

Quantitativer Vergleich der Initiator-Regionen (Ribosome Binding Sites) von 12 aus *Escherichia coli* stammenden Nukleotidsequenzen (RNA- und DNA-Phagen), die aus Triplets zusammengesetzt sind

Quantitative Comparison of Ribosome Binding Sites of Twelve Nucleotide Sequences from *Escherichia coli* (RNA- and DNA Phages) Based on Triplet Patterns

Erich Köhler

Biologische Bundesanstalt, Messeweg 11/12, D-3300 Braunschweig

Z. Naturforsch. **34 c**, 797 – 804 (1979); eingegangen am 23. April 1979

Coli phages, Conformity Indices from Alignments, Quarternary and Binary Triplet Codes

The molecular structure of ribosome binding sites of ten phage genes and two messengers of *Escherichia coli* were compared concerning the signation parts which are presumably used by ribosomes for recognition and binding. With a simple calculation based on triplet patterns so far unknown agreements between all of these sequences were found. In several cases it was shown that agreements between old sequences are easier recognizable if the purine- and pyrimidine bases are put into the triplets instead of the four A, G, C, and U(T) bases. In such cases "homologous" parts of sequences were recognized with more distinctness. This is true in our case for the double triplet (hexaplet) py-pu-pu-pu-pu-(pu) and the binding site triplet py-pu-pu, which are preceding the initiator. These triplets are in specific positions in all twelve sequences which were compared. The different course of the quaternary and the binary conformity curves (diagram 1) may show for the investigated area that the RNA phage gene-part is organized according to the well known quaternary triplet code. On the contrary the phage ϕ -gene-part seems to be organized according to a more simple, binary triplet sequence of purine and pyrimidine bases. The binary sequence seems to be the more original, the quaternary the derived one.

I. Einleitung

Nachdem die 2×3 Gene der RNA-Phagen Q β und R 17 in ihren Initiator-Regionen aufgrund einer statistischen Untersuchung [1] als homolog erkannt waren, schien die Möglichkeit nähergerückt, genauere Vorstellungen über die molekularen Strukturen der betreffenden Nukleotidstrecken zu gewinnen. Dabei stand der Gedanke im Vordergrund, daß die wahren Verhältnisse nur auf der Basis einer Triplet-einteilung zutage treten würden. Nach diesem Grundsatz wurde bei der Konstruktion eines Alignmentbündels verfahren, das sich aus den 6 Nukleotidstrecken der insgesamt 6 Gene der beiden Phagen zusammensetzt. Anstelle des Phagen R 17 wurde dabei der mit ihm nahezu identische Phage MS 2 verwendet, da über diesen zu Beginn der Untersuchung (1974) vollständigere Daten aus der Literatur vorlagen (zit. [2]).

Zur Konstruktion des Alignments wurden die vielen identischen oder teilidentischen Trinukleotide der Sequenzen als mutmaßliche Triplets in vertikale

„Kolumnen“ eingereiht. Durch ein einfaches, erstmalig angewandtes rechnerisches Verfahren wurde es möglich, die mittleren Ähnlichkeitsgrade je Tripletkolumne zu schätzen und aus den Schätzwerten eine Konformitätskurve zu erstellen. Dazu kam noch die unerwartete Erkenntnis, daß die den Kolumnen bzw. Sequenzen zugrundeliegenden Triplet-Strukturen in der Regel viel schärfer hervortreten, wenn anstelle des üblichen quaternären Schemas ein binäres aus Purinen und Pyrimidinen gewählt wird.

Bisher noch nicht veröffentlichte Vergleiche [3] mit den 6 Gen-Sequenzen der beiden RNA-Phagen bilden die Grundlage der nachstehenden Studie mit 12 verschiedenen, aus der Literatur bekannten Sequenzen. Dabei wurden die den Initiator-Triplets unmittelbar vorhergehenden, nicht Protein-codierenden Nukleotidstrecken zum Vergleich einander gegenübergestellt.

Es handelt sich dabei in Tab. I um die zu einem Alignment vereinigten 2×3 Gene der RNA-Phagen MS 2 + Q β , in Tab. II um die Konstruktion eines Alignment-Bündels aus den 4 (von insgesamt 9) ausgewählten Genen G, F, H, und D des einsträngigen DNA-Phagen ϕ X 174 [4] und schließlich in Tab. III um zwei zu einem einfachen Alignment zusammen-

Sonderdruckanforderungen an Dr. E. Köhler.

0341-0382/79/0900-0797 \$ 01.00/0



Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung“) beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition "no derivative works"). This is to allow reuse in the area of future scientific usage.

gefaßte Messenger, von denen der eine aus dem offenbar phagenfreien *Escherichia coli*, strain K, und der andere aus dem DNA-Phagen T 7 stammt [5]. Die Befunde werden anhand von drei Tabellen und drei Figuren beschrieben und kurz diskutiert.

