiographie: Gelverzögerung. Spur 1: RNA-Oligo (DIG), Spur 2: RNA-Oligo(DIG)+Proteinextrakt (Niere), Spur 3: RNA-Oligo (DIG:unmarkiert=1:50)+Proteinext., Spur 4: RNA-Oligo (DIG:unmarkiert=1:150)+Proteinext., Spur 5: RNA-Oligo (DIG:unmarkiert=1:300)+Proteinextrakt, Spur 6: nur Proteinextrakt als Negativkontrolle für den DIG-Nachweis; in den Spuren 2-5 ist Hefe-RNA als unspezifischer Kompetitor im Überschuß vorhanden.

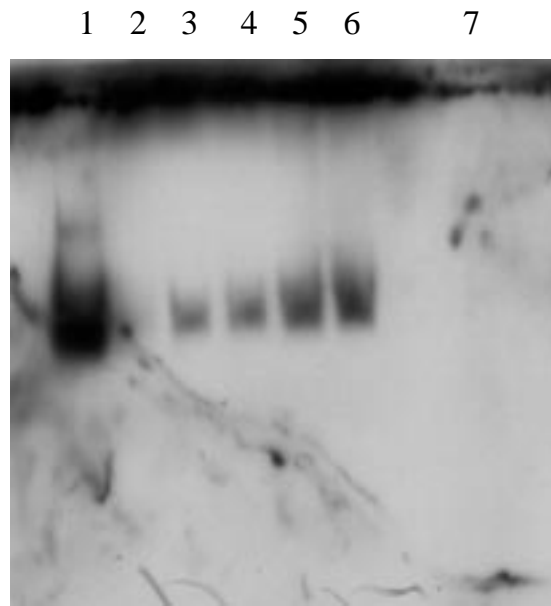


Abbildung 4.38.: Autoradiographie: Gelverzögerung. Spur 1: RNA-Oligo (Biotin), Spur 2: RNA-Oligo(Biotin)+Proteinextrakt (Herz), Spur 3: RNA-Oligo (Biotin:unmarkiert=1:50)+Proteinext., Spur 4: RNA-Oligo (Biotin:unmarkiert=1:100)+Proteinext., Spur 5: RNA-Oligo (Biotin:unmarkiert=1:200)+Proteinextrakt, Spur 6: RNA-Oligo (Biotin:unmarkiert=1:300)+Proteinextrakt, Spur 7: nur Proteinextrakt als Negativkontrolle für den Biotin-Nachweis; in den Spuren 2-6 ist Hefe-RNA als unspezifischer Kompetitor im Überschuß vorhanden.

### 4.5.2. Proteinisolation und Identifizierung

Zur Aufreinigung und Isolierung der spezifischen RNA-Bindeproteine wurde das in Abschnitt 3.3.7 beschriebene Verfahren mit magnetischen Beads eingesetzt. Für die erfolgreiche Durchführung dieses Ansatzes war zunächst eine aufwendige Optimierungsphase notwendig. Dabei wurde das Verhältnis von markierten Dynabeads zu Proteinextrakt und unspezifischer RNA (Hefe-RNA) so eingestellt, daß möglichst keine oder nur wenige unspezifische Bindeproteine mit aufgereinigt wurden. In der Abbildung 4.39 ist ein Proteingel gezeigt, in dem die Proteinbanden mittels Coomassiefärbung sichtbar gemacht worden sind. Im ersten Ansatz

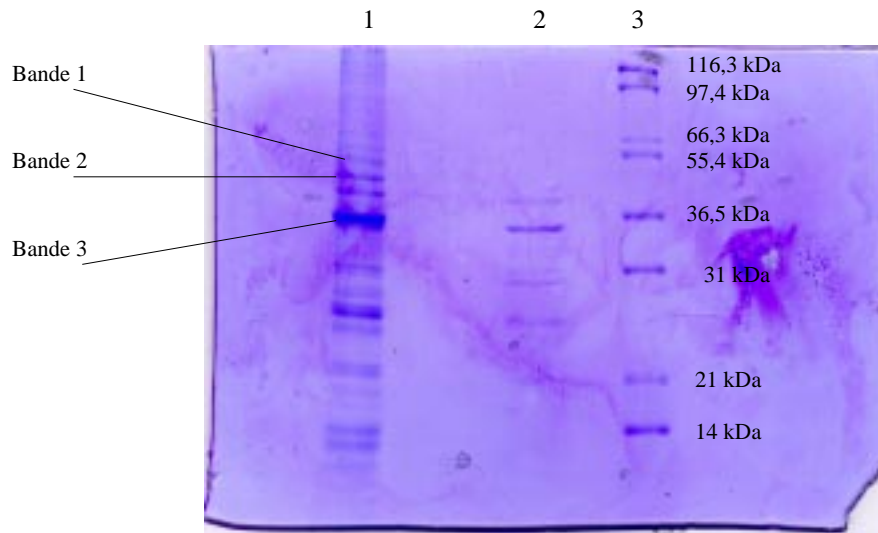


Abbildung 4.39.: 10% SDS-Proteingel mit Coomassiefärbung. Spur 1: Proteine nach Aufreinigung mit markierten Beads. Spur 2: Proteine nach Aufreinigung mit unmarkierten Beads. Spur 3: Proteingrößenstandard. Die Banden 1 bis 3 wurden mittels MALDI-TOF analysiert.

(Spur 1) wurden 500  $\mu\text{g}$  Beads (das entspricht ca. 1  $\mu\text{g}$  RNA-Oligo, bei optimaler Markierung), die mit dem RNA-Oligo aus Abbildung 4.36 (b) markiert waren, 20  $\mu\text{l}$  Proteinextrakt (Herz) und 250  $\mu\text{g}$  Hefe-RNA als unspezifischer Kompetitor eingesetzt. Der zweite Ansatz (Spur 2) unterschied sich nur durch den Einsatz unmarkierter statt markierter Beads und wurde als Negativkontrolle eingesetzt, um eine direkte Proteinbindung an die Beads nachzuweisen. In Spur 3 wurde ein Proteingrößenstandard aufgetragen. Im Vergleich zu den unmarkierten Beads konnte eine sehr viel stärkere Proteinbindung an die markierten Beads beobachtet werden. Somit kann davon ausgegangen werden, daß die meisten Proteine an die RNA-Oligonukleotide und nicht direkt an die Beads binden. Die Mehrheit dieser Proteine sollte spezifisch an das RNA-Oligonukleotid binden, da ein Überschuß an

Hefe-RNA eingesetzt wurde, um unspezifische Proteinbindungen abzufangen. Somit wäre es möglich, daß mehrere Proteine an den RNA-Stemloop binden können. Es wäre ebenfalls möglich, daß ein größerer Komplex an den Stemloop bindet, der erst durch die denaturierende Gelelektrophorese in seine einzelnen Komponenten gespalten wurde.

Nach der Proteinaufreinigung wurde versucht, spezifische Bindeproteine zu identifizieren. Dazu wurden die markierten Proteinbanden (s. Abbildung 4.39) ausgeschnitten und anschließend mittels MALDI-TOF analysiert. Das Prinzip der Proteinidentifizierung mittels MALDI-TOF Spektrometrie wurde in Abschnitt 3.3.8 erläutert. In den Abbildungen 4.40 bis 4.42 sind die Massenspektren für die aus dem Gel (Abbildung 4.39) ausgeschnittenen Proteinbanden 1, 2 und 3 abgebildet. Für das Protein der Bande 1 ergibt sich aus der Gelelektrophorese eine Größe von ungefähr 53 kDa, für das der Bande 2 von ca. 50 kDa und für das der Bande 3 eine Größe von ungefähr 37 kDa. Die durch die MALDI-TOF Spektrometrie bestimmten Fragmentmassen (s. Spektren) ergeben für alle analysierten Banden deutlich größere Proteine, so daß davon ausgegangen werden muß, daß die Banden Mischungen mehrerer Proteine sind. Die eigentliche Identifizierung der Proteine erfolgt durch einen Vergleich der ermittelten Spektren mit Spektren in einer Datenbank. Für die Auswertung der Massenspektren wurde eine Datenbank mit simulierten Proteinverdaus der Maus verwendet, da die Proteine aus Mausgeweben aufgereinigt wurden. Da es sich um Proteingemische handelt, ist es schwierig, eindeutig signifikante Treffer zu bekommen (s. Diskussion). Bei allen drei Spektren lieferte die Auswertung eine Vielzahl möglicher Treffer, wobei die jeweils wahrscheinlichsten Treffer als einzige in dem zu erwartenden Größenbereich lagen.

Die Analyse der Bande 1 lieferte den sogenannte "Rab GDP dissociation inhibitor alpha" (GDI-1) als möglichen Treffer. In der Datenbank wird für das Protein eine Größe von 50,5 kDa angegeben, während es in der Literatur mit einer Größe von 55 kDa beschrieben wird (Luan et al., 2000). Für die Bande 2 ergibt sich das Keratin 7 Protein mit einer Größe von 50,7 kDa als wahrscheinlichster Treffer. Grundsätzlich müssen Keratine als Ergebnis von MALDI Analysen kritisch betrachtet werden, da es sich um Kontaminationen handeln könnte (z.B. Hautschuppen). Daß es sich hier jedoch um einen biologisch relevanten Treffer handeln könnte, wird in Abschnitt 5.4.2 diskutiert.

Für die Bande 3 wurde eine cDNA (RIKEN cDNA 4930547K05) als wahrscheinlichster Treffer gefunden. Das von dieser cDNA kodierte Protein hat eine Größe von 41,3 kDa und einen berechneten isoelektrischen Punkt von 9,65. Das Protein enthält zwei RNA Erkennungsmotive und wird deswegen als RNA-Bindeprotein klassifiziert. Dieses Protein könnte somit direkt an den RNA-Stemloop der Rab1A 3'UTR binden. Es ist nicht bekannt, welche Funktion dieses Protein haben könnte und in welchen Geweben es überhaupt gebildet wird, da man nur die entsprechende cDNA kennt. Aufgrund seiner potentiellen RNA Bindeeigenschaften ist es dennoch ein guter Kandidat für eine mögliche posttranskriptionelle Regulation der

#### 4. Ergebnisse

Rab1A mRNA. Die Schwierigkeit der Reproduzierbarkeit der MALDI-Analysen zeigte sich bei einer weiteren Analyse der Bande 3, die aus einem anderen Proteinbindungsexperiment stammte (Gel nicht gezeigt). Dieses Spektrum (s. Abbildung 4.43) zeigte nur wenige Fragmente. Für die Auswertung dieses Spektrums mußte die NCBI Inr Datenbank verwendet werden, da in der Mausdatenbank kein Treffer gefunden werden konnte. Als einziger Treffer im zu erwartenden Größenbereich wurde das humane Protein Sirtuin 6 (36 kDa) identifiziert. Die unterschiedlichen MALDI Ergebnisse für diese Bande, sowie die mögliche biologische Bedeutung des Sirtuin 6 werden in den Abschnitten 5.4.2 und 5.4.3 ausführlich diskutiert.



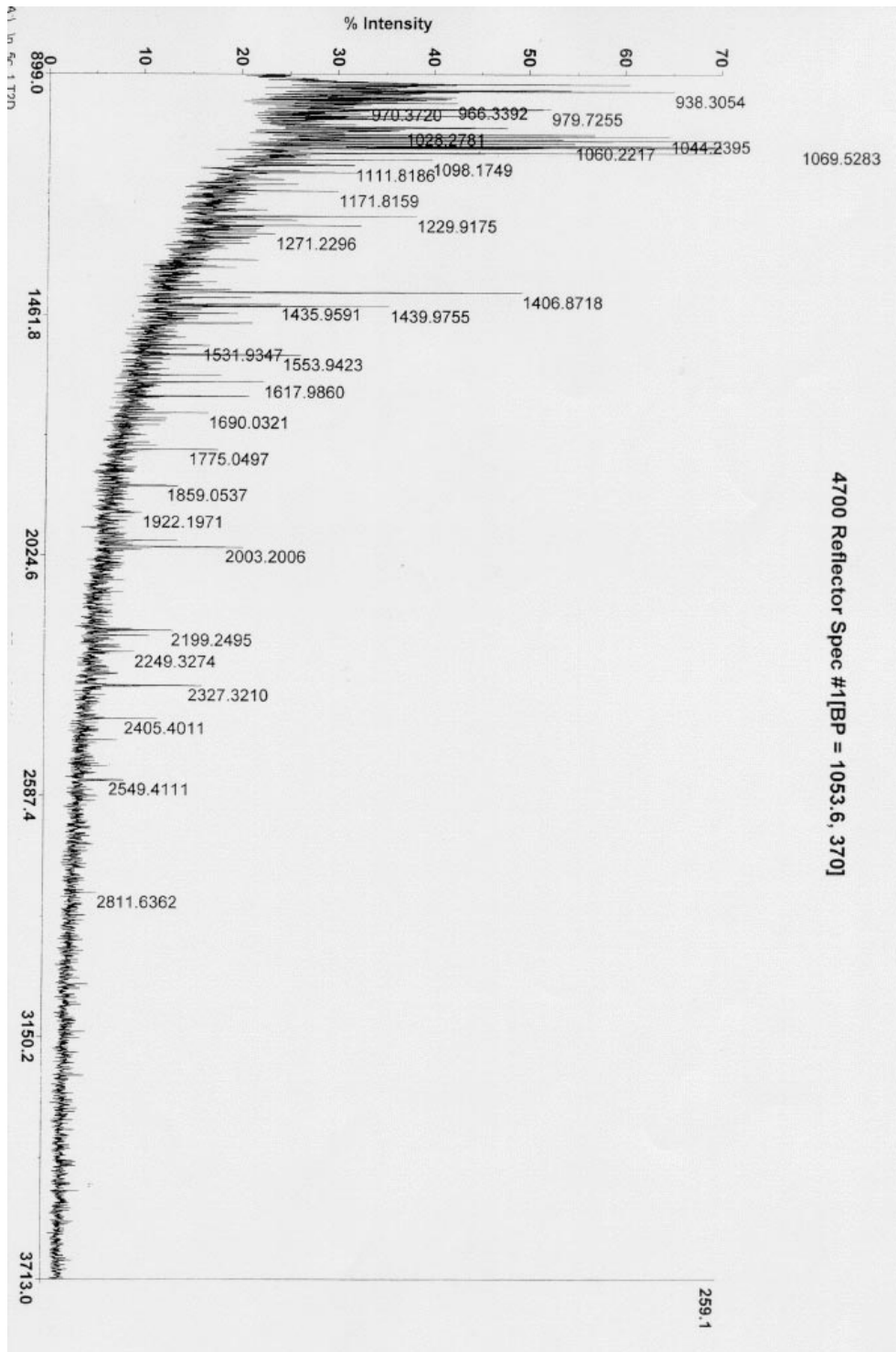


Abbildung 4.40.: Massenspektrum der Proteinbande 1 aus dem SDS-Gel (Abbildung 4.39).

#### 4. Ergebnisse

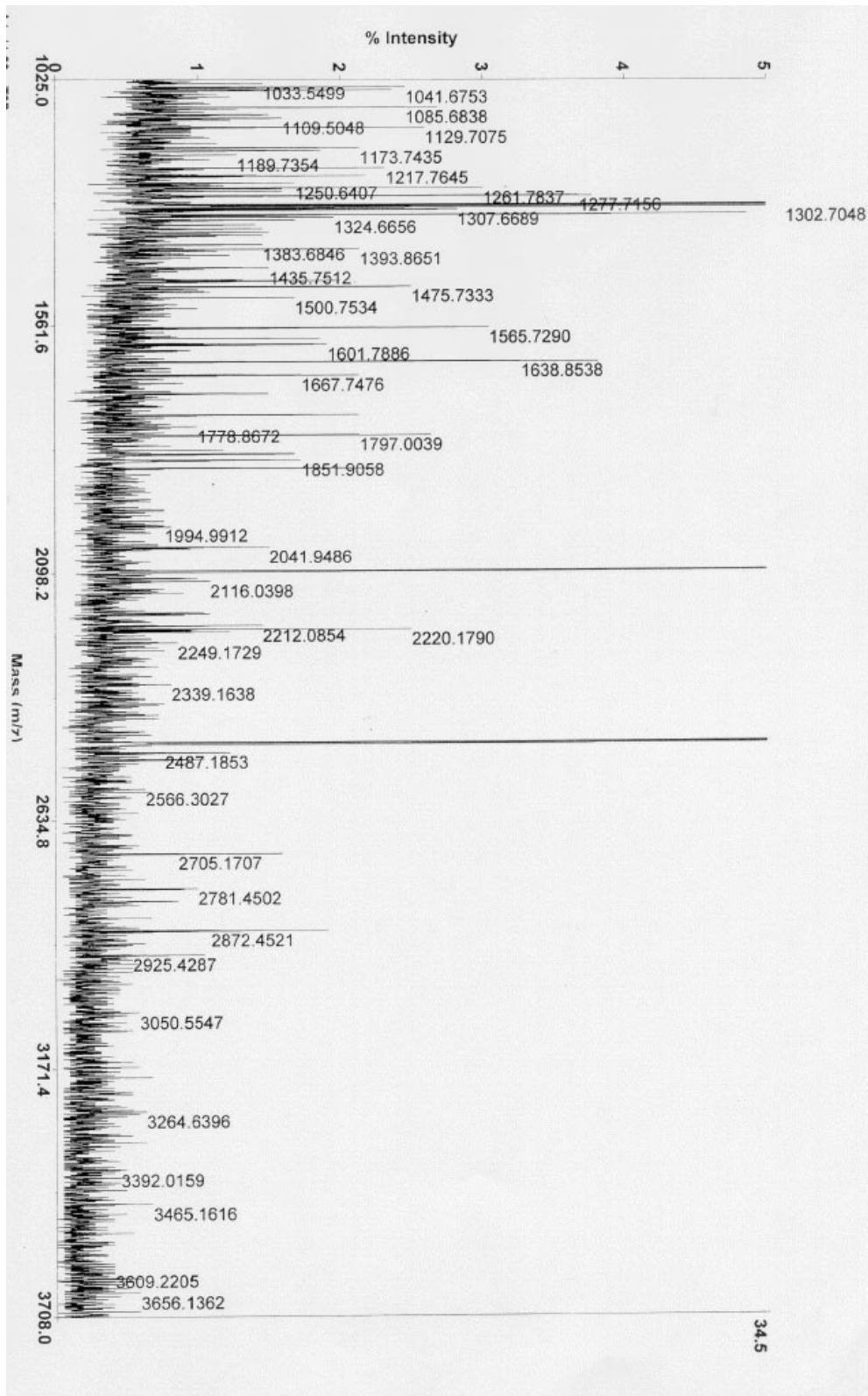


Abbildung 4.41.: Massenspektrum der Proteinbande 2 aus dem SDS-Gel (Abbildung 4.39).