II. Konstruktion eines Alignment-Bündels aus den insgesamt 6 vorcistronischen Genstrecken der RNA-Phagen MS 2 und Q β auf der Grundlage einer Triplet-Einteilung.

Aus den 6 Gen-Sequenzen der RNA-Phagen MS 2 und Q β wurden die den Initiator-Triplets unmittelbar vorhergehenden Oligonukleotide zu einem

Alignmentbündel zusammengefaßt, wobei ihre Trinukleotide (Triplets) in 11 vertikale Kolonnen eingeordnet wurden (Tab. I, Abschnitt A). Die Kolonnen wurden in der 5'-Richtung von -1 bis -10 durchnummeriert; die Initiator-Kolonne selbst erhielt die Nummer 0. Zunächst sollen die dem Alignment anhaftenden Unstimmigkeiten kurz betrachtet werden, d. h. also in erster Linie die erforderlich gewordenen Fehlstellen (Gaps): Kolonne -4 enthält außer leeren Tripleträumen nur ein einziges Triplet (UCA) in der Sequenz von Q β -C, das völlig isoliert dasteht und wohl kaum anders denn als jüngere Insertion gedeutet werden kann, zumal es auch in den Sequenzen der folgenden Tab. II und III fehlt. Lückenhaft ist

Tab. I. *Abschnitt A:* Alignment der zusammengefaßten 3×2 Initiator-Regionen der Gene A, C und R der RNA-Phagen MS 2 [2] und Q β [7-9]. Einteilung der Sequenzen nach Triplets, die ihrer Ähnlichkeit entsprechend in die Kolonnen -10 bis -1 eingereiht sind. *Abschnitt B:* Dasselbe Alignment wie bei A, jedoch mit dem Unterschied, daß die Purinbasen A und G als Kreise und die Pyrimidinbasen C und U als Kreuze eingetragen sind. Abweichende Basen oder -gruppen sind mit ovaler Umrandung versehen. *A und B:* In den Horizontalreihen 7-10 Berechnung der Konformitäts-Indices (KJ) der wahrscheinlichen Grundtriplets je Kolonne. Punkte in den Sequenzen bedeuten Fehlstellen („gaps“) in Basenpositionen, horizontale Striche bedeuten gaps von ganzen Tripleträumen, Schrägstriche in Q β -R stehen anstelle von anscheinend noch unbekannten Basen. Basen von Nonsens-Triplets in den Kolonnen -8, -6 und -2 durch Fettdruck hervorgehoben. Die Tabelle ist entnommen aus E. Köhler [3].

		Triplet-Kolonnen										
		-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
<i>Abschnitt A:</i>												
1	MS 2-A	5'GAA	UUC	CAU	UCC	UAG	GAG	---	GUU	UGA	CCU	<u>GUG</u>
2	Q β -A	CUU	CAC	UGA	GUA	UAA	GAG	---	---	GA	CAU	<u>AUG</u>
3	MS 2-C	GAG	CCC	UCA	ACC	---	GGA	---	GUU	UGA	AGC	<u>AUG</u>
4	Q β -C	AAG	CGU	UGA	AAC	UUU	GGG	UCA	AUU	UGA	U · C	<u>AUG</u>
5	MS 2-R	GCC	AUU	CAA	ACA	UGA	GGA	---	---	UUA	CCC	<u>AUG</u>
6	Q β -R	// G	CUU	AGU	AAC	UAA	GGA	---	---	UGA	UGC	<u>AUG</u>
	Hochzeilen-	abc	abc	abc	abc	abc				AA		
7	Summen	3+3+3	4+3+3	3+3+4	4+3+4	5+3+3	6+4+3		2+3+3	5+5+6	3+2+4	5+6+6
8	Grund-Seq.	GAG	CU ^U	(UGA)	ACC	UAA	GG ^A		GUU	UGA	C ^C C	<u>AUG</u>
9	Quotient	9/16	10/18	10/18	11/18	11/15	13/18		8/9	16/17	9/17	17/18
10	KJ	0,56	0,56	0,56	0,61	0,73	0,72		(0,89)	0,94	0,53	0,94
<i>Abschnitt B:</i>												
1	MS 2-A	OOO	XX ^X	XO ^X	XXX	XOO	OOO	---	OXX	XOO	XXX	<u>OXO</u>
2	Q β -A	XXX	XO ^X	XOO	OXO	XOO	OOO	---	---	·OO	XOX	<u>OXO</u>
3	MS 2-C	OOO	XXX	XXO	OXX	---	OOO	---	OXX	XOO	OOX	<u>OXO</u>
4	Q β -C	OOO	XO ^X	XOO	OXX	X ^{XX}	OOO	XXO	OXX	XOO	X · X	<u>OXO</u>
5	MS 2-R	OXX	OX ^X	XOO	OXO	XOO	OOO	---	---	XXO	XXX	<u>OXO</u>
6	Q β -R	// O	XX ^X	OOX	OXX	XOO	OOO	---	---	XOO	XOX	<u>OXO</u>
7	Summen	4+3+4	5+4+6	5+5+4	5+4+4	544	666		333	556	536	666
8	Grundsequenz	OOO	XXX	XOO	OXX	XOO	OOO		OXX	XOO	XOX	<u>OXO</u>
9	Quotient	11/16	15/18	14/18	13/18	13/15	18/18		9/9	16/12	14/17	18/18
10	KJ	0,69	0,83	0,78	0,72	0,87	1,00		(1,00)	0,94	0,82	1,00