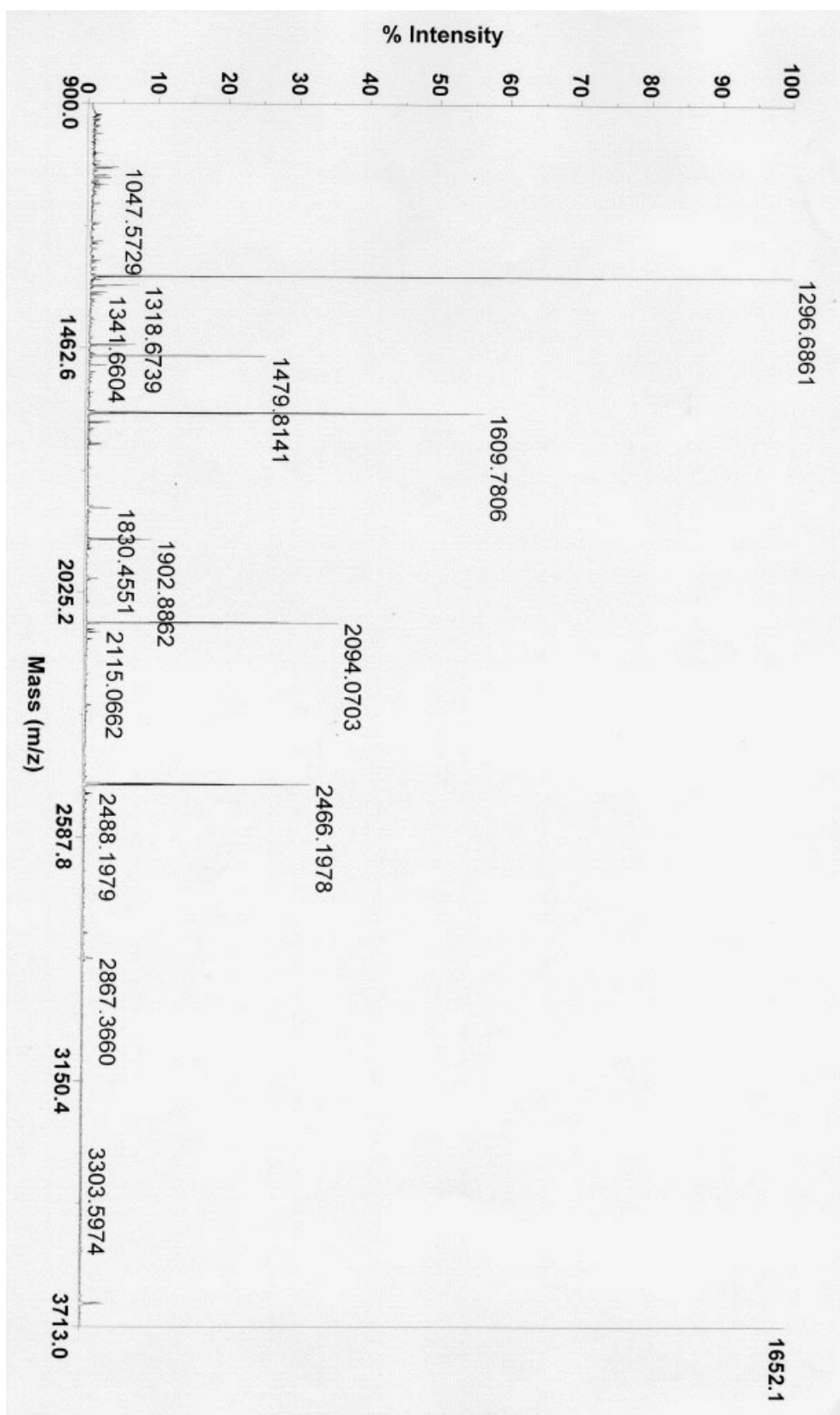


Abbildung 4.42.: Massenspektrum der Proteinbande 3 aus dem SDS-Gel (Abbildung 4.39).

#### 4. Ergebnisse

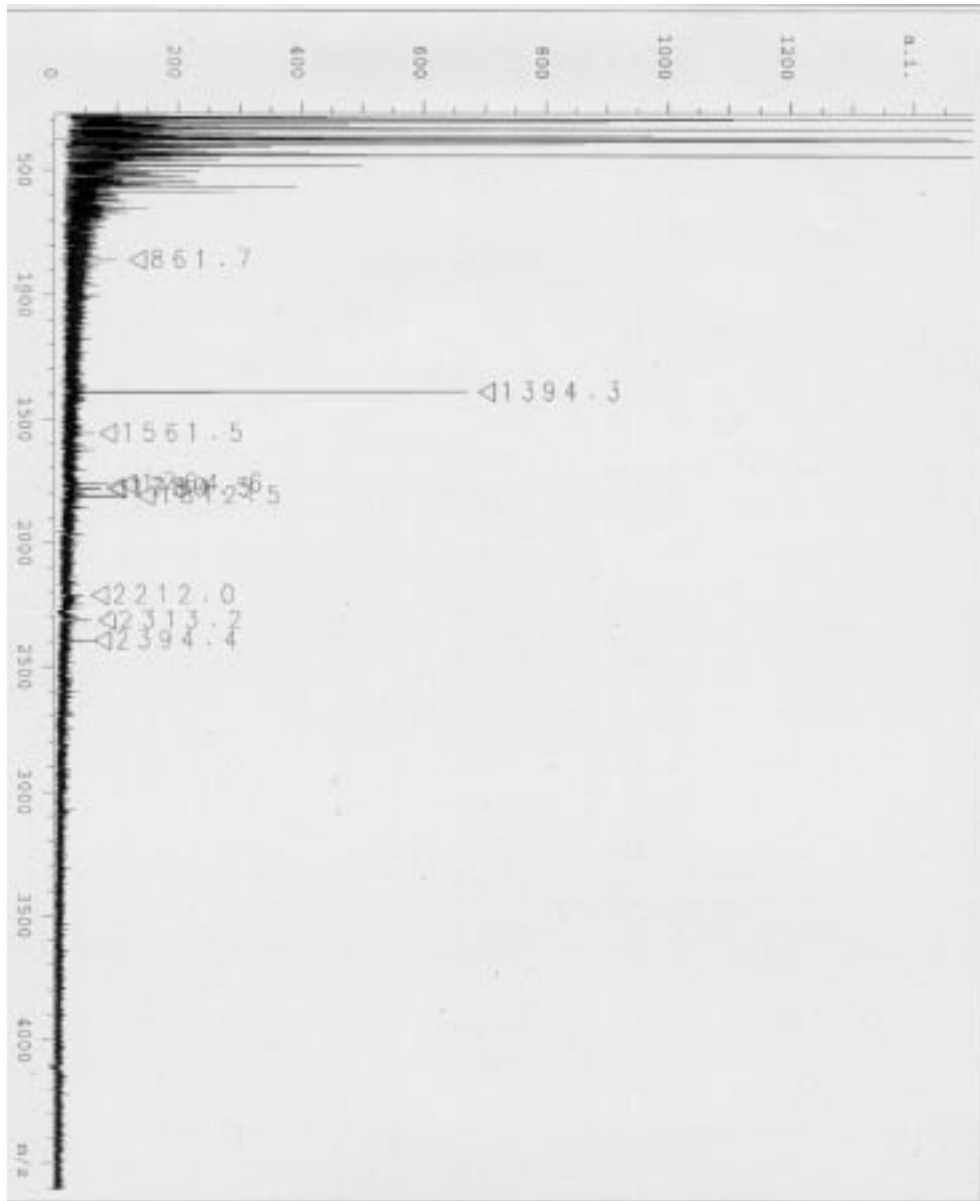


Abbildung 4.43.: Massenspektrum der Proteinbande 3 aus dem SDS-Gel eines weiteren Proteinbindungsexperimentes (Gel nicht gezeigt). Das Spektrum unterscheidet sich deutlich von dem in Abbildung 4.42 gezeigten Spektrum.

### 4.5.3. *in vivo* Lokalisationsuntersuchungen

In Abschnitt 2.4 der Einleitung wurde beschrieben, daß spezielle Motive in den 3' untranslatierten Regionen einen Einfluß auf die Lokalisation einer mRNA haben können. Da das Rab1 Protein hauptsächlich in der Umgebung des Golgi-Apparates zu finden ist, wäre es denkbar, daß auch die Rab1A mRNA dort lokalisiert und translatiert wird. Um eine mögliche Lokalisation der Rab1A mRNA in der Zelle nachzuweisen, wurden *in vivo* Hybridisierungsexperimente durchgeführt. Hierfür wurde die in Abschnitt 3.3.6 beschriebene Technik des Fluoreszenz Resonanz Energietransfers (FRET) verwendet. Im folgenden werden die Ergebnisse einer solchen Untersuchung beschrieben.

Die verwendeten humanen Myelomzellen wurden mit den Fluorescein markierten DNA-Oligonukleotiden und den Texas Red markierten DNA-Oligonukleotiden transfiziert. Die DNA-Oligonukleotide (s. Tabelle 3.2) hybridisieren im codierenden Bereich der Rab1A mRNA, wobei der Abstand so gewählt wurde, daß ein FRET möglich war. Die Abbildungen 4.44 bis 4.46 zeigen dieselbe Zelle bei unterschiedlicher Anregungs- bzw. Emissionswellenlänge. Bei der in der Abbildung 4.44 gezeigten Zelle wurde spezifisch das Fluorescein mit einer Wellenlänge von 488 nm angeregt und die Fluoreszenz des Fluoresceins in einem Bereich von 500-600 nm detektiert. In der Abbildung 4.45 wird dieselbe Zelle gezeigt, wobei hier

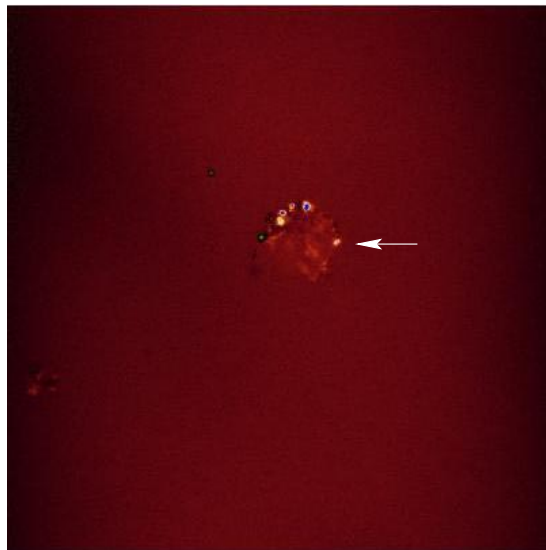


Abbildung 4.44.: Transfizierte Myelomzelle nach Anregung bei 488 nm und Detektion zwischen 500-600 nm. Die Fluoreszenz wird von Fluorescein markierten Oligonukleotiden verursacht. Der Pfeil markiert eine Stelle, an der nur Fluorescein und kein Texas Red zu finden ist (Erklärung s. Text).

#### 4. Ergebnisse

gezielt das Texas Red mit einer Wellenlänge von 543 nm angeregt und die Fluoreszenz des Texas Red in einem Bereich von 600-700 nm detektiert wurde. Der in



Abbildung 4.45.: Transfizierte Myelomzelle nach Anregung bei 543 nm und Detektion zwischen 600-700 nm. Die Fluoreszenz wird von Texas Red markierten Oligonukleotiden verursacht.

Abbildung 4.44 durch einen Pfeil markierte fluoreszierende Bereich ist dadurch gekennzeichnet, daß sich dort nur Fluorescein und kein Texas Red befindet, da dieser Bereich in der Abbildung 4.45 nicht fluoresziert. Die Abbildung 4.46 zeigt den eigentlichen FRET. Hier wurde die Zelle mit einer Wellenlänge von 488 nm angeregt. Dadurch wird nur das Fluorescein, nicht aber das Texas Red angeregt. Die Fluoreszenz wurde in einem Wellenlängenbereich von 605-625 nm detektiert. In diesem Wellenlängenbereich kann für Fluorescein keine Fluoreszenz detektiert werden, so daß es sich bei der beobachteten Fluoreszenz um die von Texas Red handeln muß. Die angeregten Fluoresceinmoleküle haben somit ihre, durch die Anregung erhaltene, Energie auf die Texas Red Moleküle übertragen, wodurch die so angeregten Texas Red Moleküle die Energie als Fluoreszenz wieder abgegeben haben. Da dies nur möglich ist, wenn die beiden Fluoreszenzfarbstoffe durch korrekte Hybridisierung an die Rab1A mRNA in der richtigen Position sind, muß es sich bei den fluoreszierenden Bereichen um solche handeln, in denen Rab1A mRNA Moleküle lokalisiert sind.



Abbildung 4.46.: Transfizierte Myelomzelle nach Anregung bei 488 nm und Detektion zwischen 605-625 nm. Die angeregten Fluorescein Moleküle haben die Energie durch Resonanzenergietransfer an die Texas Red Moleküle abgegeben, deren Fluoreszenz beobachtet wird (s. Pfeil).

#### 4.5.4. Zusammenfassung der experimentellen Ergebnisse

Die Ergebnisse der Experimente bestätigen eine mögliche biologische Funktion des vorhergesagten Stemloops in der Rab1A 3'UTR. Durch die Gelverzögerungsexperimente konnte eine spezifische Proteinbindung an den RNA-Stemloop nachgewiesen werden. Für beide verwendeten Oligonukleotide konnte gezeigt werden, daß die Zugabe von Proteinextrakt das Laufverhalten der Oligos im Gel stark beeinflusst, da ein Großteil der Oligos nicht mehr in die Gelmatrix einlaufen konnte. Dies könnte dadurch erklärt werden, daß ein großer Proteinkomplex an das RNA-Oligonukleotid bindet. Auch die Ergebnisse der anschließend durchgeführten Proteinisolierung mittels magnetischer Beads lassen solch eine Deutung zu. Hier konnte eine Vielzahl von Proteinbanden nach der Isolierung nachgewiesen werden. Bei diesen könnte es sich um die Teile eines größeren Proteinkomplexes handeln. Die MALDI-TOF Analysen haben ergeben, daß es sich bei einer dieser Banden um ein 41 kDa großes RNA-Bindeprotein handeln könnte. Die *in vivo* Lokalisationsuntersuchungen ergaben erste Hinweise darauf, daß die Rab1A mRNA im Cytoplasma der Zelle nur an bestimmten Orten vorkommt. Eine mögliche Funktion des vorhergesagten Stemloops für solch eine Lokalisation wird im folgenden Kapitel diskutiert.

#### 4. Ergebnisse



## 5. Diskussion

Im Rahmen dieser Arbeit wurde ein umfassender bioinformatischer Ansatz zur Vorhersage und Validierung regulatorischer RNA-Struktur motive entwickelt. Obwohl die Anwendung auf die Untersuchung untranslatierter Regionen von mRNAs begrenzt wurde, lassen sich die entwickelten Strategien auch auf jede andere funktionale RNA anwenden. Durch den Einsatz verschiedener experimenteller Methoden wurde die Qualität der bioinformatischen Ansätze bestätigt. Darüber hinaus konnte eine mögliche biologische Bedeutung eines im Rahmen dieser Arbeit neu vorhergesagten regulatorischen RNA-Motives durch die experimentellen Ansätze erhärtet werden. Im folgenden werden zunächst die entwickelten Strategien zur Vorhersage potentiell regulatorischer RNA Struktur motive miteinander verglichen. Anschließend erfolgt eine vergleichende Diskussion bezüglich der im Rahmen dieser Arbeit entwickelten Strategien mit anderen publizierten Vorhersagestrategien. Desweiteren wird die Bedeutung der Musterdefinition und Datenbanksuche für den Gesamtansatz bewertet. Die biologische Bedeutung der neu vorhergesagten RNA-Motive in der BMP7 3'UTR und der Rab1A 3'UTR wird ausführlich diskutiert, letzteres vor allem im Hinblick auf die hier durchgeführten Experimente.

### 5.1. Vorhersagestrategien im Vergleich

Die Erkenntnisse über die Bedeutung regulatorischer RNA-Struktur motive in UTRs für die posttranskriptionelle Genregulation haben in den letzten Jahren stark zugenommen (Pesole et al., 2001). Dennoch ist die Anzahl bekannter regulatorischer Motive in UTRs relativ klein (Pesole et al., 2002), vor allem wenn man diese Motive mit der Anzahl bekannter Transkriptionsfaktorbindestellen (Wingender et al., 2000) auf DNA Ebene vergleicht. Die Hauptursache hierfür liegt darin, daß regulatorische Motive in UTRs in den seltensten Fällen reine Sequenz motive sind, die mit entsprechenden Programmen relativ einfach gefunden werden können (Conclin et al., 2002). Die Tatsache, daß es sich bei regulatorischen Motiven in UTRs häufig um kombinierte Sequenz- und Struktur motive handelt, macht die Vorhersage sehr viel anspruchsvoller. Im Rahmen meiner Arbeit wurden Strategien für solch eine Vorhersage entwickelt, die im folgenden verglichen werden.

### 5.1.1. Kombination von Sequenzalignment und Strukturvorhersage

In dieser Arbeit wurden zwei unterschiedliche Strategien zur Vorhersage regulatorischer RNA-Motive entwickelt. Es konnte gezeigt werden, daß die erste Strategie, die auf einem Sequenzalignment mit DIALIGN2 und anschließender Strukturvorhersage beruht, unter bestimmten Voraussetzungen gute Ergebnisse liefert. Wenn ein potentiell regulatorisches RNA-Motiv eine ausreichende Sequenzkonservierung aufweist, wie es bei dem Stemloop-Motiv in den Histon 3'UTRs der Fall ist, kann dieses Motiv durch ein multiples Sequenzalignment gefunden werden. Durch die nachfolgende Strukturvorhersage für einen durch das Sequenzalignment definierten Bereich läßt sich zeigen, ob die sequenzkonservierten Bereiche auch eine gemeinsame Struktur ausbilden können. So konnte mit dieser Strategie der regulatorische Stemloop in den Histon 3'UTRs, welcher zur Stabilisierung der Histon mRNAs dient, vorhergesagt werden. Auch das im Rahmen dieser Arbeit neu vorhergesagte potentiell regulatorische Motiv in der Rab1A 3'UTR ließ sich durch das Alignment der 3'UTRs von Mensch und Zitterrochen identifizieren. Der Einsatz des Programmes DIALIGN2 zur Berechnung des Sequenzalignments war bei der Entwicklung dieser ersten Vorhersagestrategie der entscheidende Faktor. Während man für ein paarweises Alignment auch den Smith-Waterman Algorithmus (Smith und Waterman, 1981) zum Auffinden lokaler Ähnlichkeiten einsetzen könnte, ermöglicht DIALIGN2 die Berechnung multipler globaler Alignments, in denen sich aber auch lokal ähnliche Bereiche finden lassen, wenn sie an ähnlichen Positionen in den Sequenzen liegen. Somit ist es einem Programm wie ClustalW, das global verwandte Sequenzen korrekt alignieren kann, für den Einsatz zur Motivvorhersage klar überlegen.

Diese erste Vorhersagestrategie stößt an ihre Grenzen, wenn die untersuchten Sequenzen global zu ähnlich sind, das Motiv zu wenig Sequenzkonservierung enthält, oder das regulatorische Motiv in unterschiedlichen UTRs an völlig verschiedenen Positionen lokalisiert ist. Der letzte Fall läßt sich durch den Einsatz eines lokalen Sequenzalignmentprogramms, wie dem Smith-Waterman Algorithmus umgehen, wobei hier nur paarweise Alignments durchgeführt werden können. Das Testsystem mit den modifizierten IREs zeigte, daß die erste Strategie bei regulatorischen Motiven mit variabler Länge und sehr wenig Sequenzkonservierung nicht mehr erfolgreich eingesetzt werden kann. Das Sequenzalignment der Rab1A 3'UTRs von Mensch und Maus verdeutlichte, daß bei zu ähnlichen Sequenzen ebenfalls kein lokales Motiv gefunden werden kann. Durch das Alignment mit der Rab1A 3'UTR des Zitterrochen, die global sehr viel weniger Sequenzkonservierung mit den anderen Rab1A 3'UTRs aufweist, ließ sich jedoch ein lokales Motiv vorhersagen. Die erste Vorhersagestrategie läßt sich somit dann am erfolgreichsten einsetzen, wenn ein potentiell regulatorisches RNA-Motiv ein hohes Maß an Sequenzkonservierung aufweist und sich damit deutlich vom Rest der UTR unterscheidet, der

möglichst keine Sequenzkonservierung mehr haben sollte. Da bei regulatorischen RNA-Motiven sowohl in UTRs als auch in anderen funktionalen RNA Molekülen häufig die Struktur der eigentliche funktionelle Faktor ist, läßt sich die zweite in dieser Arbeit entwickelte Vorhersagestrategie sehr viel universeller einsetzen. Wie in den folgenden Abschnitten diskutiert wird, bietet diese auf dem Strukturalignment basierende zweite Strategie auch im Vergleich mit anderen publizierten Strukturmotivvorhersagestrategien entscheidende Vorteile.

### 5.1.2. RNA-Strukturalignment als universelle Methode zur Strukturmotivvorhersage

Der Einsatz des RNA-Strukturalignmentprogramms *RNAforester* ermöglichte die Entwicklung einer extrem flexiblen Strategie zur Strukturmotivvorhersage. So konnten mit dieser Strategie regulatorische Motive gefunden werden, die an völlig verschiedenen Positionen in UTRs lokalisiert sind. Hierfür war die Berechnung eines paarweisen, lokalen Strukturalignments ausschlaggebend, wie das Beispiel der Vorhersage des *Iron Responsive Elements* in der Ferritin 5'UTR des Menschen und der Succinatdehydrogenase 5'UTR von *Drosophila* in Abschnitt 4.2.2 zeigt. Obwohl die Berechnung des Strukturalignments aufgrund der verwendeten Scorewerte die Primärsequenz nicht berücksichtigte, zeigte die Visualisierung dieses lokalen Alignments viele kompensatorische Basenpaare in der Helixregion des IRE (s. Abbildung 4.12). Dies ermöglicht die Bewertung der biologischen Relevanz des gefundenen Motives, da kompensatorische Basenaustausche ein starker Hinweis auf das tatsächliche Vorhandensein einer Struktur sind. Dies wird dadurch bestätigt, daß die verlässlichste Methode zur Sekundärstrukturvorhersage auf der Analyse solcher kompensatorischer Basenaustausche beruht (s. Abschnitt 2.5.1). Die Art der Visualisierung des Strukturalignments mit *RNAforester* ermöglicht somit neben der Vorhersage eines Strukturmotives gleichzeitig eine Beurteilung seiner biologischen Relevanz, obwohl kompensatorische Basenaustausche bei der Berechnung der Strukturalignments nicht berücksichtigt wurden.

Der Einsatz des Testsystems mit den modifizierten IREs zur Strategieentwicklung zeigte, daß die Berechnung eines paarweisen lokalen Strukturalignments auch zur Vorhersage von Motiven mit variabler Länge in einzelnen Strukturelementen geeignet ist. In dem verwendeten Beispiel konnten IREs mit variabler Länge im Loop gefunden werden (Abbildung 4.16). Wie weiter unten diskutiert wird, galt das bis dahin als einer der exklusiven Vorteile des *FOLDALIGN* Ansatzes von Gorodkin und Stormo. Eine weitere Stärke des lokalen Strukturalignments liegt in der Unabhängigkeit von der Größe der als Eingabe verwendeten Strukturen. Wie die Abbildung 4.14 zeigt, konnte das IRE auch als ähnlichstes lokales Strukturmotiv in der Ferritin 5'UTR (208 Nukleotide) und der sehr viel größeren Transferrinrezeptor 3'UTR (2464 Nukleotide) gefunden werden. Die Vorhersage regulatorischer Struktur motive in solch großen Strukturen wurde durch die im Rahmen dieser

## 5. Diskussion

Arbeit entwickelte Strategie des lokalen Strukturalignments zur Motivvorhersage erstmalig möglich und stellt somit einen enormen Vorteil gegenüber den bisher publizierten Vorhersagestrategien dar (s.u.).

Neben dem paarweisen lokalen Strukturalignment wurde auch das multiple globale Strukturalignment erfolgreich zur Motivvorhersage eingesetzt. Die Berechnung multipler Strukturalignments ist deshalb von großem Nutzen, da es sich bei den eingesetzten Strukturen in der Regel um solche handelt, die mit Hilfe von Energieminimierung vorhergesagt worden sind. Daß es sich bei der energetisch günstigsten Struktur nicht um die biologisch korrekte handeln muß, ist schon lange bekannt und konnte durch die Berechnung der Ferritin 5'UTR des Schweins im Rahmen dieser Arbeit bestätigt werden (s. Abschnitt 4.2.1). In der energetisch günstigsten Struktur dieser UTR ist das IRE nicht enthalten. Somit ist es erforderlich, neben der energetisch günstigsten Struktur auch energetisch suboptimale Strukturen zu analysieren. Während das beim paarweisen Strukturalignment nur nacheinander durchgeführt werden kann, ermöglicht das multiple Alignment die gleichzeitige Analyse vieler verschiedener suboptimaler Strukturen. Dadurch wird die Wahrscheinlichkeit, ein regulatorisches Motiv tatsächlich zu finden, deutlich erhöht. Die thermodynamischen Parameter, die zur Berechnung von RNA-Sekundärstrukturen verwendet werden, sind inzwischen soweit optimiert worden, daß im Durchschnitt 73% der tatsächlich vorhandenen Basenpaare in der energetisch optimalen Struktur korrekt vorhergesagt werden (Mathews et al., 1999). Obwohl die Wahrscheinlichkeit, daß ein kleines, stabiles Strukturmotiv bereits in der energetisch günstigsten Struktur enthalten ist, somit durchaus gut ist, erhöht die Betrachtung energetisch suboptimaler Strukturen die Chancen zur korrekten Vorhersage eines regulatorischen Motives. Daher sollten energetisch suboptimale Strukturen immer berücksichtigt werden, was durch den Einsatz des multiplen Strukturalignments erheblich vereinfacht wird.

Eine besondere Einsatzmöglichkeit des paarweisen und multiplen globalen RNA-Strukturalignments zeigte sich bei der Untersuchung der beiden Varianten der BMP7 3'UTR (s. Abschnitt 4.4.2). Die Berechnung multipler Strukturalignments mehrerer energetisch suboptimaler Strukturen der jeweiligen Variante zeigte die Ausbildung eines verzweigten Stemloops in der zweiten Variante, der in der ersten nicht auftrat. Das paarweise globale Alignment der energetisch günstigsten Strukturen der beiden Varianten (s. Abbildung 4.34) hebt diesen Unterschied deutlich hervor. Somit läßt sich das Strukturalignment mit **RNAforester** aufgrund der Visualisierungseigenschaften auch hervorragend dazu einsetzen, kleine Unterschiede in Gesamtstrukturen zu detektieren. Wie bei der ausführlichen Diskussion der Ergebnisse für die Untersuchung der BMP7 3'UTR in Abschnitt 5.3 beschrieben wird, ist bei alternativen Spleißvarianten gerade die Entdeckung solcher Unterschiede entscheidend. Während Ansätze zur Motivvorhersage normalerweise darauf ausgelegt sind, gemeinsame Motive in verschiedenen Strukturen zu entdecken (s.u.), bietet die Strategie des Strukturalignments mit **RNAforester**

auf einfache Weise die Möglichkeit, Unterschiede in ansonsten ähnlichen Strukturen zu finden. Damit eignet sich diese Strategie ausgezeichnet zur Untersuchung alternativ gespleißter UTRs.

### 5.1.3. Vergleiche mit anderen Vorhersagemethoden

In den folgenden Abschnitten werden die im Rahmen dieser Arbeit entwickelten Vorhersagestrategien zur Analyse regulatorischer RNA Struktur motive mit anderen publizierten Strategien verglichen.

#### Assoziierung von Sequenzmotiven mit funktionellen Genklassen

In einer von Darrell Conclin und Kollegen veröffentlichten Arbeit (Conclin et al., 2002) wurde in humanen 3'UTRs nach überrepräsentierten Sequenzmotiven gesucht. Dabei wurden die 3'UTRs zunächst in funktionelle Gruppen eingeteilt, in denen anschließend mit einer neu entwickelten, sogenannten "association rule mining" Methode, nach kurzen überrepräsentierten Nukleotidmustern gesucht wurde. Diese Muster variierten in der Größe von 5 bis 12 Nukleotiden. Für Pentamere und Hexamere konnten keine signifikanten Muster gefunden werden, während für Heptamere und Oktamere bestimmte GC-reiche Muster in Genen gefunden wurden, die für die Transkriptionskontrolle verantwortlich sind. Für die Oligonukleotide von 9 bis 12 Nukleotiden Länge wurden Sequenzmuster gefunden, die den AU-reichen Elementen entsprechen, welche eine Rolle bei der Destabilisierung einer mRNA spielen (s. Abschnitt 2.4). Diese sogenannten "AREs" wurden mit den 3'UTRs von Cytokinen assoziiert, wo ihre Bedeutung für die posttranskriptionelle Genregulation bereits bekannt war (Chen und Shyu, 1995). Während die Autoren das Auffinden der AU-reichen Elemente als Bestätigung für ihre Methode diskutierten, konnte über die GC-reichen Elemente, die stark mit UTRs assoziiert waren deren mRNAs für Proteine kodieren, welche eine sogenannte Homeodomäne besitzen, nur spekuliert werden. Hier wurde eine ähnliche Funktion wie bei den AU-reichen Elementen diskutiert, allerdings mit dem Hinweis, daß experimentelle Nachweise erforderlich wären. Aufgrund der Beschränkung auf Sequenzmotive kann dieser Ansatz für sich alleine genommen nur begrenzt zur Analyse regulatorischer Motive in UTRs eingesetzt werden, da diese wie bereits häufig erwähnt meist eine Kombination aus Sequenz- und Strukturanteilen sind. Durch eine Kombination mit der von mir entwickelten Strategie des Einsatzes eines Strukturalignments ließe sich der Ansatz jedoch funktionell erweitern. Hierzu wäre es zum Beispiel möglich, den Einfluß bestimmter kurzer, überrepräsentierter Sequenzabschnitte auf die Struktur der entsprechenden UTRs zu untersuchen, ähnlich zu dem Vorgehen bei den Spleißvarianten der BMP7 3'UTRs (Abschnitt 4.4.2). Die Kernidee des Ansatzes von Conclin war die Assoziierung bestimmter Nukleotidmuster mit funktionellen Genklassen. Die Untersuchung der Assoziierung bestimmter Strukturmuster mit funktionellen Genklassen könnte eine sehr

## 5. Diskussion

interessante Erweiterung des Ansatzes sein. Allerdings müßte die Signifikanzüberprüfung von Sequenzmustern auf Strukturmuster übertragen werden, wobei der von mir verwendete Ansatz von Meyer und Giegerich (Abschnitt 4.3.1) ein möglicher Ausgangspunkt für solch eine Signifikanzberechnung wäre. Die Erweiterung auf andere Organismen (nicht nur humane 3'UTRs) würde einen zusätzlichen Informationsgewinn bringen und den Ansatz variabler machen.

### **"Genetische Programmierung" zur RNA-Strukturvorhersage**

Im Jahr 2002 wurde ein sogenannter "Genetischer Programmierungsansatz" zum Auffinden lokaler Strukturmotive in coregulierten RNA Sequenzen publiziert (Hu, 2002). Der Ansatz wurde als GPRM (Genetic Programming for RNA Motifs) bezeichnet und kann als Optimierungsverfahren beschrieben werden, in dem bestimmte "genetische Operatoren" iterativ angewendet werden, um die "Fitness" vorläufiger Lösungen zu optimieren. GPRM besteht aus drei Bestandteilen, einer Population möglicher Strukturmotive, einer Fitnessfunktion zur Qualitätsmessung jedes einzelnen Motivs und genetischen Operatoren zur Veränderung der Motive. Motive werden dabei als Segmente beschrieben, wobei ein Segment entweder einzelsträngig oder ein Watson-Crick Komplement sein kann. Die Segmentanzahl sowie der Längenbereich einzelner Segmente muß vom Benutzer angegeben werden. Die Fitnessfunktion bewertet die Qualität einzelner Individuen (Motive) in der Population. Dabei werden solchen Motiven höhere Fitnesswerte zugeordnet, die häufig in den vom Benutzer vorgegebenen RNA Sequenzen und selten in Zufallssequenzen vorkommen. Durch die genetischen Operatoren werden zum Beispiel die Längen von Helices und einzelsträngigen Bereichen geändert. Das GPRM Verfahren wiederholt kontinuierlich die drei Phasen Fitnessbewertung, Individuumsauswahl und Populationsbildung, bis keine Verbesserung bei den Fitnesswerten mehr erzielt werden kann, oder die maximale Anzahl von Generationen erreicht ist. Die Funktionalität des Verfahrens wurde unter anderem mit den modifizierten IRE Sequenzen überprüft, die ursprünglich für die Evaluation des FOLDALIGN Ansatzes (s.u.) generiert wurden, und die auch für die Entwicklung der Vorhersagestrategien in dieser Arbeit verwendet wurden (Abschnitt 4.1.1). Für diese Sequenzen konnte das IRE als gemeinsames Strukturmotiv durch den GPRM Ansatz gefunden werden. Dazu mußten jedoch zuvor Angaben über die maximale Anzahl möglicher Helices und einzelsträngiger Bereiche sowie deren Größenbereich gemacht werden. Solche vorab zu definierenden Bedingungen sind bei den von mir entwickelten Strategien nicht erforderlich. Dem Autor zufolge kann die Zeitkomplexität des Ansatzes unter günstigen Bedingungen auf  $O(L^3N)$  reduziert werden, wobei L die maximale Sequenzlänge und N die Anzahl der Sequenzen ist. Obwohl diese Zeitkomplexität günstiger ist als bei dem FOLDALIGN Ansatz (s.u.), können dennoch nur kürzere Sequenzen effektiv analysiert werden.