auch Kolumne - 3, die neben drei offensichtlich homologen Triplets drei leere Tripleträume enthält, nämlich in den Gen-Sequenzen A und R von Q β , und in der Gen-Sequenz R von MS 2. Bemerkenswert ist, daß dieses Triplet (UCA) auch in einem der 4 ϕ -Gene von Tab. II zu finden ist und zwar in der Sequenz von Gen G. Man findet es an homologer Stelle (Kol. - 3) auch bei dem *Escherichia coli*-Messenger von Tab. III wieder. Sein unregelmäßiges Vorkommen bleibt einstweilen unverständlich. Ein leerer Tripletraum („Triple-Gap“) findet sich ferner in der MS 2-C-Sequenz der im übrigen vollständigen Kol. - 6; das Triplet ist augenscheinlich durch Deletion ausgeschieden. In dem gesamten Alignment mußten nur zwei Punkt gaps eingesetzt werden, nämlich in in Kol. - 2 bei Q β -A und in Kol. - 1 bei Q β -C. Das Adenin-Paar zwischen den Spalten 2 und 1 in der Replicase-Sequenz von Q β läßt sich einfach als „Bunch“, d. h. Ausstülpung in der Sekundärstruktur der Nukleotidsequenz deuten. (Tab. I ist einer noch unveröffentlichten, größeren Studie [3] entnommen, die auch den anschließenden Protein kodierenden Plusteil bis zur Kol. + 15 umfaßt.)

Die einzelnen Spalten enthalten offenbar Triplets von sehr unterschiedlichem Ähnlichkeitsgrad, darunter nicht wenige von Voll- oder Zweidrittelidentität. Aus den meisten Spalten lassen sich Triplets ableiten, die eine für die einzelne Spalte charakteristische Durchschnittskonformität aufweisen und im Folgenden als „Grundtriplets“ bezeichnet werden. Aus ihnen setzt sich die mutmaßliche „Grund-Sequenz“ (Horizontalreihe 10) zusammen. Für jedes Grundtriplet läßt sich ein Konformitätsindex (KJ) berechnen, wie die folgenden Beispiele zeigen sollen.

1. In Spalte - 8 ist in der ersten der drei Hochzeilen a, b, c die Base U dreimal, in der mittleren Hochzeile (b) die Base G gleichfalls dreimal und in der letzten (c) die Base A viermal die häufigste. Die Summe aus den drei Häufigkeiten $a + b + c = 10$ ergibt dividiert durch die Summe (= 18) der insgesamt besetzten Basenpositionen der Kol. - 8 den sehr niedrigen Konformitätsindex $10 : 18 = 0,56$. 2. In Spalte - 2 sind in den Hochzeilen a und b die Basen U und G beide mit dem Höchstwert 5 die häufigsten, in Zeile c sind sogar alle 6 Positionen von der Base A besetzt. Die Summe $a + b + c = 16$ ergibt, mit der Summe der insgesamt besetzten Positionen (= 17) dividiert, den hohen KJ = 0,94. 3. In Kol. - 5 ist in Zeile a die Base G sechsmal und in b

die Base G viermal die häufigste. In c sind die Basen G und A beide dreimal die häufigsten; der Wert 3 darf nur einmal gerechnet werden, da in jeder Hochzeile nur eine Basenhöchstzahl möglich ist. Die Summe der Häufigkeiten $6 + 4 + 3 = 13$ ergibt dividiert mit der Gesamtzahl der besetzten Basenpositionen den mittelhohen Konformitätsindex 0,72. In unserem Diagramm (Abb. 1) ist die ganze Sequenz der berechneten Indexwerte als Kurve I A eingetragen. Sie verläuft zunächst bei drei Spalten horizontal (KJ = 0,56), steigt dann ab Kol. - 7 (0,61) ziemlich stetig an und erreicht ihren ersten Gipfel in Kol. - 2 mit dem hohen Wert 0,94, den sie nach einem eindrucksvollen, sehr tiefen Absturz (auf 0,53) in Kol. - 1 in der Initiator-Spalte (0) wieder erreicht.

In Abschnitt B von Tab. I wird nur zwischen Purinbasen und Pyrimidinbasen unterschieden. Die Purinbasen (A und G) sind als Kreise, die Pyrimidinbasen (C und U) als Kreuze in die Triplet-Positionen eingetragen. Als vollkommen identisch erweist sich jetzt die reine Purin-Spalte - 5 (KJ = 1,00). Von vollkommener Identität weichen die Triplets von Kol. - 6 (KJ = 0,87) und von Kol. - 2 (KJ = 0,94) nur wenig ab. Sie zeichnen sich beide dadurch aus, daß sie überwiegend aus Terminatoren-Triplets bestehen, Kol. - 2 sogar ausschließlich aus solchen vom Typ UGA („opal“). Man kann die Spalten - 6 und - 5 auch als eine Einheit aus Hexaplets auffassen, die als solche eine bedeutende Rolle bei der Signalgebung für die Ribosomen spielen. Eine besonders starke Häufung von Pyrimidinen fällt in den Koln. - 9 und - 8 auf. Man könnte annehmen, daß diese Häufung oder doch das in allen Sequenzen vorhandene Triplet X-X-O in Vorzeiten gleichfalls als Erkennungssignal diente. Die aus den Indices der Grundsequenz konstruierte binär gefaßte Kurve hat auf unserem Diagramm die Bezeichnung I B erhalten.

III. Alignment-Bündel aus den vier Genen G, F, H und D des DNA-Phagen ϕ X 174 im gleichen Aufbau-Schema wie bei den RNA-Genen des vorhergehenden Kapitels

In dem oberen Teil (A) der Tab. II sind die Oligonukleotide der vier ϕ -Gene G, F, H und D [4] zu einem Alignment vereinigt, dem das gleiche Aufbau-Schema wie in Tab. I zugrunde gelegt ist. Zunächst fällt auf, daß die ganze Spalte - 4 leer ist, die in

Tab. II. Sammel-Alignment aus Trinukleotiden der vier Gene G, F, H und D des DNA-Phagen ϕ X 174 [4] im Schema von Tab. I mit den dort gegebenen Erläuterungen. Überlappung in Gen G zwischen den Koln. – 5 und – 3 von Teil B.

		Triplet-Kolumnen											
		- 10	- 9	- 8	- 7	- 6	- 5	- 4	- 3	- 2	- 1	0	
A Vierbasen-Sequenzen													
1	φ-Gene	G	AGG	TTT	TCT	GCT	T · A	GGA		GTT	TAA	· TC	ATG
2		F	CGG	CCC	CTT	ACT	TGA	GGA		---	TAA	ATT	ATG
3		H	CAG	CCA	CTT	AAG	TGA	GG ·		---	TGA	TTT	ATG
4		D	ACC	ACT	AAT	AGG	TAA	GAA		---	---	ATC	ATG
Hochzeilen		abc	abc	abc	etc.								
5	Grundsequenz	$\overset{A}{C}GG$	CCT	CTT	$AC\overset{G}{T}$	TGA	GGA			TAA	$AT\overset{C}{T}$	ATG	
6	{	Index-	2+2+3	2+3+2	2+2+4	3+2+2	4+2+4	4+3+3		3+2+3	2+4+2	4+4+4	
7		Berechnung	7/12	7/12	8/12	7/12	10/11	10/11		8/9	8/11	12/12	
8		KJ	0,58	0,58	0,67	0,58	0,91	0,91		0,89	0,73	1,00	
B Zweibasen-Sequenzen													
1	φ-Gene	G	OOO	XXX	XXX	OXX	XOO	OO ·		OXX	XOO	· XX	OXO
2		F	XOO	XXX	XXX	OXX	XOO	OOO		---	XOO	OXX	OXO
3		H	XOO	XXO	XXX	OOO	XOO	OO ·		---	XOO	XXX	OXO
4		D	OXX	OXX	OOX	OOO	XOO	OOO		---	---	OXX	OXO
5	Grundsequenz	$\overset{X}{O}OO$	XXX 0,83	XXX 0,83	$O\overset{XX}{O}O$	XOO	OOO			XOO	OXX	OXO	
6	{	wie oben	2+3+3	3+4+3	3+3+4	4+2+2	4+4+4	4+4+2		3+3+3	2+4+4	4+4+4	
7			8/12	20/24		8/12	12/12	10/10		9/9	10/11	12/12	
8			KJ	0,67	0,83		0,67	1,00	1,00		1,00	0,91	1,00
9	ohne D: 17/18 0,94												

Tab. I nur mit dem dort zum wahrscheinlichen Singulartriplet erklärten Triplet UCA besetzt ist. Ihre Nachbarkolumne – 3 enthält das einzige, aber reelle Triplet GTT, das mit dem Triplet GUU in Kol. – 3 von Tab. I „identisch“ ist, wie bereits zu Tab. I bemerkt wurde. Auch Kol. – 2 ist nicht vollständig, es fehlt ihr ein Triplet in Gen D, so daß hier im ganzen nur drei, vermutlich homologe Triplets vorhanden sind. Außer den genannten leeren Tripleträumen waren zwei Punktgaps bei der G-Sequenz in den Koln. – 6 und – 1 und ein drittes in Kol. – 5 bei der H-Sequenz einzusetzen.