### Optimierung von Sequenz und Struktur: der FOLDALIGN Ansatz

Ein sehr erfolgversprechendes Verfahren zur Vorhersage regulatorischer Motive in UTRs wurde 1997 von Jan Gorodkin und Kollegen veröffentlicht (Gorodkin et al., 1997). In dem in Abschnitt 2.6 erwähnten FOLDALIGN Ansatz wird gleichzeitig ein Sequenzalignment und die Struktur der Sequenzen optimiert. Wie bereits erwähnt wurde, stellt der Ansatz eine Vereinfachung des von Sankoff (Sankoff, 1985) vorgeschlagenen Algorithmus zum optimalen Falten und Alignieren mehrerer RNA Sequenzen dar. Obwohl die Autoren ihr Verfahren als strukturelles Alignment bezeichnen, darf es nicht mit dem in dieser Arbeit verwendeten Programm `RNAforester` verwechselt werden. Bei FOLDALIGN wird von Sequenzen ausgegangen, die so aligniert werden, daß gleichzeitig überprüft wird, ob Basen an bestimmten Positionen in zwei Sequenzen auch Basenpaarungen bilden können. Hier werden also nicht wirklich Strukturen aligniert, im Gegensatz zu `RNAforester`, das Strukturen als Eingabe erhält und ein tatsächliches, auf verschiedenen Edit-Operationen aufgebautes Strukturalignment berechnet. Ein Vorteil von FOLDALIGN gegenüber der Strategie des Strukturalignments mit `RNAforester` könnte darin liegen, daß `RNAforester` auf die Qualität der erforderlichen RNA-Sekundärstrukturvorhersagen angewiesen ist, während FOLDALIGN seine Optimierungen ausgehend von den Primärsequenzen durchführt. Die Vereinfachung des FOLDALIGN Ansatzes, nur die Anzahl von Basenpaarungen zu maximieren und die Energien eines damit erhaltenen Motives nicht zu berücksichtigen, kann jedoch dazu führen, daß Motive vorhergesagt werden, die an sich nicht stabil sind. Desweiteren berücksichtigt FOLDALIGN die Energie der Gesamtstruktur nicht, so daß ein vorhergesagtes Motiv zwar für sich genommen stabil sein kann, aber die Energie der Gesamtstruktur der RNA so stark beeinflußt, daß es in der Natur nicht vorkommen kann. Das Strukturalignment mit `RNAforester` berücksichtigt diese Energie der Gesamtstruktur, sofern die alignierten Strukturen zuvor durch Energieminimierung vorhergesagt wurden. Aufgrund der immer besser werdenden Qualität dieser RNA-Sekundärstrukturvorhersagen bietet die Strategie der Motivvorhersage mit `RNAforester` hier Vorteile. Auch die Betrachtung energetisch suboptimaler Sekundärstrukturen kann diesen vermeintlichen Nachteil gegenüber FOLDALIGN ausgleichen. Die Zeitkomplexität, die für die Analyse von  $N$  Sequenzen mit einer Maximallänge  $L$  bei FOLDALIGN  $O(L^4 N^4)$  beträgt, führt dazu, daß Sequenzen bis maximal 300 Nukleotiden Länge von dem FOLDALIGN-Server akzeptiert werden. Das Auffinden des IRE als ähnlichstes Motiv in der Ferritin 5'UTR und der Transferrinrezeptor 3'UTR ist mit diesem Ansatz somit nicht möglich, während das paarweise lokale Strukturalignment mit einer Zeitkomplexität von  $O(L^2 k^2)$  hierzu in der Lage war (Abschnitt 4.2.2), wobei  $L$  die Sequenzlänge und  $k$  die maximale Loopgröße der Struktur ist. Um die angegebene Zeitkomplexität zu erreichen, wurde bei FOLDALIGN eine Restriktion auf einfache Stemloop-Strukturen eingeführt. Mit der Strategie des Strukturalignments ist dagegen auch eine Vorhersage verzweigter Struktur motive möglich.

## 5.2. Datenbanksuche zur Analyse regulatorischer RNA-Motive

Die Vorhersage potentiell regulatorischer RNA-Struktur motive ist der erste Schritt zur Analyse posttranskriptioneller Regulationsmechanismen. Um die Qualität der Vorhersagen zu überprüfen, muß eine Validierung der Vorhersagen erfolgen. Im Rahmen dieser Arbeit wurden sowohl computergestützte als auch experimentelle Methoden zur Validierung von Strukturmotivvorhersagen durchgeführt. Die bioinformatische Validierung erfolgte durch die Motivsuche in Datenbanken. Hierzu mußte zunächst eine Strategie zur Definition eines Suchmusters entwickelt werden. Die Anforderungen an ein geeignetes Suchmuster umfassen dabei zwei gegensätzliche Eigenschaften des Musters. Zum einen muß es allgemein genug beschrieben sein, um möglichst viele Vorkommen des Musters durch die Datenbanksuche zu entdecken. Zum anderen muß es aber auch spezifisch genug sein, um möglichst wenige falsch positive Treffer zu liefern. Daher wurde zur Entwicklung des Suchmusters das in Abschnitt 4.3.1 beschriebene Verfahren der Musterdefinition durch Signifikanzberechnungen angewendet. Mit Hilfe dieser Methode wurden die Muster so beschrieben, daß höchstens ein Treffer in einer Zufallssequenz zu erwarten wäre. Dies stellt einen guten Kompromiß zwischen Spezifität und Variabilität des Suchmusters dar. Die tatsächlichen Datenbanksuchen, sowohl für das *Iron Responsive Element* als auch für den vorhergesagten Stemloop in der Rab1A 3'UTR, lieferten sehr viel mehr Treffer, als per Zufall zu erwarten gewesen wären. Während die biologische Funktion für die meisten Treffer der IRE Suche bereits bekannt war, wurde eine mögliche biologische Bedeutung des Rab1A Stemloops im Rahmen dieser Arbeit experimentell überprüft. Beide Beispiele verdeutlichen jedoch die Funktionalität der Signifikanzberechnung zur Musterdefinition, da die Tatsache, daß deutlich mehr Treffer gefunden wurden, als per Zufall zu erwarten waren, ein wichtiger Hinweis auf eine tatsächliche biologische Funktion ist. Wie weiter unten diskutiert wird, ermöglichten die Ergebnisse der Datenbanksuche für den vorhergesagten Stemloop in der Rab1A 3'UTR eine umfassende Deutung der experimentellen Ergebnisse bei der Proteinidentifizierung. Die Datenbanksuche ist somit der entscheidende bioinformatische Ansatz zur Validierung vorhergesagter Motive, und die Musterdefinition durch Signifikanzberechnungen erleichtert die Bewertung der gefundenen Treffer, besonders in Kombination mit experimentellen Ergebnissen.

### 5.2.1. RNAMotif und HyPa im Vergleich

Die Datenbanksuchen wurden in dieser Arbeit auch dazu genutzt, um das von Thomas Macke und Kollegen entwickelte Motivsuchprogramm *RNAMotif* mit dem Programm *HyPa*, das zeitgleich zu meiner Arbeit in der Doktorarbeit von Dirk Strothmann entwickelt wurde, zu vergleichen.



Während die Suche nach reinen Sequenzmotiven bereits durch einfache UNIX Programme wie `AGREP` zufriedenstellend, wenn auch nicht optimal durchgeführt werden kann, stellt die Suche nach RNA-Motiven besondere Anforderungen an die Programme. Obwohl bestimmte Programme das Problem der Suche nach speziellen funktionalen RNAs, wie zum Beispiel tRNAs, sehr gut lösen (Lowe und Eddy, 1997), sind für die Datenbanksuche nach neu vorhergesagten Strukturmotiven Programme erforderlich, die eine einfache und schnelle Beschreibung auch komplexerer Struktur motive ermöglichen. Sowohl `HyPa` als auch `RNAMotif` entsprechen diesen Anforderungen, unterscheiden sich allerdings teilweise durch ihre Funktionalität und, wie Experimente im Rahmen dieser Arbeit zeigen konnten, vor allem durch die benötigte Zeit zur Mustersuche. Beiden Programmen ist eine einfache Art der Musterbeschreibung gemeinsam, wie der Abschnitt 4.3.2 verdeutlicht. `HyPa` ist dabei in der Art der Beschreibung einzelner Elemente des Musters variabler, da zum Beispiel eine bestimmte Anzahl von Insertionen und Deletionen in einer Helixregion definiert werden können. Dadurch ist es möglich, Bulge loops zu beschreiben, deren genaue Position in einer Helix man nicht mit angeben möchte. Um mit `RNAMotif` Bulge loops zu beschreiben, muß die Helix, in der der Bulge enthalten sein soll, in zwei Helices aufgespalten werden. Dadurch wird die Position des Bulges zumindest teilweise festgelegt (variable Längen der Helices lassen ein gewisses Maß an Variabilität zu). Im Rahmen dieser Arbeit war die Definition solcher Bulge loops jedoch nicht erforderlich, da sowohl im IRE Muster als auch im Rab1A Muster die Position und Sequenz des Bulge loops genau vorgegeben war. Die Analyse sogenannter Stemloop-Precursor von micro-RNAs (Lee et al., 2002), deren Entdeckung durch verschiedene bioinformatische Ansätze zurzeit ein sehr aktuelles Forschungsthema ist (Lim et al., 2003), zeigt jedoch, daß sich diese Strukturen durch eine Vielzahl von Bulge loops auszeichnen, deren Position in den Precursoren verschiedener micro-RNAs sehr unterschiedlich sein kann. Für die Beschreibung eines gemeinsamen Suchmusters für die Stemloop-Precursor verschiedener micro-RNAs wäre `HyPa` aufgrund der erwähnten Insertions- und Deletionsoptionen somit sehr viel besser geeignet als `RNAMotif`.

Die Vergleiche der zur Mustersuche benötigten Zeiten für das Rab1A Muster und vor allem für das BMP7 Muster ergaben deutliche Vorteile für `HyPa`. Für das Rab1A Muster waren die ermittelten Suchzeiten für `HyPa` zu klein, um gemessen werden zu können, während `RNAMotif` für die Suche nach diesem einfachen Muster in der größten 3'UTR Kollektion (Human) bereits 2,8 Sekunden benötigte. Das in dieser Arbeit definierte BMP7 Muster stellt generell sehr viel höhere Anforderungen an ein Suchprogramm als einfache Stemloop-Strukturen, wie das IRE oder der Rab1A Stemloop. Das verzweigte BMP7 Muster (Abbildung 4.35) enthält fast keine Sequenzinformationen und läßt darüberhinaus variable Helixlängen zu. Während `RNAMotif` für die Suche nach dem einfachen Rab1A Stemloop in der 3'UTR Kollektion des Menschen 2,8 Sekunden benötigte, dauerte die Suche nach dem BMP7 Muster in dem gleichen Datensatz 88,4 Sekunden. `HyPa` benötigte

dagegen nur 11,2 Sekunden für die gleiche Suche (s. Tabelle 4.7). Dieser Vorteil von HyPa resultiert aus der verwendeten Indexstruktur. HyPa führt die Suche mit Hilfe eines sogenannten Affix-Arrays durch. Die Verwendung dieser Indexstruktur ermöglicht es, daß die Suche nicht mit dem ersten Teilmotiv beginnen muß, sondern mit einem beliebigen Abschnitt des Musters beginnen kann. So wird bei der Suche nach dem IRE zum Beispiel zunächst nach der festgelegten Loopsequenz gesucht, um anschließend zu überprüfen, ob die umgebenden Sequenzen den geforderten Stemloop ausbilden können. Dazu ist es somit auch nicht erforderlich, daß jede Position der zu durchsuchenden Sequenz betrachtet wird. Dies führt zu einem enormen Zeitvorteil bei der Suche nach kombinierten Sequenz und Strukturmotiven.

### 5.3. Alternativ gespleißte UTRs: die BMP7 3'UTR

Die neuesten Bioinformatikanalysen haben ergeben, daß 40-60% aller menschlichen Gene alternative Spleißformen haben (Modrek und Lee, 2002). Die Methoden beruhen hauptsächlich darauf, daß EST-Sequenzen an genomischen Sequenzen ausgerichtet werden, um dann Konsensussequenzen für die verschiedenen Spleißformen zu bilden (Kan et al., 2001). Um nicht auf das Vorhandensein genomischer Daten angewiesen zu sein, entwickelt Alexander Sczyrba in seiner laufenden Doktorarbeit ein Verfahren, daß es ermöglicht, alternative Spleißformen ausschließlich durch EST-Clusterung zu finden. Dabei wurden die beiden potentiellen Spleißvarianten der BMP7 3'UTR bei *Xenopus laevis* entdeckt, die in dieser Arbeit näher analysiert wurden (Abschnitt 4.4.2). Während in einer früheren Veröffentlichung davon ausgegangen wurde, daß 80% aller alternativen Spleißvorgänge die 5'UTRs der entsprechenden Gene betreffen (Mironov et al., 1999), zeigen neuere Analysen, daß 74% die codierende Region, 22% die 5'UTRs und nur 4% die 3'UTRs betreffen (Modrek et al., 2001). Die Tatsache, daß der Anteil der 3'UTRs so gering ist, wird unter anderem dadurch erklärt, daß alternativ gespleißte 3'UTRs destabilisierende Elemente enthalten könnten. Dies würde dazu führen, daß diese Spleißformen seltener beobachtet werden (Modrek et al., 2001).

Es ist sehr wahrscheinlich, daß mRNAs, die alternativ gespleißte UTRs enthalten, auf unterschiedliche Weise posttranskriptionell reguliert werden. Das Einfügen eines Destabilisierungselementes in eine 3'UTR durch alternatives Spleißen kann beispielsweise gerade dazu führen, daß eine mRNA in einem bestimmten Zelltyp oder in einem speziellen Entwicklungsstadium gezielt abgebaut wird. Genauso ist es denkbar, daß alternativ gespleißte 5'UTRs unterschiedliche Struktur motive enthalten, die Auswirkungen auf die Translationseffizienz der mRNA haben könnten.

Die Analyse der beiden BMP7 3'UTR Varianten verdeutlicht, wie kleine Unterschiede in der Sequenz zu einem neuen, energetisch sehr stabilen Strukturmotiv

führen können. Durch die Berechnung von Strukturalignments mit `RNAforester` wird die Untersuchung alternativer Spleißvarianten in UTRs verbessert. Das paarweise Alignment der beiden BMP7 3'UTR Varianten (Abbildung 4.34) zeigt sofort den neu eingefügten Stemloop in der zweiten Variante. Auch die Berechnung multipler Strukturalignments hat die verzweigte Stemloopstruktur in der zweiten Variante hervorgehoben. Durch die anschließende Analyse der Basenpaarwahrscheinlichkeiten mit Hilfe der Dotplots konnte gezeigt werden, daß dieses Strukturmotiv stabil und auf die zweite Variante beschränkt ist. Geht man davon aus, daß die Untersuchung alternativ gespleißter UTRs und ihre Rolle für die posttranskriptionelle Genregulation in Zukunft weiter an Bedeutung gewinnen wird, so stellt die vergleichende Strukturuntersuchung mit `RNAforester` einen enormen Vorteil für die Entdeckung regulatorischer Struktur motive in unterschiedlichen Spleißvarianten dar.

Um die mögliche biologische Relevanz der zweiten Spleißvariante der BMP7 3'UTR zu analysieren, wurde eine Datenbanksuche mit dem verzweigten Stemloop als Suchmuster durchgeführt. Das Muster wurde dabei sehr allgemein beschrieben, um möglichst viele Treffer zu finden. Durch die Datenbanksuche sollte überprüft werden, ob das Strukturmotiv in anderen BMP 3'UTRs und darüberhinaus in weiteren mRNAs vorkommt. Die Ergebnisse der Datenbanksuche zeigten, daß das Motiv in keiner weiteren BMP UTR enthalten war. Wenn man spekuliert, daß das gefundene Motiv einen Einfluß auf die posttranskriptionelle Genregulation der BMP7 mRNA hat und daß die zweite Variante deshalb nur in ganz bestimmten Zelltypen oder zu ganz bestimmten Zeitpunkten gebildet wird, wäre es ebenfalls möglich, daß auch andere BMP mRNAs selten exprimierte Spleißformen haben, die dieses Motiv enthalten könnten, und die nicht in der UTR Datenbank enthalten sind. Dies zeigt sich auch für die BMP7 3'UTR, deren in der UTR-Datenbank vorhandene Variante das Motiv ebenfalls nicht enthält.