Ein Überblick über die Sequenz der errechneten Konformitäts-Indices (Reihe 8) zeigt folgendes. Die Anfangswerte (Kol. – 10 und – 9) sind beide sehr niedrig (0,58). Der gleiche Wert findet sich auch wieder bei Kol. – 7. Dazwischen erscheint bei Kol. – 8 ein zwar kurzer aber deutlicher Anstieg. Bei den

Koln. – 6 und – 5 tritt übereinstimmend der sehr hohe Wert 0,91 auf. In die gleiche Größenordnung fällt auch der Wert 0,89 der Terminatoren-Kolumne – 2. Bei Kol. – 1 zeigt sich wieder ein vergleichsweise niedriger Wert (0,73). Die Initiator-Kol. 0 schließlich enthält 4 gleiche Triplets und erreicht damit als einzige den absoluten Höchstwert = 1,00.

Wenn man anstelle der vier Einzelbasen deren Purin- und Pyrimidinbasen in die Triplets einsetzt, erhält man die in Abschnitt B von Tab. II eingetragenen binären Tripletsequenzen. Durch dieselbe Berechnung, wie sie für Tab. I angegeben wurde, erhält man die in der unteren Horizontalreihe 8 verzeichneten Index-Werte. Sie beginnen in Kol. – 10 mit dem niedrigen Wert 0,67 und erreichen schon bei den beiden benachbarten Koln. – 9 und – 8 die ansehnliche Höhe von 0,83. In Kol. – 6, die sich aus verschiedenen Terminatoren zusammensetzt, wird

die völlige Identität aller Triplets und damit der höchstmögliche Wert 1,00 erreicht. Dasselbe trifft auch auf die Purinkolumne – 5 zu, wo die zwei unbesetzten Positionen (Punkt-Gaps) in die Berechnung nicht eingehen. Identität besteht auch bei der Terminatoren-Kolumne – 2, jedoch bleibt der ermittelte Höchstwert in Anbetracht der relativen Kürze der Kolumne unsicher. Ganz eindeutig ist die absolute Übereinstimmung in der Initiator-Kolumne 0. Auch in den Kolumnen – 9 und – 8 zeichnet sich überraschend völlige Übereinstimmung ab, wenn man lediglich die beiden einander sehr ähnlichen Gene F und G berücksichtigt. Es sind dann zwei ausschließlich aus Pyrimidinen bestehende Hexaplets festzustellen. Unter Berücksichtigung aller Gene bekommen die beiden Koln. den gleichhohen Index 0,83. Schließt man außerdem noch das Gen D aus, so steigt der Wert sogar auf 0,94. Es hat danach den Anschein, daß hier auch die Koln. – 9 und – 8 mit ihren Pyrimidinen als Signalgeber für die Ribosomen fungieren. Was Kol. – 6 anbetrifft, so mag ihr vielleicht eine aufschiebende Funktion zukommen, wie dies für die ähnlich gebaute Terminator-Kol. – 2 gleichfalls vermutet werden kann.

IV. Näherer Vergleich der oben durch Berechnung gewonnenen vier Konformitätskurven der 6 RNA-Phagen-Gene und der vier DNA- ϕ -Gene.

Im Folgenden werden die vier Konformitätskurven (Abb. 1) einer vergleichenden Betrachtung unterzogen:

Der Vergleich der beiden binären Sequenzen 1 B und 2 B läßt einen vorwiegend kongruenten Verlauf erkennen; bei beiden zeichnet sich deutlich ein Vorbau zwischen den Koln. – 10 und – 7 ab, der die Funktion eines vordersten Signalbereichs für die Ribosomen vermuten läßt. 1 B erreicht die oberste Grenze (das „Plateau“, $KJ = 1,00$) jedoch (bei Kol. – 5) später als 2 B und steigt auch früher ab. Schließlich verlaufen die beiden Kurven von Kol. – 2 bis zum Initiatortriplet wieder kongruent. Voneinander auffällig verschieden verhalten sich dagegen die Vierbasen-Kurven von 1 A und 2 A. Die ϕ -Kurve 2 A zeigt zwar den Vorbau bei Kol. – 8 noch deutlich, jedoch ist er in Kurve 1 A (RNA-Phagen) nicht einmal angedeutet. Während dann Kurve 2 A den Verlauf von Kurve 2 B bis zu Kol. – 2 in einigem Abstand kopiert (vergl. insbesondere den stei-

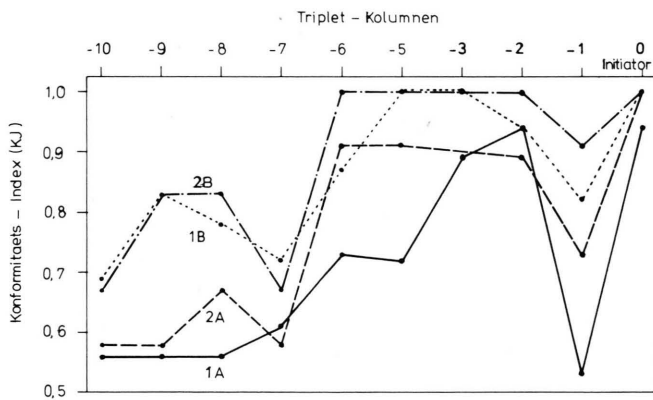


Abb. 1. Vergleich der Konformitätskurven der beiden Alignmente aus den RNA-Phagen MS 2 und Q β und den vier ϕ -Genen G, F, H und D':

1 A RNA-Abschnitt	} quaternär	(mit vier Basen)
2 A ϕ -Abschnitt		
1 B RNA-Abschnitt	} binär	(mit zwei Basen, Purine und Pyrimidine)
2 B ϕ -Abschnitt		

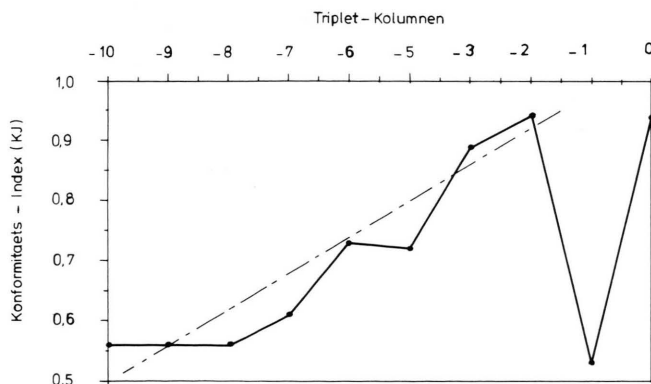


Abb. 2. Wiederholung der Kurve 1 A von Abb. 1 mit berechneter Regressionsgeraden, im Maßstab von Abb. 1.

Tab. III. Alignment aus den Vierbasen-Sequenzen (A) und den Zweibasen-Sequenzen (B) eines Messengers aus *Escherichia coli* (6) und eines Messengers aus dem DNA-Phagen T 7 [5]. Auf der nachstehenden Übersicht tritt bei B wieder die starke Zunahme der identischen Paarlinge und die starke Abnahme der differenten Paarlinge gegenüber A eindeutig in Erscheinung:

	A	B
identische Paarlinge	11	21
differente Paarlinge	11	3
fragliche Paarlinge	2	0
	24	24

	Triplet-Kolumnen									
	- 10	- 9	- 8	- 7	- 6	- 5	- 4	- 3	- 2	0
A. Vierbasen-Sequenzen										
1. <i>Escherichia coli</i> , messenger:	AUU	UUG	AAA	GCA	CG ^A _G	GG ^A _G	GAA	AUC	UG ^A _{AUG}	(6)
	Ile	Leu	Lys	Ala	Arg	Gly	Glu	Ile	opal Met	
2. Phage T 7, Gene 0,3, site b messenger:	 GUA	CGA	 GGA	 GGA	 UGA	 AGA	 G · ·	· · ·	 UA ^A _{AUG}	(9)
									ochre Met	
B. Zweibasen-Sequenzen										
3. <i>Escherichia coli</i> messenger:	OXX	XXO	OOO	OXO	XOO	OOO	OOO	OXX	XOOXO	
4. Phage T 7, etc. messenger:	 OXO	 XOO	 OOO	 OOO	 XOO	 OOO	 O · ·	· · ·	 XOOXO	

len Parallelanstieg von Kol. - 7 bis - 6), zeigt 1 A einen langsamen, fast stetigen Anstieg bis zu ihrer höchst-erreichbaren Spitze (0,94) in Kol. - 2, wo sie auf die absteigende Kurve 1 B trifft.

Weiterhin zeigt 1 A einen steilen und tiefen Abstieg auf 0,53 in Kol. - 1, um dann sogleich in steilem Anstieg die schon in Kol. - 2 erreichte Höhe (0,94) im Initiator wieder zu gewinnen. Kurve 2 A hingegen hat einen weniger tiefen Abstieg (nach 0,73) und erreicht danach den Index 1,00 der Initiator-Kolumne. Die quaternäre Kurve 1 A fällt demnach auf der Strecke von Kol. - 9 bis Kol. - 2 völlig aus dem Rahmen. Offenbar ist sie als einzige der 4 Sequenzen von allen vier Basen streng quaternär determiniert und folgt damit ihrer eigenen Gesetzmäßigkeit. Die Kurve 2 A (φ) gibt zu einer solchen Vermutung keinen Anlaß. Sie ist der binären Kurve 1 B praktisch kongruent, was darauf hindeutet, daß die φ -Kurven 2 A und 2 B im Zweibasenschema hinreichend determiniert und also auf eine Vierbasen-Sequenz vielleicht gar nicht angewiesen sind. Andererseits läßt der Umstand, daß die Kurve 1 B ein stark