BMP7 spielt unter anderem eine Rolle bei der Induzierung der Epidermis bei dem Krallenfrosch *Xenopus laevis* (Suzuki et al., 1997). Neben BMP7 sind auch BMP4 und BMP2 an dieser Induzierung beteiligt. Somit wäre es interessant, auch bei BMP2 und BMP4 nach alternativen Spleißformen zu suchen, die möglicherweise ebenfalls das in der BMP7 3'UTR gefundene Strukturmotiv enthalten. Dazu müßten jedoch gezielt Experimente durchgeführt werden, da die in den Datenbanken vorhandenen BMP2 und BMP4 UTR Sequenzen das Motiv nicht enthalten.

Bei der Datenbanksuche wurde das Motiv in fünf 3'UTRs gefunden. Darüber, ob diese UTRs das Motiv auch tatsächlich ausbilden, kann derzeit nur spekuliert werden. Auffällig war jedoch, daß in der größten 3'UTR Kollektion, der humanen, keine Treffer gefunden wurden. Zufällige Treffer hätte man am ehesten in dieser großen Kollektion erwartet. Auch in den gesamten 5'UTR Kollektionen konnte kein Treffer gefunden werden, so daß das Motiv auf 3'UTRs beschränkt zu sein scheint. Das macht eine biologische Funktion zumindest möglich. Interessant ist dabei der Treffer in der Xepsin 3'UTR von *Xenopus*. Bei Xepsin handelt es sich um eine Serinprotease, die spezifisch in der Epidermis von *Xenopus* exprimiert

## 5. Diskussion

wird (Yamada et al., 1999). Ein direkter funktioneller Zusammenhang zwischen Xepsin und BMP7 kann allerdings nicht abgeleitet werden.

Die Entdeckung der zweiten Variante der BMP7 3'UTR und des darin enthaltenen verzweigten Stemloops bietet den Ausgangspunkt für eine Reihe experimenteller Ansätze zur Überprüfung einer möglicherweise vorliegenden biologischen Bedeutung. So ließe sich zum Beispiel mit PCR-Experimenten überprüfen, in welchen Zellen und in welchen Entwicklungsstadien von *Xenopus* die zweite Variante tatsächlich vorkommt. Der Einsatz von Reportergenkonstrukten könnte zeigen, ob die beiden UTRs unterschiedliche Einflüsse auf die Stabilität oder die Translationseffizienz der Reporter-gen-mRNA haben. Durch chemisches Mapping könnte darüberhinaus in der Xepsin 3'UTR das tatsächliche Vorkommen des verzweigten Stemloops nachgewiesen werden.

Das Beispiel der BMP7 3'UTR zeigt somit, wie bioinformatische Analysen der Ausgangspunkt für vielfältige biologische Experimente sein können. Die in den folgenden Abschnitten diskutierten bioinformatischen und experimentellen Analysen der Rab1A 3'UTR betonen dieses Zusammenspiel noch deutlicher.

### 5.4. Die Rab1A 3'UTR

Die entwickelten bioinformatischen Methoden wurden zur Analyse der Rab1A 3'UTR eingesetzt. Diese UTR eignete sich vor allem aus zwei Gründen. Zum einen konnte in früheren Arbeiten eine hohe Konservierung der Rab1A 3'UTR in zahlreichen Vertebraten nachgewiesen werden (Wedemeyer et al., 2000). Dies machte eine biologische Funktion dieser UTR bzw. eines Motives innerhalb der UTR wahrscheinlich. Außerdem handelt es sich bei Rab1A um ein ubiquitäres Protein, was die experimentellen Analysen, die in dieser Arbeit durchgeführt wurden, erheblich vereinfachte.

#### 5.4.1. Die Vorhersage eines potentiell regulatorischen Stemloops

Entscheidend für die Vorhersage des potentiell regulatorischen Stemloops in der Rab1A 3'UTR war der Einsatz der UTR Sequenz des Zitterrochen. Wie die Abbildung 4.21 zeigt, waren die 3'UTRs, die in der Arbeit von Wedemeyer et al. untersucht wurden, zu ähnlich, um ein einzelnes Strukturmotiv vorherzusagen. Die 3'UTR des Zitterrochen war im Vergleich zu den übrigen Rab1A 3'UTRs sehr viel weniger konserviert. So konnte durch beide Vorhersagestrategien der Stemloop (s. Abbildung 4.22) als gemeinsames Strukturmotiv der Rab1A 3'UTR des Zitterrochen und des Menschen vorhergesagt werden. Dieser Stemloop ist in beiden Spezies über einen weiten Temperaturbereich (von 10°C bis 37°C) stabil. Da Fische in der Regel eine deutlich niedrigere Körpertemperatur als Menschen haben, ist das ein wichtiger Hinweis auf die tatsächliche biologische Bedeutung

des Stemloops. Auch die Analyse der Basenpaarwahrscheinlichkeiten zeigte, daß dieses Strukturmotiv mit großer Wahrscheinlichkeit in der Natur existiert. In der Abbildung 5.1 ist der Stemloop als Resultat des paarweisen lokalen Strukturalignments der Rab1A 3'UTRs von Mensch und Zitterrochen noch einmal dargestellt. Die Sequenz innerhalb des Stemloops ist zwischen Mensch und Zit-

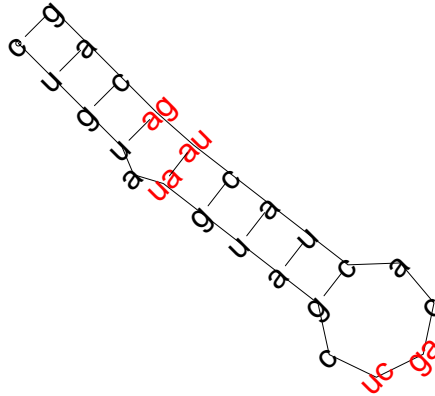


Abbildung 5.1.: Vorhergesagter Stemloop in der Rab1A 3'UTR. Die rot gezeichneten Basen zeigen Sequenzunterschiede zwischen Mensch und Zitterrochen.

terrochen hoch konserviert. Der kompensatorische Basenaustausch nach dem Bulge-loop sowie der Austausch des Adenins durch ein Guanin vor dem Bulge-loop sind deutliche Hinweise darauf, daß die Struktur in diesem Bereich erhalten bleiben mußte, was wiederum für eine biologische Funktion des Stemloops spricht. Innerhalb des Loops ist an der zweiten Position ein Pyrimidin durch ein Pyrimidin (U gegen C) und an der dritten Position ein Purin gegen ein Purin (G gegen A) ausgetauscht worden. Diese Austausche wurden bei der Definition des Suchmusters entsprechend berücksichtigt.

### Das Rab1A Stemloop-Suchmuster und die Datenbanksuche

Ausgehend von der Motivvorhersage wurde ein möglichst allgemeines Suchmuster für den Rab1A Stemloop definiert. Die zu erwartende Trefferanzahl von 0,81 in einer entsprechenden Zufallssequenz wurde dabei als hinreichend signifikant eingestuft. Obwohl der Stemloop in den Helixbereichen zwischen Mensch und Zitterrochen hoch konserviert war, wurde die Sequenz mit Ausnahme des schließenden Basenpaares in diesen Bereichen variabel gehalten. Dies resultierte daraus, daß bereits vor den experimentellen Untersuchungen anzunehmen war, daß es eine spezifische Proteinbindung an den Stemloop geben müßte, falls er tatsächlich eine

## 5. Diskussion

regulatorische Funktion haben sollte. Für solch eine spezifische Proteinbindung sind neben dem eigentlichen Strukturmotiv einzelsträngige Sequenzen wichtig, da sie normalerweise von Proteinen für eine spezifische Bindung genutzt werden. Daher wurden bei der Musterdefinition die Basen im Loop vorgegeben, wobei auf die Unterschiede an Position zwei und drei geachtet wurde. Der Bulge-Loop wurde sowohl als Sequenz- und auch als Strukturmerkmal definiert. Bulge-Loops, die aus einer Base bestehen, führen zu einem Abknicken der Helix, wobei die große Furche noch stärker erweitert wird, so daß sie für Bindeproteine besser zugänglich ist (Zacharias und Sklenar, 1999). Daher wurde der Adenin-Bulge als ein wichtiges Element des Musters definiert.

Die durchgeführte Datenbanksuche zeigte, daß das Vorkommen des vorhergesagten Stemloops auf sehr wenige verschiedene mRNA 3'UTRs beschränkt ist. Das und die Tatsache, daß das Muster in den 5'UTR Kollektionen der UTRdb nicht auftrat, ist ein starker Hinweis auf eine tatsächliche biologische Funktion. Neben zahlreichen Rab1A 3'UTRs wurde der Stemloop auch in der 3'UTR der Sir2 $\alpha$  mRNA der Maus sowie der dazu homologen Sirtuin 1 mRNA (diese wird alternativ auch als Sir2 $\alpha$  bezeichnet) des Menschen gefunden. Die Sequenz des Stemloops in der Sirtuin 1 und Sir2 $\alpha$  3'UTR ist bei Mensch und Maus identisch, unterscheidet sich aber von der Sequenz des Stemloops in der Rab1A 3'UTR. Die Abbildung 5.2 zeigt den Stemloop, wie er in den Sirtuinen zu finden ist. Desweiteren zeigt die Abbildung das Strukturalignment des humanen Sirtuin 1 und Rab1A Stemloops, um die Unterschiede auf der Sequenzebene zu verdeutlichen. Die Rolle des Stemloops bei den Sirtuinen, vor allem im Hinblick auf das experimentell gefundene Sirtuin 6 als "Bindeprotein" des Stemloops, wird weiter unten diskutiert.

Zu dem Stemloop in der Methyltransferase 3'UTR kann kein Zusammenhang hergestellt werden, so daß es sich hierbei möglicherweise um einen falsch positiven Treffer handelt.

### 5.4.2. Die biologische Bedeutung des Stemloops

Geht man davon aus, daß der vorhergesagte Stemloop in der Rab1A 3'UTR einen Einfluß auf die posttranskriptionelle Genregulation der Rab1A mRNA hat, so könnte die Stabilität, die Translationseffizienz, oder aber auch die Lokalisation der mRNA kontrolliert werden. Obwohl man grundsätzlich keinen der Mechanismen ausschließen kann, ist eine Lokalisation der Rab1A mRNA doch am wahrscheinlichsten. Es wäre allerdings auch möglich, daß die Translation an eine mögliche Lokalisation gekoppelt ist, wie es für verschiedene andere mRNAs bereits nachgewiesen werden konnte (Lipshitz und Smibert, 2000). Da das Rab1A Protein ein ubiquitäres Protein ist, das in allen Zelltypen für den Vesikeltransport vom Endoplasmatischen Retikulum zum Golgi-Apparat verantwortlich ist, ist es unwahrscheinlich, daß die Rab1A mRNA in Abhängigkeit von bestimmten inneren oder äußeren Einflüssen destabilisiert wird. Eine Lokalisation der mRNA

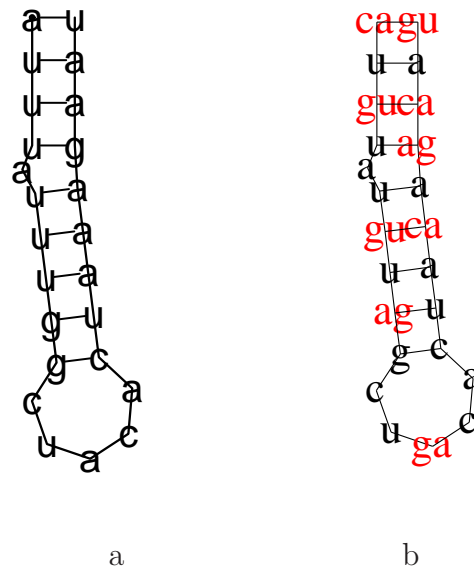


Abbildung 5.2.: Stemloop, wie er in der Sir2 $\alpha$  3'UTR von Mensch und Maus zu finden ist (a). Strukturalignment des Stemloops in der Rab1A und Sir2 $\alpha$  3'UTR des Menschen (b).

wäre hingegen denkbar, da auch das Rab1A Protein im Bereich des *cis*-Golgi-Netzwerkes lokalisiert ist (Saraste et al., 1995). Diese Lokalisation des Proteins könnte durch eine vorhergehende Lokalisation der mRNA verursacht werden.

Die Ergebnisse der im Rahmen dieser Arbeit durchgeführten *in vivo* Lokalisationsexperimente (Abschnitt 4.5.3) geben erste Hinweise darauf, daß die Rab1A mRNA tatsächlich nur an bestimmten Orten in der Zelle in größeren Mengen vorkommt. Dies muß jedoch durch weitere Experimente verifiziert werden (s. Ausblick). Die hier durchgeführten FRET-Experimente zeigen zwar eine Lokalisation der Rab1A mRNA, allerdings ist der genaue Ort der Lokalisation in der Zelle (z.B. die Nähe zum Golgi-Apparat) durch die Experimente nicht nachgewiesen worden. Wie später beschrieben wird, sind hierzu aufwändigere Co-Lokalisationsstudien erforderlich.

Ob der Stemloop als solcher einen Einfluß auf eine mögliche posttranskriptionelle Regulation der Rab1A mRNA hat, sollte durch die Gelverzögerungsexperimente sowie die Proteinisolierung nachgewiesen werden.

### Diskussion der Gelverzögerungsexperimente

Der vorhergesagte Stemloop in der Rab1A 3'UTR ist vom Aufbau her mit dem *Iron Responsive Element* vergleichbar. Es handelt sich um einen einfachen Stemloop mit einem Einzelnukleotid-Bulge. Wie oben bereits beschrieben wurde, ist gerade das Auftreten eines solchen Bulges ein deutlicher Hinweis auf eine Pro-

## 5. Diskussion

teininteraktion mit dem Stemloop, da er die Zugänglichkeit für Bindeproteine zur großen Furche des Doppelstrangbereiches erleichtert. Daher sollten die Gelverzögerungsexperimente genutzt werden, um eine spezifische Proteinbindung an den Stemloop nachzuweisen. Die in Abschnitt 4.5.1 dargestellten Ergebnisse zeigen, daß es zu solch einer spezifischen RNA-Protein-Interaktion kommt. Das Auffällige bei den Gelverzögerungen war, daß entweder keine (Abbildung 4.38) oder nur eine sehr schwache (Abbildung 4.37) Verzögerungsbande im Gel beobachtet werden konnte. Der größte Anteil der markierten RNA-Oligonukleotide war nach Proteinzugabe nicht mehr nachweisbar. Zum einen wäre es theoretisch möglich, daß die RNA-Oligonukleotide durch Zugabe des Proteinextraktes abgebaut wurden. Da die unverzögerte Bande nach Zugabe der unmarkierten Oligonukleotide jedoch wieder auftrat, ist solch ein Abbau eigentlich auszuschließen. Es ist sehr viel wahrscheinlicher, daß sehr große Proteinkomplexe mit dem Stemloop interagierten, so daß diese RNA-Protein-Komplexe nicht mehr ins Gel einwandern konnten. Dies wird auch durch die Ergebnisse der Proteinisolierungsexperimente bestätigt.

### **Diskussion der Proteinisolierungsexperimente**

Die Proteinisolierung mit den Dynabeads zeigte eine unerwartet hohe Anzahl von Proteinbanden. Obwohl davon ausgegangen werden muß, daß einige der isolierten Proteine durch unspezifische Bindung an das verwendete RNA-Oligonukleotid mit aufgereinigt wurden, so handelt es sich bei der Mehrzahl der isolierten Proteine wahrscheinlich um solche, die spezifisch direkt oder indirekt in Form von größeren Proteinkomplexen mit dem Stemloop interagiert haben. Dafür spricht, daß auch bei der Proteinisolierung ein Überschuß an Hefe-RNA eingesetzt wurde, um unspezifische Proteinbindungen soweit wie möglich zu verhindern. Aus experimentellen Gründen ist dieser Überschuß jedoch deutlich geringer als bei den Verzögerungsexperimenten, so daß ein geringer Anteil unspezifischer Bindungen nicht ausgeschlossen werden kann. Desweiteren ist anzunehmen, daß nur ein geringer Teil der spezifisch aufgereinigten Proteine direkt mit dem Stemloop interagiert. Ein Kandidat hierfür wäre das durch die MALDI-Analysen gefundene RNA-Bindeprotein. Da die Ergebnisse der Gelverzögerungsexperimente nahelegen, daß ein großer Proteinkomplex mit dem Stemloop interagiert, ist davon auszugehen, daß ein Großteil der Banden zu spezifischen Proteinen eines solchen Komplexes gehört. Die durchgeführten MALDI-Analysen liefern mögliche Hinweise auf das Vorhandensein eines solchen Komplexes.