verkürztes Plateau aufweist, die volle Funktionsfähigkeit der Signalstrecke von - 6 bis - 2 fraglich erscheinen. Damit steht Kurve 1 B offenbar nicht im Widerspruch zur Kurve 1 A, in der die Abkehr von der binären Sequentierung klar zum Ausdruck kommt. Kurve 1 A läßt sich befriedigend als lineare Regression darstellen*. Diese Feststellungen gelten zunächst nur für die vorcistronischen Oligonukleotide und nicht für den anschließenden Plusteil der Sequenzen in 3'-Richtung.

V. Vorschlag der Konstruktion eines einfachen Alignments zwischen zwei aus *Escherichia coli* stammenden Messengern unter Berücksichtigung der vorstehend gemachten Erfahrungen.

Ein unerwartetes Ergebnis hatte die Gegenüberstellung des von Platt und Yanofsky [6] analysierten

* Deren Berechnung (Diagramm Abb. 2) verdanke ich der Freundlichkeit von Herrn Dr. H. Rohloff.

Messengers aus *Escherichia coli* mit einem aus dem DNA-Phagen T 7, Gen 0,3 stammenden Messenger, dessen Sequenz von Steitz und Bryan [5] aufgestellt wurde. In Tab. III wird ein aus den entsprechenden Oligonukleotiden der beiden Messenger konstruiertes Alignment in unserem Kolumnensystem vorschlagen, oben (A) wieder in quaternärer und unten (B) in binärer Fassung. Die zweifellos homologe Struktur der beiden Sequenzen tritt erst in der binären Fassung klar zutage: Auf der Strecke von Kol. – 10 bis Kol. – 5 fallen von insgesamt 18 Basen nur 3 als nicht identisch aus dem Rahmen. Eindeutig tritt auch hier wieder die Identität der beiden Hexaplets X–O–O–O–O–O der Koln. – 6 und – 5 hervor. Bei dem Purintriplet in Kol. – 4 handelt es sich vermutlich um eine Zufalls-Insertion; sie könnte in der Vorgeschichte durch Duplikation aus dem identischen Nachbartriplet von Kol. – 5 entstanden sein. In Kol. – 3 ist beim Coli-Messenger (obere Sequenz) das unregelmäßig auftretende Triplet O–X–X wieder zu finden, das außer beim φ -Gen G auch bei den RNA-Genen MS 2-A, MS 2-C und Q β -C vorkommt. An seiner Stelle befindet sich in der Phagen T 7-Sequenz ein Triple-Gap.

Über die gesamte Strecke sind deutlich drei homologe Basengruppen verteilt: eine Gruppe aus 5 Purinen (O–O–O–O–O) im Bereich der Koln. – 9 bis – 7, eine andere aus 6 Basen (X–O–O–O–O–O) der Kolumnen – 6 und – 5, und eine dritte (X–O–O–X–O), die offenbar durch Überlappung der Kol. – 2 mit dem Initiator-Codon O entstanden ist (s. unten). Die homologen Gruppen sind auf Tab. III durch geschlängelte Linien eingegrenzt, sie entsprechen nach Lage und Inhalt gleichgelagerten Gruppen in den Sequenzen der Tabn. I B und II B. Den beiden Messengern fehlt die Kolumne – 1. Man kann aber in Analogie zu den Tabn. I und II vermuten, daß die beiden fehlenden Triplets ursprünglich die Kol. – 1 gebildet hatten und daß die von Platt und Yanofsky bei dem Coli-Messenger nachgewiesene Überlappung zwischen dem Terminortriplet UGA (Kol. – 2) und dem Initiator-AUG erst möglich wurde, nachdem das Triplet von Kol. – 1 aus der Sequenz ausgeschieden war. Der Schluß liegt nahe, daß das homologe Pentaplet X–O–O–X–O des Phagen T 7-Messengers ebenfalls auf diesem Wege entstanden ist, was vermutlich schon zu einem Zeitpunkt geschah, bevor die sehr ähnlich gebauten Messenger sich differenzierten.

Die Deletionen der Kolumnen-Triplets und die Überlappungen verursachten eine nicht unwesentliche Verkürzung der Signalstrecken.

Schlußbemerkungen

Nach den vorhergehenden Ausführungen hat es den Anschein, daß auch die nicht Protein kodierenden Initiatorregionen als ursprüngliche Tripletsequenzen zu verstehen sind und zwar mindestens in dem Kolumnen-Abschnitt von – 6 bis 0, wenn nicht schon von – 9 oder – 10 an.