### **Generelle Bewertung der MALDI-TOF Analysen**

Die Ergebnisse der MALDI-Analysen verschiedener Proteinbanden müssen als Ausgangspunkt für weitere Experimente angesehen werden. Insgesamt erwiesen sich die MALDI-Ergebnisse als schwer auswertbar. Einer der Gründe hierfür ist,



daß einzelne Proteinbanden in den SDS-Gelen anscheinend Mischungen verschiedener Proteine enthielten. Dies macht eine Auswertung mit der MASCOT-Software extrem schwierig, da die erhaltenen Treffer in der Regel schlechte Signifikanzwerte hatten, was auf das Vorhandensein von Proteingemischen zurückzuführen war. Den eindeutigsten Hinweis, daß ein isoliertes Protein durch die MALDI-Analyse korrekt identifiziert wurde, lieferte die nach dem denaturierenden Gel zu erwartende Größe des Proteins.

Überraschend war auch die Tatsache, daß die Analyse der Proteinbande mit einer Größe von 37 kDa bei verschiedenen Experimenten unterschiedliche Massenspektren und somit auch unterschiedliche Proteine lieferte (s. Abschnitt 4.5.2). Wie weiter unten diskutiert wird, sind sowohl das potentielle RNA-Bindeprotein als auch das Sirtuin 6 funktionell interessante Kandidaten. Man kann spekulieren, daß beide Proteine direkt an den Stemloop binden können (obwohl für das Sirtuin 6 keine RNA-Bindedomäne beschrieben ist, läßt zumindest der pI-Wert von 9,78 auf eine Nukleinsäurebindefähigkeit schließen) und somit in Konkurrenz zueinander stehen. Das könnte die verschiedenen Ergebnisse bei den MALDI-Analysen erklären. Es wäre jedoch auch möglich, daß beide Proteine in einem größeren Komplex assoziiert sind und somit beide mit dem Stemloop interagieren. In diesem Zusammenhang ist vor allem das Vorkommen des Stemloops in den 3'UTRs der Sirtuin 1 und Sir2 $\alpha$  mRNAs zu diskutieren.

Zunächst werden jedoch die Ergebnisse der Proteinidentifizierungen im Hinblick auf die mögliche posttranskriptionelle Genregulation der Rab1A mRNA diskutiert.

### **Funktionelle Bedeutung der identifizierten Proteine**

Für die Bande 3 aus Abbildung 4.39 konnte zum einen ein 41 kDa großes potentiell RNA-Bindeprotein und (in einem weiteren Experiment) das Sirtuin 6 identifiziert werden. Obwohl das Bindeprotein mit 41 kDa etwas zu groß ist, da Proteine in der analysierten Bande eine Größe von ungefähr 37 kDa haben müßten, ist es dennoch ein guter Kandidat für eine spezifische Interaktion mit dem Stemloop. Da von dem Protein nur die cDNA bekannt ist, könnte das Laufverhalten durch posttranslationale Modifikationen so beeinflusst werden, daß das Protein etwas schneller im Gel läuft. Über eine mögliche Funktion des Proteins kann allerdings nur spekuliert werden, da außer der cDNA Sequenz nichts über das Protein bekannt ist. Das Vorhandensein von potentiellen RNA-Bindedomänen kann ein Hinweis darauf sein, daß das Protein direkt an den Stemloop bindet und somit die Verbindung zu weiteren Proteinen in einem größeren Komplex darstellt.

Neben dem RNA-Bindeprotein und Sirtuin 6 zeigte auch das Ergebnis der Datenanalyse für die Bande 2 aus Abbildung 4.39 mit dem Keratin 7 eine gute Übereinstimmung bezüglich der zu erwartenden Größe. Mit 50 kDa liegt das

## 5. Diskussion

Keratin 7 genau in dem zu erwartenden Größenbereich. Wie bereits im Ergebnisteil beschrieben wurde, müssen Keratine als Ergebnis von MALDI-Analysen mit großer Vorsicht interpretiert werden, da es sich hierbei um Verunreinigungen, zum Beispiel durch Hautschuppen, handeln kann. Diese Verunreinigungen würden allerdings keine Bande im SDS-Gel ergeben, sondern müßten nachträglich bei der Behandlung der ausgeschnittenen Gelbande verursacht worden sein. Wenn also tatsächlich Keratin in der Gelbande enthalten war, was aufgrund der Proteingröße möglich ist, würde es sich nicht um eine Verunreinigung, sondern allenfalls um ein unspezifisch aufgereinigtes Protein handeln.

Keratine gehören zu der Klasse der Intermediärfilamente, die zum Aufbau des Cytoskeletts in Epithelgeweben beitragen. Keratin 7 zählt dabei zu den sogenannten "weichen" Keratinen und wird zum Beispiel in Epithelien der Lunge und Niere (Ramaekers et al., 1987), aber auch in Epithelien von Venen exprimiert (Miettinen und Fetsch, 2000). Eine neuere Studie zeigte, daß Keratin 7 darüberhinaus in vielen weiteren Geweben nachweisbar ist und mit dem Keratin 8 co-exprimiert wird (Smith et al., 2002). Das ist insofern interessant, als eine weitere Studie zeigte, daß Keratin 8 in bestimmten Zellen bei Kaninchen Bestandteil des Golgi-assoziierten Filament-Netzwerkes ist (Ogawa et al., 2002). Das bedeutet natürlich nicht, daß das bei den Mausgeweben, aus denen die Proteine aufgereinigt wurden, auch der Fall ist und daß Keratin 7 ebenfalls zu diesem Golgi-assoziierten Filament-Netzwerk gehört. Obwohl es spekulativ ist, wäre es dennoch möglich, daß das Keratin 7 als Bestandteil des Cytoskeletts in Blutgefäßen (die Proteine wurden aus Herzgewebe aufgereinigt) mit dem Stemloop aufgereinigt wurde, da dieser möglicherweise mit dem Cytoskelett assoziiert ist.

Die Analyse des MALDI-Spektrums für die Bande 1 aus Abbildung 4.39 ergab unter anderem das Rab GDI-Protein (GDP dissociation inhibitor) als wahrscheinlichen Treffer. Dieses Protein ist ein regulatorisches Protein für alle Rab Proteine, die beim Vesikeltransport in der Zelle eine Rolle spielen (Ishizaki et al., 2000; Ullrich et al., 1994). Dabei bindet GDI an Rab Proteine, die GDP gebunden haben. Diese heterodimeren Komplexe dienen als cytosolisches Reservoir für Rab Proteine, die durch Austausch des GDPs durch GTP wieder aktiviert werden können (Luan et al., 2000). Durch eine Interaktion mit der Rab1A 3'UTR könnte dieses regulatorische Protein somit direkt in die Nähe des neu translatierten Rab Proteins gebracht werden, um mit ihm zu interagieren. Dies wäre ein sehr effizienter Weg für eine mögliche Regulation des Rab Proteins.

Obwohl sich die erhaltenen MALDI-Spektren aufgrund der wahrscheinlich vorhandenen Proteingemische in den einzelnen analysierten Banden nicht optimal auswerten lassen, ist eine funktionelle Bedeutung für die Interaktion der oben beschriebenen Proteine mit der Rab1A 3'UTR biologisch durchaus vorstellbar. Das gefundene RNA-Bindeprotein könnte dabei direkt an den Stemloop binden, während das GDI Protein wahrscheinlich in einem größeren funktionellen Komplex

mit der Rab1A 3'UTR in Verbindung steht. Das Keratin 7 könnte mit aufgereinigt worden sein, da die Rab1A 3'UTR möglicherweise mit dem Cytoskelett der Zelle assoziiert ist. Das könnte auch eine Erklärung dafür sein, daß die Oligonukleotide bei den Verzögerungsgelen nicht richtig ins Gel einlaufen konnten, da sie durch die Assoziierung mit dem Cytoskelett praktisch unlöslich waren. Nimmt man an, daß die Rab1A 3'UTR tatsächlich im Cytoplasma an bestimmten Positionen lokalisiert wird (wie es die FRET-Experimente vermuten lassen), wäre solch eine Interaktion mit dem Cytoskelett wahrscheinlich. Die Lokalisation einer mRNA erfolgt in der Regel dadurch, daß spezielle Motive innerhalb der 3'UTR von Proteinen erkannt werden, und die mRNA mit Hilfe von Motorproteinen entlang den Bestandteilen des Cytoskeletts zu ihrem Zielort in der Zelle transportiert und dort verankert wird (Kloc et al., 2002). Es konnte außerdem gezeigt werden, daß lokalisierte mRNAs mit vielen weiteren Proteinen, wie zum Beispiel Komponenten der Translationsmaschinerie, in großen Komplexen lokalisiert werden können (Jansen, 2001). Die durchgeführten Experimente im Rahmen meiner Arbeit konnten eine Lokalisation der Rab1A mRNA zwar nicht endgültig nachweisen, die Ergebnisse der FRET Untersuchungen sowie die Hinweise auf eine Assoziation der Rab1A mRNA mit dem Cytoskelett über den vorhergesagten Stemloop sind aber ein guter Ausgangspunkt für weitere Experimente, um eine Lokalisation letztendlich zu beweisen (s. Ausblick). Auch die folgende Diskussion über die mögliche biologische Bedeutung des vorhergesagten Rab1A Stemloops in den 3'UTRs der Sirtuin 1 und Sir2 $\alpha$  mRNAs gibt weitere Hinweise auf eine Funktion des Stemloops für eine mRNA Lokalisation.

### 5.4.3. Der Rab1A Stemloop in der Sirtuin 3'UTR

Sirtuine gehören zu einer Proteinklasse, die sich sowohl in Prokaryoten als auch Eukaryoten finden läßt und deren gemeinsames Merkmal die sogenannte Sirtuin-Kern-Domäne ist, die konservierte Sequenzmotive enthält (Frye, 2000). Sir2 wurde ursprünglich in Hefe entdeckt, wo es unter anderem beim sogenannten "transcriptional silencing" der "silent mating type" Loci beteiligt ist (Gartenberg, 2000) und daher auch seinen Namen hat (silent information regulator). Sir2 fungiert dabei als eine NAD-abhängige Histondeacetylase (Imai et al., 2000). Das humane Sirtuin 1, das auch als Sir2 $\alpha$  bezeichnet wird, und das Sir2 $\alpha$  der Maus sind aufgrund von Sequenzähnlichkeiten die engsten Homologen zum Sir2 der Hefe (Frye, 1999). Zwei neuere Untersuchungen konnten zeigen, daß Sir2 $\alpha$  bei Mensch und Maus einen regulatorischen Einfluß auf p53 hat (Luo et al., 2001; Langley et al., 2002), wobei p53 ein wichtiger Tumorsuppressor ist (Levine, 1997). Das Sir2 $\alpha$  Protein ist sowohl beim Menschen als auch bei der Maus ausschließlich im Kern lokalisiert (Luo et al., 2001). Hier zeigt sich eine spannende Parallele zum Rab1A Protein, daß zwar nicht im Kern, aber auch an einem klar definierten Ort im Cytoplasma lokalisiert ist. Für mehrere Proteine, die im Kern lokalisiert sind, konnte gezeigt werden, daß zuvor die entsprechende mRNA im Bereich des

## 5. Diskussion

sogenannten perinuklearen Cytoskeletts lokalisiert wird, so daß die Proteine nach ihrer Translation direkt in den Kern transportiert werden (Levadoux et al., 2000; Hesketh et al., 1994). Es wäre also durchaus denkbar, daß auch die Sir2 $\alpha$  mRNA zunächst in der Nähe des Kerns lokalisiert wird, bevor das Protein gebildet wird.

Neben dem Sirtuin 1 (bzw. Sir2 $\alpha$ ) sind 6 weitere Sirtuine beim Menschen beschrieben, nämlich die Sirtuine SIRT2-SIRT7 (Frye, 2000). Über die Funktion dieser Sirtuine ist noch nicht viel bekannt, wobei eine aktuelle Arbeit eine mögliche Funktion für SIRT2 beschreibt (s.u.). Im Zusammenhang mit den Sirtuinen ist das Ergebnis eines der MALDI-Spektren besonders interessant. Wie in Abschnitt 4.5.2 beschrieben wurde, ergab eine MALDI-Analyse für die Proteinbande mit der Größe von 37 kDa ein Spektrum mit nur wenigen Fragmentmassen. Die Auswertung dieses Spektrums lieferte für die Mausdatenbank keinen Treffer in dem zu erwartenden Größenbereich. Bei einer Auswertung des Spektrums mit Hilfe einer anderen Datenbank ergab sich das humane SIRT6 als möglicher Kandidat für die Proteinbande. Sirtuin 6 hat eine Größe von 36 kDa und paßt damit sehr gut in den nach der SDS-Gelelektrophorese zu erwartenden Größenbereich. Dennoch muß dieses Ergebnis aufgrund der geringen Fragmentanzahl des MALDI-Spektrums vorsichtig diskutiert werden. Da die beschriebenen humanen Sirtuine eine hohe Sequenzhomologie zu den Sirtuinen der Maus haben, ist es durchaus denkbar, daß auch in der Maus das Sirtuin 6 mit dem Rab1A Stemloop aufgereinigt werden kann. Wie Abbildung 5.2 zeigt, unterscheidet sich der Stemloop in der Rab1A 3'UTR von dem in der Sir2 $\alpha$  3'UTR innerhalb des Loops nur durch eine Base. Daher ist es vorstellbar, daß die mit dem Rab1A Stemloop aufgereinigten Proteine auch funktionell mit dem Stemloop in der Sir2 $\alpha$  3'UTR interagieren können. Somit könnte ein Sirtuin (unter der Annahme, daß es sich bei dem aufgereinigten Protein tatsächlich um Sirtuin 6 handelt) bei der posttranskriptionellen Regulation eines anderen Sirtuins, nämlich Sir2 $\alpha$ , eine Rolle spielen. Über die biologische Funktion des Sirtuin 6 ist noch nichts bekannt. Dagegen ist in einer aktuellen Arbeit für das SIRT2 gezeigt worden, daß es, anders als Sir2 $\alpha$ , im Cytoplasma der Zelle lokalisiert ist (North et al., 2003). Außerdem wurde nachgewiesen, daß SIRT2 zusammen mit anderen Proteinen in einem Multiproteinkomplex entlang des Mikrotubuli-Netzwerkes der Zelle co-lokalisiert ist. Aufgrund der Ähnlichkeit der Sirtuine untereinander wäre es somit durchaus denkbar, daß auch SIRT6 Teil eines solchen Multiproteinkomplexes ist und die Co-Lokalisation mit Mikrotubuli könnte als ein weiterer Hinweis auf eine mögliche Lokalisation der Sir2 $\alpha$  mRNA und auch der Rab1A mRNA angesehen werden.

Obwohl gerade im Hinblick auf die Sirtuin-Ergebnisse vieles Spekulation ist, haben sich aus den bioinformatischen Analysen und experimentellen Ergebnissen der vorliegenden Arbeit viele spannende Hinweise auf mögliche posttranskriptionelle Regulationsmechanismen der untersuchten mRNAs ergeben. Diese bilden den Ausgangspunkt für weitere experimentelle Untersuchungen der beschriebenen mRNAs. Dies wird im folgenden Abschnitt noch einmal abschließend diskutiert.

## 5.5. Ausblick

Die entwickelten Strategien zur Strukturmotivvorhersage und Validierung wurden in dieser Arbeit ausschließlich zur Analyse von untranslatierten Regionen in mRNAs eingesetzt. In den letzten Jahren sind jedoch eine Vielzahl sogenannter "noncoding RNAs", also nicht für Protein codierende RNAs entdeckt worden und es wird angenommen, daß die Zahl dieser RNAs weiter ansteigen wird (Storz, 2002). Viele dieser "noncoding RNAs" enthalten bestimmte Struktur motive, die zum einen für die Funktion wichtig sind und zum anderen die jeweilige "noncoding RNA" Klasse charakterisieren. Ein typisches Beispiel hierfür sind die sogenannten H/ACA-Box snoRNAs (small nucleolar RNAs) (Kiss, 2002).

Die Durchführung multipler Strukturalignments solcher "noncoding RNAs" könnte eingesetzt werden, um eine Konsensusstruktur oder zumindest bestimmte Struktur motive, die einer Klasse von "noncoding RNAs" gemeinsam sind, zu bestimmen. Diese könnten dann in ein Suchmuster umgesetzt werden, um durch den Einsatz von HyPa neue Vorkommen der jeweiligen RNA zu entdecken. So ließen sich zum Beispiel auch durch die Berechnung multipler Strukturalignments bereits bekannter Stemloop-Precursor von micro-RNAs Konsensusstrukturen solcher Precursor ermitteln, die dann zur Vorhersage neuer miRNAs eingesetzt werden könnten.

Durch die Kombination von bioinformatischen Strategien zur Entdeckung sogenannter "conserved noncoding sequences" (CNS), wie sie zum Beispiel in der Doktorarbeit von Jomuna Choudhuri entwickelt wurden (Choudhuri, 2003), mit den in dieser Arbeit entwickelten Strategien zur Vorhersage gemeinsamer Struktur motive, könnten Regionen, die nicht für Proteine kodieren und dennoch zwischen verschiedenen Spezies konserviert sind, daraufhin überprüft werden, ob sie funktionale RNAs enthalten können.

Da funktionale RNAs häufig durch ihre Sekundärstruktur bzw. bestimmte Sekundärstruktur motive charakterisiert sind, können die in dieser Arbeit entwickelten Strategien in Zukunft somit zur Analyse von "noncoding RNAs" eingesetzt werden.