Unser Befund, daß sich der Strukturabschnitt py-pu-pu-pu-pu (-pu) bei allen 12 untersuchten Initiator-Regionen – wenn auch nicht durchweg vollständig – wiederfindet, dürfte ziemlich eindeutig für seine funktionelle Bedeutung im Dienste der Ribosomenanbindung sprechen. Dieser Befund trifft sich mit der von Steitz und Jakes [7] postulierten Wasserstoff-Bindung, die zwischen dem Gen A des RNA-Phagen R 17 und der 3'-Endregion der ribosomalen 16 S-RNA anzunehmen ist. In unserer Schreibweise würde sich diese Bindung entsprechend der Zeichnung in Tab. IV darstellen. Da hiernach die erste Position U (Pfeil) des amber-Triplets – 6 nicht in die Bindung miteinbezogen ist, scheint dieses Triplet noch eine Sonderfunktion als Terminator auszuüben.

Die binäre Beschränkung auf die Pyrimidin- und Purinbasen bei der φ -Sequenz dürfte ziemlich eindeutig für die Vorstellung sprechen, daß hier für die Determinierung stereospezifische Wechselwirkungen maßgebend sind, die vermutlich auch die Aufrechterhaltung des starken selektiven Druckes ermöglichen, dem die gesamte Initiatorregion notwendigerweise unterworfen ist. Ein Nachlassen dieser Wechselwirkungen könnte die allmähliche Umstellung auf das quaternäre Schema bei den RNA-Phagen eingeleitet haben, wie dies in Kurve I B infolge des teilweisen Verlustes der Terminortriplets bei MS₂-C und Q β -C (Kol. – 6) zum Ausdruck

Tab. IV.

Kolumnen-Nr.:	– 7	– 6	– 5	– 3
A-Gen von Phage R 17:	5'UCC	UAG UC U A OH	GAG CUC	GUU CAC
3'-Ende der ribosomalen 16 S-RNA:				

kommt, ein Verlust, der möglicherweise durch das Tief von Kol. – 1 kompensiert wird.

Unverständlich ist noch die Lückenhaftigkeit im Auftreten des Triplets O–X–O in den Kolumnen – 3. Man könnte vermuten, daß das Triplet ursprünglich in allen Sequenzen vorhanden war, dann aber weitgehend abortierte und sich beim ϕ -Genom schließlich nur noch im Gen G gehalten hat.

Acknowledgements

Dem großen Entgegenkommen von Herrn Prof. M. Eigen (Göttingen) verdanke ich die gründliche Überprüfung des Manuskriptes in seinem Institut. Den Braunschweiger Kollegen Dr. B. Lerch und Dr. H. Rohloff danke ich für ihr lebhaftes Interesse an meiner Arbeit.

- [1] E. Köhler u. H. Rohloff, Z. Naturforsch. **29 c**, 433 – 436 (1974).
- [2] W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. van den Berghe, C. Volckaert u. M. Ysebaert, Nature **260**, 500 – 507 (1967).
- [3] E. Köhler (unveröffentlicht), Ableitung einer gemeinsamen auf Triplets aufgebauten Grundsequenz aus den 6 Initiator-Regionen der 2×3 Gene der RNA-Phagen MS 2 + Q β , und Vergleich der Grundsequenz mit der entsprechenden Region aus dem Spike-Gen G des DNA-Phagen ϕ X 174. Nachweis einer hohen Konformität zwischen Teilen der RNA-Grundsequenz und der ϕ -Sequenz.
- [4] F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchinson III, P. M. Slocombe u. M. Smith, Nature **265**, 687 – 695 (1977).
- [5] J. A. Steitz u. R. A. Bryan, J. Mol. Biol. **114**, 527 – 543 (1977).
- [6] T. Platt u. C. Yanofsky, Proc. Nat. Acad. Sci. USA **72**, 2399 – 2403 (1975).
- [7] J. A. Steitz u. K. Jakes, Proc. Nat. Acad. Sci. USA **72**, 4734 – 4738 (1975).
- [8] A. G. Porter, J. Hindley u. M. Billeter, Europ. J. Biochem. **41**, 413 – 420 (1974).
- [9] T. Maita u. W. Konigsberg, J. Biol. Chem. **246**, 5003 – 5024 (1971).
- [10] C. Bresch u. R. Hausmann, Klassische und Molekulare Genetik, 3. Aufl., Lange & Springer Verlag, Berlin 1972.
- [11] C. Weissmann, M. A. Billeter, H. M. Goodman, J. Hindley u. H. Weber, Structure and Function of Phage RNA, Ann. Rev. Biochem. **42**, 303 – 328 (1973).
- [12] J. A. Steitz, Ribosome Recognition of Initiator Regions in the Bacteriophage Genome. In RNA-Phages (ed. N. D. Zinder), Cold Springs Harbor Laboratory, 319 – 352 (1975).