Die Ergebnisse der experimentellen Analysen für den vorhergesagten Stemloop in der Rab1A 3'UTR haben gezeigt, daß dieser mit einem großen Proteinkomplex interagiert. Aufgrund der MALDI-Analysen ist es ziemlich wahrscheinlich, daß sowohl das regulatorische Rab GDI-Protein als auch ein potentiell RNA-Bindeprotein in diesem Komplex enthalten sind. Auch eine Assoziierung der Rab1A mRNA mit dem Cytoskelett der Zelle scheint möglich zu sein. Da jedoch sowohl die Identifizierung des Keratin 7 als auch des Sirtuin 6 durch die MALDI-Experimente mit einem gewissen Unsicherheitsfaktor bewertet werden muß, bieten sich zur Überprüfung der Assoziation der Rab1A mRNA mit dem Cytoskelett weitere Experimente an. Hierfür würden sich Co-Lokalisationsexperimente eignen. Dabei würde man versuchen, zunächst die Lokalisation der Rab1A mRNA

## 5. Diskussion

zu zeigen, wofür man theoretisch wieder die FRET Experimente einsetzen könnte, allerdings an Gewebeschnitten. In diesen Gewebeschnitten könnten dann durch spezifische Antikörper Teile des Cytoskeletts sichtbar gemacht werden, um nachzuweisen, ob die Rab1A mRNA in demselben Bereich lokalisiert ist. Mit solchen Co-Lokalisationsstudien ließe sich natürlich auch direkt überprüfen, ob die identifizierten Proteine tatsächlich mit der Rab1A mRNA assoziiert sind. Hierzu müßten spezielle Antikörper gegen die identifizierten Proteine verwendet werden, um ihre Lokalisation in Gewebeschnitten nachzuweisen. Hier wäre vor allem die Interaktion mit dem Rab GDI-Protein von Interesse, da das einen möglichen neuen Regulationsmechanismus für das Rab1A Protein darstellen könnte. Dabei würde ein regulatorisches Protein direkt mit der mRNA des zu regulierenden Proteins interagieren.

Ähnliche Fragestellungen wie für die Rab1A mRNA ergeben sich auch für die Sir2 $\alpha$  mRNA, deren 3'UTR ebenfalls den vorhergesagten Stemloop enthält. Die mögliche Interaktion mit dem Sirtuin 6 ist dabei besonders spannend. Für die Sir2 $\alpha$  mRNA könnte man auch versuchen, eine Lokalisation, möglicherweise im perinuklearen Raum, nachzuweisen. Aufgrund des nicht eindeutig auswertbaren Spektrums müßte die Interaktion des Stemloops bzw. der Sir2 $\alpha$  mRNA oder auch der Rab1A mRNA mit Sirtuin 6 jedoch mit anderen experimentellen Methoden (hierfür könnten Co-Immunopräzipitationsexperimente eingesetzt werden) verifiziert werden.

Obwohl sich mit den oben beschriebenen Experimenten nachweisen ließe, daß die Rab1A mRNA tatsächlich mit dem Cytoskelett der Zelle assoziiert ist und dadurch lokalisiert wird, könnte man zur Aufklärung der tatsächlichen Bedeutung des vorhergesagten Stemloops für die posttranskriptionelle Genregulation einer mRNA sogenannte Reporter-gen-Analysen durchführen. Hierzu würde man die 3'UTR, einmal mit und einmal ohne den Stemloop, hinter die codierende Sequenz eines Reportergens, zum Beispiel *lacZ*, klonieren. Da man sowohl die mRNA (durch *in situ* Hybridisierung) als auch das *lacZ* Proteinprodukt in Gewebeschnitten nachweisen kann, könnte überprüft werden, ob die 3'UTR Einfluß auf die Lokalisation oder auch die Translationseffizienz des Reportergens hat und ob sich dieser Einfluß durch den fehlenden Stemloop ändert.

# Literaturverzeichnis

M.D. Adams et al. The genome sequence of *Drosophila melanogaster*. *Science*, 287:2185–2195, 2000.

A. Apostolico, M. Bock, S. Lonardi, und X. Xu. Efficient detection of unusual words. *J. Comput. Biol.*, 7:71–94, 2000.

C.A. Beelman und R. Parker. Degradation of mRNA in eukaryotes. *Cell*, 81:179–183, 1995.

R. Bellman und R. Kabala. On kth best policies. *J. Soc. Indust. Appl. Math.*, 8:582–588, 1960.

S.L. Berger. Histone modifications in transcriptional regulation. *Curr. Opin. Genetics Dev.*, 12:142–148, 2002.

G. Brewer. Regulation of c-myc mRNA decay in vitro by a phorbol ester-inducible, ribosome associated component in differentiating megakaryoblasts. *J. Biol. Chem.*, 275:33336–33345, 2000.

E. Carballo, W.S. Lai, und P.J. Blackshear. Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony stimulating factor messenger RNA deadenylation and stability. *Blood*, 95:1891–1899, 2000.

L. Cartegni, S.L. Chew, und A.R. Krainer. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev Genet*, 3:285–298, 2002.

J.L. Casey, B. Di Jeso, K. Rao, R.D. Klausner, und J.B. Harford. Two genetic loci participate in the regulation by iron of the gene for the human transferrin receptor. *Proc. Natl. Acad. Sci. USA*, 85:1787–1791, 1988.

J.L. Casey, D.M. Koeller, V.C. Ramin, R.D. Klausner, und J.B. Harford. Iron regulation of transferrin receptor mRNA levels requires iron-responsive elements and a rapid turnover determinant in the 3' untranslated region of the mRNA. *EMBO J.*, 8:3693–3699, 1989.

C. Chen und A. Shyu. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci*, 20:465–470, 1995.

## LITERATURVERZEICHNIS

- J.L. Chen, M. Blasco, und C.W. Greider. Secondary structure of vertebrate telomerase RNA. *Cell*, 100:503–514, 2000.
- J.V. Choudhuri. *Bioinformatics approaches to large scale genome comparison, including the identification of Conserved Noncoding Regions*. Doktorarbeit, Universität Bielefeld, 2003.
- D. Conclin, I. Jonassen, R. Aasland, und W.R. Taylor. Association of nucleotide patterns with gene function classes: application to human 3' untranslated sequences. *Bioinformatics*, 18:182–189, 2002.
- T. Dandekar und W. Hentze. Finding the hairpin in the haystack: searching for RNA motifs. *Trends Genet.*, 11:45–50, 1995.
- T.K. Darlington, K. Wager-Smith, M.F. Ceriani, D. Stakni, N. Gekakis, T.D.L. Steeves, C.J. Weitz, J.S. Takahashi, und S.A. Kay. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science*, 280:1599–1603, 1998.
- M. Deutsch und M. Long. Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res*, 27:3219–3228, 1999.
- G. Dirheimer, P. Keith, P. Dumas, und E. Westhof. Primary, secondary and tertiary structures of tRNAs. In D. Söll und U. Rajbahandry, Herausgeber, *RNA: Structure, Biosynthesis and Function*, Seiten 93–126. American Society for Microbiology, Washington DC, 1995.
- W. Driever und C. Nüsslein-Volhard. The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell*, 54:95–104, 1988.
- S. Eddy und R. Durbin. RNA sequence analysis using covariance models. *Nucleic Acids Res.*, 22:2079–2088, 1994.
- R.A. Frye. Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have ADP-ribosyltransferase activity. *Biochem. Biophys. Res. Commun.*, 260:273–279, 1999.
- R.A. Frye. Phylogenetic classification of Prokaryotic and Eukaryotic Sir2-like proteins. *Biochem. Biophys. Res. Commun.*, 273:793–798, 2000.
- M.R. Gartenberg. The Sir proteins of *Saccharomyces cerevisiae*: mediators of transcriptional silencing and much more. *Curr. Opin. Microbiol.*, 3:132–137, 2000.



- E.R. Gavis und R. Lehmann. Localization of nanos RNA controls embryonic polarity. *Cell*, 71:301–313, 1992.
- R. Giegerich und C. Meyer. Algebraic Dynamic Programming. In H. Kirchner und C. Ringeissen, Herausgeber, *Algebraic Methodology And Software Technology, 9th International Conference, AMAST 2002*, Seiten 349–364. Saint-Gilles-Bains, Reunion Island, France, 2002.
- A.C. Gingras, B. Raught, und N. Sonnenberg. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev. Biochem.*, 68:913–963, 1999.
- J. Gorodkin, L.J. Heyer, und G.D. Stormo. Finding the most significant common sequence and structure motifs in a set of RNA sequences. *Nucleic Acids Res.*, 25:3724–3732, 1997.
- J. Gorodkin, S.L. Stricklin, und G.D. Stormo. Discovering common stem-loop motifs in unaligned RNA sequences. *Nucleic Acids Res.*, 29:2135–2144, 2001.
- B.R. Graveley. Sorting out the complexity of SR protein functions. *RNA*, 6:1197–1211, 2000.
- M.R. Green. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci*, 25:59–63, 2000.
- J. Guhaniyogi und G. Brewer. Regulation of mRNA stability in mammalian cells. *Gene*, 265:11–23, 2001.
- R.R. Gutell. Collection of small subunit (16 S and 16 S-like) ribosomal RNA structures. *Nucleic Acids Res*, 22:3502–3507, 1994.
- J.B. Harford, J.L. Casey, D.M. Koeller, und R.D. Klausner. Structure, function and regulation of the transferrin receptor: insights from molecular biology. In C.J. Steer und J.A. Hanover, Herausgeber, *Intracellular Trafficking of Proteins*, Seiten 302–334. Cambridge University Press, Cambridge, 1990.
- J.B. Harford und R.D. Klausner. Coordinate post-transcriptional regulation of ferritin and transferrin receptor expression: the role of regulated RNA-protein interaction. *Enzyme*, 44:28–41, 1990.
- M.L. Hastings und A.R. Krainer. Pre-mRNA splicing in the new millenium. *Curr. Opin. Cell Biol.*, 13:302–309, 2001.
- M.W. Hentze, S.W. Caughman, T.A. Rouault, J.G. Barriocanal, A. Dancis, J.B. Harford, und R.D. Klausner. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science*, 238:1570–1573, 1987.

- J. Hesketh, G. Campbell, M. Piechaczyk, und J.M. Blanchard. Targeting of c-myc and  $\beta$ -globin coding sequences to the cytoskeletal-bound polysomes by c-myc 3'untranslated region. *Biochem. J.*, 298:143–148, 1994.
- M. Höchsmann, T. Töller, R. Giegerich, und S. Kurtz. Local Similarity in RNA Secondary Structures. In *Proc. of the IEEE Computer Society Bioinformatics Conference (CSB) in press.* 2003.
- I.L. Hofacker, W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, und P. Schuster. Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte f. Chemie*, 125:167–188, 1994.
- F.C. Holstege, E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, und R.A. Young. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95:717–728, 1998.
- I. Hu, B. Wightman, und G. Ruvkun. A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes Dev.*, 10:3041–3050, 1996.
- Y.J. Hu. Prediction of consensus structural motifs in a family of coregulated RNA sequences. *Nucleic Acids Res.*, 30:3886–3893, 2002.
- M.J. Hughes und D.W. Andrews. A single nucleotide is a sufficient 5'untranslated region for translation in an eukaryotic in vitro system. *FEBS Lett*, 414:19–22, 1997.
- S. Imai, C.M. Armstrong, M. Kaerberlein, und L. Guarente. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature*, 403:795–800, 2000.
- H. Ishizaki, J. Miyoshi, H. Kamiya, A. Togawa, M. Tanaka, T. Sasaki, K. Endo, A. Mizoguchi, S. Ozawa, und Y. Takai. Role of Rab GDP dissociation inhibitor alpha in regulating plasticity of hippocampal neurotransmission. *Proc. Natl. Acad. Sci. USA*, 97:11587–11592, 2000.
- R.P. Jansen. mRNA localization: Message on the move. *Nature Rev Mol Cell Biol*, 2:247–256, 2001.
- Z. Kan, E.C. Rouchka, W.R. Gish, und D.J. States. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.*, 11:889–900, 2001.
- J. Kato, K. Fujikawa, M. Kanda, N. Fukuda, K. Sasaki, T. Takayama, M. Kobune, K. Takada, R. Takimoto, H. Hamada, T. Ikeda, und Y. Niitsu. A mutation, in the iron-responsive element of H ferritin mRNA, causing autosomal dominant iron overload. *Am. J. Hum. Genet.*, 69:191–197, 2001.

- L.P. Keegan, A. Gallo, und M.A. O'Connell. The many roles of an RNA Editor. *Nature Rev Genet*, 2:869–878, 2002.
- C.A. Keller, M.A. Grill, und S.M. Abmayr. A role for nautilus in the differentiation of muscle precursors. *Dev. Biol.*, 202:157–171, 1998.
- T. Kiss. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, 109:145–148, 2002.
- R.D. Klausner, T.A. Rouault, und J.B. Harford. Regulating the fate of mRNA: The control of cellular iron metabolism. *Cell*, 72:19–28, 1993.
- M. Kloc, R. Zearfoss, und L.D. Etkin. Mechanisms of subcellular mRNA localization. *Cell*, 108:533–544, 2002.
- C.G. Korner, M. Wormington, M. Muckenthaler, S. Schneider, E. Dehlin, und E. Wahle. The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of *Xenopus* oocytes. *EMBO J.*, 17:5427–5437, 1998.
- M. Kozak. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*, 15:8125–8184, 1987.
- U.K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227:680–685, 1970.
- M. Lagos-Quintana, R. Rauhut, W. Lendeckel, und T. Tuschl. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–858, 2001.
- M. Lagos-Quintana, R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel, und T. Tuschl. Identification of tissue-specific MicroRNAs from mouse. *Curr. Biol.*, 12:753–739, 2002.
- E. Langley, M. Pearson, M. Faretta, U.M. Bauer, R.A. Frye, S. Minucci, P.G. Pelicci, und T. Kouzarides. Human Sir2 deacetylates p53 and antagonizes PML/p53-induced cellular senescence. *EMBO J.*, 21:2383–2396, 2002.
- N.C. Lau, L.P. Lim, E.G. Weinstein, und D.P. Bartel. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294:858–862, 2001.
- R.C. Lee und V. Ambros. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 294:862–864, 2001.
- R.C. Lee, R.L. Feinbaum, und V. Ambros. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14* RNA. *Cell*, 75:843–854, 1993.

## LITERATURVERZEICHNIS

- Y. Lee, K. Jeon, J.T. Lee, S. Kim, und V.N. Kim. microRNA maturation: stepwise processing and subcellular localization. *EMBO J.*, 21:4663–4670, 2002.
- E.P. Lei und P.A. Silver. Protein and RNA Export from the nucleus. *Dev Cell*, 2:261–272, 2002.
- M. Levadoux, C. Mahon, J.H. Beattie, H.M. Wallace, und J.E. Hesketh. Nuclear import of metallothionein requires its mRNA to be associated with perinuclear cytoskeleton. *J. Biol. Chem.*, 274:34961–34966, 2000.
- A.J. Levine. p53, the cellular gatekeeper for growth and division. *Cell*, 88:323–331, 1997.
- L.P. Lim, N.C. Lau, E.G. Weinstein, A. Abdelhakim, S. Yekta, M.W. Rhoades, C.B. Burge, und D.P. Bartel. The microRNAs of *Caenorhabditis elegans*. *Genes Dev*, 17:991–1008, 2003.
- H.D. Lipshitz und C.A. Smibert. Mechanisms of RNA localization and translational regulation. *Curr. Opin. Genetics Dev.*, 10:476–488, 2000.
- T.M. Lowe und S.R. Eddy. tRNAscan-se: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, 25:955–964, 1997.
- P. Luan, A. Heine, K. Zeng, B. Moyer, S.E. Greasely, P. Kuhn, W.E. Balch, und I.A. Wilson. A new functional domain of guanine nucleotide dissociation inhibitor( $\alpha$ -GDI) involved in Rab recycling. *Traffic*, 1:270–281, 2000.
- J. Luo, S. Nikolaev, A.Y. Imai, D. Chen, F. Su, A. Shiloh, L. Guarente, und W. Gu. Negative control of p53 by Sir2 $\alpha$  promotes cell survival under stress. *Cell*, 107:137–148, 2001.
- P.M. Macdonald, K. Kerr, J.L. Smith, und A. Leask. RNA regulatory element BLEI directs the early steps of bicoid mRNA localization. *Development*, 118:1233–1243, 1993.
- T.J. Macke, D.J. Ecker, R.R. Gutell, D. Gautheret, D.A. Case, und R. Sampath. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res*, 29:4724–4735, 2001.
- T. Maniatis und T. Bosiljka. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, 418:236–243, 2002.
- O. Martinez und B. Goud. Rab proteins. *Biochim. Biophys. Acta*, 1404:101–112, 1998.
- J. Massague. How cells read TGF- $\beta$  signals. *Nature Rev Mol Cell Biol.*, 1:169–178, 2000.

- D.H. Mathews, J. Sabina, M. Zuker, und D.H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.
- C. Meyer und R. Giegerich. Matching and Significance Evaluation of combined sequence-structure motifs in RNA. *Z. Phys. Chem.*, 216:193–216, 2002.
- M. Miettinen und J.F. Fetsch. Distribution of keratins in normal endothelial cells and a spectrum of vascular tumors: implications in tumor diagnosis. *Hum Pathol.*, 31:1062–1067, 2000.
- F. Mignone, C. Gissi, S. Liuni, und G. Pesole. Untranslated regions of mRNAs. *Genome Biol.*, 3:1–10, 2002.
- A.A. Mironov, J.W. Fickett, und M.S. Gelfand. Frequent alternative splicing of human genes. *Genome Res.*, 9:1288–1293, 1999.
- B. Modrek und C. Lee. A genomic view of alternative splicing. *Nature Genet.*, 30:13–19, 2002.
- B. Modrek, A. Resch, C. Grasso, und C. Lee. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.*, 29:2850–2859, 2001.
- B. Morgenstern. DIALIGN2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics*, 15:211–218, 1999.
- D.P. Morse und B.L. Bass. Long RNA hairpins that contain inosine are present in *Caenorhabditis elegans* poly(A)+ RNA. *Proc. Natl. Acad. Sci. USA*, 96:6048–6053, 1999.
- K.L. Mowry und D.A. Melton. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science*, 255:991–993, 1992.
- C. W. Müller. Transcription factors: global and detailed views. *Curr. Opin. Struct. Biol.*, 11:26–32, 2001.
- B.J. North, B.L. Marshall, M.T. Borra, J.M. Denu, und E. Verdin. The human Sir2 ortholog, SIRT2, is an NAD-dependent tubulin deacetylase. *Mol Cell*, 11:437–444, 2003.
- C. Nuoffer, H.W. Davidson, J. Metteson, J. Meinkoth, und W.E. Balch. A GDP-bound form of rab1 inhibits protein export from the endoplasmic reticulum and transport between Golgi compartments. *J. Cell Biol.*, 125:225–237, 1994.

- C. Ogawa, H. Iwatsuki, M. Suda, und K. Sasaki. Golgi-associated filament networks in duct epithelial cells of rabbit submandibular glands: immunohistochemical light and electron microscopic studies. *Histochem Cell Biol.*, 118:35–40, 2002.
- N.R. Pace, B.C. Thomas, und C.R. Woese. Probing RNA structure, function and history by comparative analysis. In R.F. Gesteland, T.R. Cech, und J.F. Atkins, Herausgeber, *The RNA World*, Seiten 113–141. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2. Auflage, 1999.
- J. Parsch, J.M. Braverman, und W. Stephan. Comparative sequence analysis and patterns of covariation in RNA secondary structures. *Genetics*, 154:909–921, 2000.
- A.E. Pasquinelli et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408:86–89, 2000.
- J. Pelletier, G. Kaplan, V.R. Racaniello, und N. Sonenberg. Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Mol. Cell Biol.*, 8:1103–1112, 1988.
- G. Pesole, G. Fiormarino, und C. Saccone. Sequence analysis and compositional properties of untranslated regions of human mRNAs. *Gene*, 140:219–225, 1994.
- G. Pesole, S. Liuni, G. Grillo, F. Licciulli, F. Mignone, C. Gissi, und C. Saccone. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002. *Nucleic Acids Res*, 30:335–340, 2002.
- G. Pesole, S. Liuni, G. Grillo, und C. Saccone. Structural and compositional features of untranslated regions of eukaryotic mRNAs. *Gene*, 205:95–102, 1997.
- G. Pesole, F. Mignone, C. Gissi, G. Grillo, F. Licciulli, und S. Liuni. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 276:73–81, 2001.
- L.M. Powell, S.C. Wallis, R.J. Pease, Y.H. Edwards, T.J. Knott, und J. Scott. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell*, 50:831–840, 1987.
- F.C.S. Ramaekers, A. Huysmans, G. Schaart, O. Moesker, und P. Vooijs. Tissue distribution of keratin 7 as monitored by a monoclonal antibody. *Exp Cell Res.*, 170:235–249, 1987.
- R. Reed. Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.*, 12:340–345, 2000.

- R. Reed und E. Hurt. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*, 108:523–531, 2002.
- B.J. Reinhart, F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, und G. Ruvkun. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403:901–906, 2000.
- M.W. Rhoades, B.J. Reinhart, L.P. Lim, C.B. Burge, B. Bartel, und D.P. Bartel. Prediction of plant MicroRNA targets. *Cell*, 110:513–520, 2002.
- L. Rowen, J. Young, B. Birditt, A. Kaur, A. Madan, D.L. Phillips, S. Qjn, P. Minx, R.K. Wilson, L. Hood, und B.R. Graveley. Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics*, 79:587–597, 2002.
- G.M. Rubin et al. Comparative genomics of the eukaryotes. *Science*, 287:2204–2215, 2000.
- S.M. Rueter, T.R. Dawson, und R.B. Emeson. Regulation of alternative splicing by RNA editing. *Nature*, 399:75–80, 1999.
- J.E. Russel, J. Morales, und S.A. Liebhaber. The role of mRNA stability in the control of globin gene expression. *Prog. Nucleic Acid Res. Mol. Biol.*, 57:249–287, 1997.
- C. Saavedra, C.M. Hammel, C.V. Heath, und C.N. Cole. Yeast heat shock mRNAs are exported through a distinct pathway defined by Rip1p. *Genes Dev.*, 11:2845–2856, 1997.
- J. Sambrook, E. Fritsch, und T. Maniatis. *Molecular Cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2. Auflage, 1989.
- D. Sankoff. Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J. Appl. Math.*, 45:810–825, 1985.
- J. Saraste, U. Lahtinen, und B. Goud. Localization of the small GTP-binding protein rab1p to early compartments of the secretory pathway. *J. Cell Sci.*, 108:1541–1552, 1995.
- F.J.D. Smith, R.M. Porter, L.D. Corden, D.P. Lunny, B.E. Lane, und I. McLean. Cloning of human, murine, and marsupial keratin 7 and a survey of K7 expression in the mouse. *Biochem. Biophys. Res. Commun.*, 297:818–827, 2002.
- T.F. Smith und M.S. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147:195–197, 1981.

## LITERATURVERZEICHNIS

- J.P. Staley und C. Guthrie. Mechanical devices of the spliceosome: motors, clocks, springs and things. *Cell*, 92:315–326, 1998.
- J.M. Staton, A.M. Thomson, und P.J. Leedman. Hormonal regulation of RNA stability and RNA-protein interactions in the pituitary. *J. Mol. Endocrinology*, 25:17–34, 2000.
- S.W. Stevens, D.E. Ryan, H.Y. Ge, R.E. Moore, M.K. Young, T.D. Lee, und J. Abelson. Composition and functional characterisation of the Yeast spliceosomal Penta-snRNP. *Mol Cell*, 9:31–44, 2002.
- G. Stoesser et al. The EMBL Nucleotide Sequence Database: major new developments. *Nucleic Acids Res*, 31:17–22, 2003.
- G. Storz. An expanding universe of noncoding RNAs. *Science*, 296:1260–1263, 2002.
- A. Suzuki, E. Kaneko, N. Ueno, und A. Hemmati-Brivanlou. Regulation of epidermal induction by BMP2 and BMP7 signaling. *Dev. Biol.*, 189:112–122, 1997.
- Y.V. Svitkin, A. Pause, A. Haghighat, S. Pyronnet, G. Witherell, G.J. Belsham, und N. Sonenberg. The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA*, 7:382–394, 2001.
- J.D. Thompson, D.G. Higgins, und T.J. Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22:4673–4680, 1994.
- A.M. Thomson, J.T. Rogers, und P.J. Leedman. Iron-regulatory proteins, iron-responsive elements and ferritin mRNA translation. *Int J. Biochem. Cell Biol.*, 31:1139–1152, 1999.
- O. Ullrich, H. Horiuchi, C. Bucci, und M. Zerial. Membrane association of Rab5 mediated by GDP-dissociation inhibitor and accompanied by GDP/GTP exchange. *Nature*, 368:157–160, 1994.
- J. van Helden, M. del Omo, und J. Perez-Ortin. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic Acids Res.*, 28:1000–1010, 2000.
- R. Walczak, P. Carbon, und A. Krol. An essential non-Watson-Crick base pair motif in 3'UTR to mediate selenoprotein translation. *RNA*, 4:74–84, 1998.



N. Wedemeyer, T. Schmitt-John, D. Evers, C. Thiel, D. Eberhard, und H. Jokusch. Conservation of the 3'-untranslated region of the Rab1a gene in amniote vertebrates: exceptional structure in marsupials and possible role for posttranscriptional regulation. *FEBS Letters*, 477:49–54, 2000.

M. Wickens, E.B. Goodwin, J. Kimble, S. Strickland, und M. Hentze. Translational control of developmental decisions. In N. Sonenberg, J.W.B. Hershey, und M.B. Mathews, Herausgeber, *Translational Control of Gene Expression*, Seiten 295–370. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000.

A.S. Williams und W.F. Marzluff. The sequence of the stem and flanking sequences at the 3'end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. *Nucleic Acids Res*, 23:654–662, 1995.

E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruss, I. Reuter, und F. Schacherer. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res*, 28:316–319, 2000.

C.R. Woese und N.R. Pace. Probing RNA structure, function and history by comparative analysis. In R.F. Gesterland und J.F. Atkins, Herausgeber, *The RNA World*, Seiten 91–117. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993.

C. Wu. Heat shock transcription factors: structure and regulation. *Annu Rev. Cell. Dev. Biol.*, 11:441–469, 1995.

S. Wuchty, W. Fontana, I.L. Hofacker, und P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49:145–165, 1999.

K. Yamada, Y. Takabatake, T. Takabatake, und K. Takeshima. The early expression control of Xepsin by nonaxial and planar posteriorizing signals in *Xenopus* epidermis. *Dev. Biol.*, 214:318–330, 1999.

M. Zacharias und H. Sklenar. Conformational analysis of single-base bulges in A-form DNA and RNA using a hierarchical approach and energetic evaluation with a continuum solvent model. *J. Mol. Biol.*, 289:261–275, 1999.

M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244:48–52, 1989.

M. Zuker, D.H. Mathews, und D.H. Turner. Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide. In J. Barciszewski und B.F.C. Clark, Herausgeber, *RNA Biochemistry and Biotechnology*, Seiten 11–43. NATO ASI Series, Kluwer Academic Publishers, Dordrecht, NL, 1999.

## LITERATURVERZEICHNIS

M. Zuker und P. Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, 9:133–148, 1981.

# A. Sequenzen

## Histon 3'UTRs

>3MMU005868 3'UTR in M.domesticus (CD-1) mRNA for histone H2A (partial) 108 bp

```
ggccacugcucgucggccccgucucugugauauaacccccgagcucccaaaaagg  
cucuuucagagccaccacugaaucagauaaagaguugugucacggu
```

>3MMU000265 3'UTR in M.domesticus (CD-1) mRNA for histone H3 (partial) 110 bp

```
auuagguagugagugaauuuggaccccaaggcucuuucagagccaccacauuuucu  
auaaaaggcugauaucgaaagcuuuuauaacccccacucagcaacucc
```

>3RN0001862 3'UTR in Rat testis-specific histone UH2B mRNA, complete cds. 45 bp

```
gcccacagagcucacaacagacccaaaggcucuuucagagccac
```

>3HSA014419 3'UTR in Homo sapiens histone H2A.2 mRNA, complete cds. 118 bp

```
ggcugacgucggccaagugggcccagcccggcccgcgucucgaaggggaccugugaa  
cucaaaaaggcucuuucagagccaccacguuuucaaaauaaaagaguuguaauncguc
```

## Ferritin 5'UTRs

>5HSA015337 5'UTR in Human mRNA for ferritin heavy chain, 5'UTR (sequence from the 5'cap to the start codon)

```
ccagacguucucgcccagagucgucgggguuuccugcuucaacagugcuuggacggaac  
ccggcgcucguuccccaccccggcggcggcccacauagccagcccuccgucaccucuuac  
cgcaccucgagcugccccaaggcccccgccgcccagcgcgcccagccaccgccc  
ccgcccggccucuccuagucgcccgc
```

>5MMU002159 5'UTR in Mouse ferritin heavy chain gene, complete cds.

```
cagacguucucgcccagagucgcccgggguuuccugcuucaacagugcuugaacggaacc  
ggugcucgaccccuccgacccccgcccggcggcggcccagccugagccuuugcaacuucg  
cguuccgcccucagcugcggaccgcccgcgcccgcgcccgcgcccacc
```







>3HSA008842 3'UTR in Human mRNA for transferrin receptor  
2464 bp

augugauacccauagcuuccaugagaacagcaggguagucugguuucuagacuugugcuga  
ucgugcuaaauuuucaguagggcuacaaaaccugauguuaaaauuccaucccaucaucuug  
guacuacuagugucuuuaggcagcagcuuuaauacaggguagauaaccuguacuucaag  
uuaaagugaaaaccacuuaaaaauguccauguggauuuucccuaucucuagauuu  
uaagugcuuuguaaugggaaacugccucuuuuccuguguuuguaaaugaaaaugcagaaacc  
aguuaugugaaugaucucucugaauccuaagggcuggucucugcugaagguguaaguggu  
ucgcuuacuugagugauccuccacuucauuugaugcuaaauagggagauaccagguugaa  
agaccucuccaaaugagaucuaagccuuuccauaaggaauguagcagguuuccucauucc  
gaagaaacaguuaacuucagaagagaugggcuuguuucuugccaaugaggucugaau  
ggagguccuucugcuggauaaaaugagguucaacugugauugcaggaauaaggccuuaau  
auguuaaccucagugcauuuaugaaaagaggggaccagaagccaaagacuuaguauuuu  
ucuuuuccucuguccucccccauagccucauuuagucuuuguauuuuuguuucu  
ccaaagcauugaaagagaaccaguuccagguuuagugugcagacucaguugucagac  
uuuaagaauaugugcccaauuuggccaaaguguaaacuuaggggagagcuucu  
guccuuuggcacugagauuuuuauguuuuuuaucagugacagaguucacuaaaaagg  
uguuuuuuaagaauaauuaucggaagcagugccuccauaauuaugacaguuuac  
ugucgguuuuuuuaauuaaagcagcaucugcuauuaaacccaacagauacuggaguu  
uugcauuuuaggucaaccuaaagguuuuagaaaacagccgucagccaaaugaauugaa  
uaaaguugagcuaaguuuagagaugaauuaauuuuuuuuagggugcuaagagcg  
cacugaccagauaaagaugcugguuuccaaaugcagugaauugaccaaguuuaauu  
caugucuuaaaggcuguguguaguacuccugcaaauuuuuaugcuaguuuuauccaa  
guguaacucauuuccauugcaaaauuuccaguaccuugucacaauccuaacacauua  
ucgggagcagugucuuccaaauguaaagaacaagguaguuuuaccuaccagugu  
uguaucgggagacagugaucccauaguuaccuaaggguguaaguuuuuucgggagcag  
uguuucccaaauuucucaugcaaugcaucuucaaagcuugaagaucguuaguaucua  
acauguauccaacuccuauaaucccuaucuuuuaguuuagugcagaacauuuugu  
gucauuagcauggggugguaaauucaaccacuguaaaaugaauuacuacaaauuuga  
aauuuagcuggguuuguaccuuuaugguuuccaggguccucuacuuaaugagauag  
cagcauacauuuaaauguugcuauugcaagucauuuuuaauuuuaccauuuuugca  
guuaccuccuauaacuuagugcggacaguuuaauccagaauugaccuuugacuuaaa  
gcagagggacuuguauagaagguugggggcuguggggagagucccugaaggu  
gacacgucugccuacccauucguggugaucauuaauuaguagguaugaauaaguucgaag  
uccgugagugaaccaucauaaaacguguaguacagcuguugucauagggcaguuggaa  
cggccuccuagggaaaguucauagggucucuucagguucuuagugucacuuaccuagau  
uacagccucacuugaaugugucacuacucacagucucuuuaaucuucaguuuuacuuua  
ucuccuuuuuaucuuggacugacauuuagcguagcuaagugaaaggucauagcugagau  
uccgguucggguguuacgcacacguacuuaauuagaaagcauguggcaugucaucuguaa  
acacaauagaauacagggcaugcauuuugcagcagugagucucuucagaaaaccuuuuc  
uacaguuaggguaguacuccuaucaagccaguacgugcuaacagggcucauuuccu  
gaaugaauaucagacuagugacaagcuccgugucuugagagucuucguuaaagggagu  
gggcuuuuggagguaaagguaa

## A. Sequenzen



## B. Abkürzungsverzeichnis

APS	Ammoniumperoxodisulfat
ARE	AU-reiches Element
BMP	Bone Morphogenetic Protein
CPE	cytoplasmatisches Polyadenylierungselement
DIG	Digoxigenin
DNA	Desoxyribonukleinsäure
DTT	Dithiothreitol
EDTA	Ethylendiamintetraacetat
EST	Expressed sequence tag
FRET	Fluoreszenz Resonanz Energie Transfer
GPRM	Genetic Programming for RNA Motifs
h	Stunde
IRE	<i>Iron Responsive Element</i>
IRES	Internal ribosome entry site
kDa	kilo-Dalton
l	Liter
M	molar
MALDI	Matrix assisted laser desorption ionisation
min	Minute
mRNA	messenger RNA
miRNA	micro RNA
PAA	Polyacrylamid
PBS	Phosphate-buffered saline
RAB	Ras-like in rat brain
RNA	Ribonukleinsäure
rpm	rounds per minute
SDS	Sodiumdodecylsulfat
SIR	silent information regulator
TBE	Tris-Borat-EDTA Puffer
TOF	time of flight
Tris	Trishydroxymethylaminomethan
tRNA	transfer RNA
UTR	untranslatierte terminale Region
V	Volt
v	volume
w	weight