



Doctoral Thesis

Comparative analysis of the thermodynamic RNA binding signatures of different types of RNA recognition motifs

Author(s):

Samatanga, Brighton

Publication Date:

2012

Permanent Link:

<https://doi.org/10.3929/ethz-a-007586428> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No 20603

**Comparative analysis of the thermodynamic RNA
binding signatures of different types of RNA
Recognition Motifs**

DISSERTATION

Submitted to

ETH ZURICH

for the degree of

DOCTOR OF SCIENCES

by

BRIGHTON SAMATANGA

MSc, Jacobs University Bremen

06 March 1985

Citizen of
ZIMBABWE

accepted on the recommendation of

Prof. Dr. Frédéric Allain
PD. Dr. Ilian Jelesarov
Prof. Dr. Rudolf Glockshuber
Prof. Dr. Ben Schuler

2012

Summary

Protein-RNA interactions are important for the regulation of several biological processes in the cell which include alternative splicing which occurs in about 98% of human genes. Alteration of splicing activity by proteins can cause diseases such as spinal muscular atrophy. RNA binding is regulated by various RNA binding proteins mostly through the RNA recognition motif (RRM). The RRM contains ~ 70 –90 amino acids consisting of two α helices packed against four β -strands. Canonical binding of the RNA to the β -sheet surface of the RRM is mediated by aromatic and basic residues in conserved RNP1 and RNP2 sequences. RRMs containing non-conserved RNP1/RNP2 sequences may exhibit structural versatility by binding to RNA using other surfaces which include α/β or β/β loops and α -helices. However, the energetics of this structural versatility remain poorly understood. Understanding these energetics helps explain regulation of complex stability and protein function. This study comparatively analyzes the thermodynamic RNA binding signatures of different types of RRMs exhibiting different binding modes. This is the first such comprehensive study. The RRMs considered are hnRNP A1 RRM1, hTra2 β 1 RRM, hnRNP F qRRM1 or qRRM3 and SRSF1 pseudo-RRM in which the RNA is bound to the β -sheet surface, β -sheet surface with involvement of the N- and C- terminal segments, the loops and α -helix regions respectively. The sequences and numbers of residues of cognate RNAs recognized by the RRMs are similar. 5'-UAGG-3', 5'-GAA-3', 5'-GGG-3' and 5'-GGA-3' RNA sequences are specifically recognized by hnRNP A1 RRM1, hTra2 β 1 RRM, hnRNP F qRRM1/3 and SRSF1 pseudo-RRM respectively.

In the first part, because the qRRMs of hnRNP F specifically recognize sequences of successive guanines (G-tracts) that spontaneously assemble into stable higher order conformations termed G-quadruplexes at moderate salt concentration, we investigated the effects of the G-quadruplex stability on the energetics and kinetics of qRRM-ssRNA association. To date, these effects have not been thoroughly investigated. Formation of G-quadruplexes both in vitro and in vivo is established, and effects on the regulation of events such as telomere biology and alternative splicing are well documented. To understand the thermodynamic and kinetic effects of the stability of G-quadruplexes on formation of the qRRM-ssRNA complex we studied by isothermal titration calorimetry (ITC), differential scanning calorimetry (DSC), circular dichroism (CD) and nuclear magnetic resonance spectroscopy (NMR) the binding equilibrium between qRRM3 and 5'-AGGGAU-3' and 5'-GGA-(GGGUUA)₄-3' G-tract RNAs under conditions that variably stabilized the G-quadruplex. An individual qRRM can mimic the function of hnRNP F. The 5'-AGGGAU-3' sequence was used in the solution NMR structural studies, whereas 5'-GGA-(GGGUUA)₄-3' resem-

bles telomeric RNA. The obtained results show that G-tract RNAs exist as an equilibrium between single-stranded RNA (ssRNA) and G-quadruplexes. qRRM3 binds to the ssRNA but not the G-quadruplex. This fact contradicts other authors. The intrinsic dissociation of the G-quadruplex is rate-limiting and slows complex formation. Consequently, qRRM3 binds to G-tract RNA co-transcriptionally preventing G-quadruplex formation.

To study the binding equilibrium in the absence of G-quadruplexes, the RNA sequence was modified by 7-deazaguanine (5'-AGG*G*AU-3' where G* is 7-deazaguanine). 5'-AGG*G*AU-3' and 5'-AGGGAU-3' are recognized in the same manner by the qRRMs. The binding equilibria between hnRNP A1 RRM1, hTra2 β 1 RRM, hnRNP F qRRM1 and SRSF1 pseudo-RRM and cognate RNAs were investigated by ITC and DSC experiments. qRRM1 was used because qRRM3 is not soluble in the same buffer as the other three proteins. Firstly, the thermodynamic binding parameters governing association and their temperature-dependence were determined. Secondly, the contribution of electrostatic effects to the overall free energy of association (ΔG_{obs}) was evaluated according to the counterion-condensation (CC) theory. Complex formation is driven by net negative enthalpy change and opposed by net negative entropy change throughout the physiological temperature range. The changes in enthalpy and entropy vary widely but compensate to give a ΔG_{obs} that is almost constant as function of temperature. Interestingly, the magnitude of ΔG_{obs} is the same in all considered complexes, which seems to be due to the dominance of intermolecular hydrogen bonds. The non-electrostatic component of ΔG_{obs} (ΔG_{nel}) contributes $\sim 75\%$ to complex stability. The remaining $\sim 25\%$ is contributed by the electrostatic component (ΔG_{ele}) due to formation of two ionic bonds between positively charged arginine side-chains and negatively charged phosphate oxygens (except in hnRNP F qRRM1-ssRNA). The arginine residues are always located in the β 2- β 3 loop and the N- or C-terminal segment. Furthermore, complex formation is accompanied by large negative heat capacity change of association ($\Delta C_{p,obs}$), which is characteristic of sequence-specific interactions. However, the magnitudes of $\Delta C_{p,obs}$ for the RRM-ssRNA complexes are different despite the recognition of the same number of RNA residues, except an additional uracil in the case of hnRNP A1 RRM1. Hence, the magnitude of $\Delta C_{p,obs}$ was dissected into its components. The combination of structural analysis, analyses of semi-empirical ΔC_p predictions and direct heat capacity measurements led to the following conclusions: (1) Complex formation proceeds with incomplete dehydration of the binding interfaces, (2) The side-chains and terminal segments of the RRMs undergo significant structural transitions, and (3) Non-parallel changes in thermal motions/dynamics experienced by the reaction partners and complex cause net increase in thermal motions (except in SRSF1). The net increase in thermal motions appears to correlate with the expected struc-

tural flexibility of the binding interfaces.

Zusammenfassung

Protein-RNA-Interaktionen regulieren verschiedene biologische Vorgänge der Zelle, darunter alternatives Spleißen. Mehr als 90-95% der menschlichen messenger RNA (mRNA) wird alternativ gespleißt. Durch Proteine veränderte Spleiß-Aktivität kann Krankheiten wie zum Beispiel spinale Muskelatrophie auslösen. Die RNA-Bindung ist durch verschiedene RNA-bindende Proteine geregelt und häufig durch das RNA Recognition Motif (RRM) vermittelt. Das RRM besteht aus ca. 70-90 Aminosäuren und enthält zwei α -Helices welche gegen ein viersträngiges β -Faltblatt gepackt sind. RRM's mit konservierten RNP1- und RNP2-Aminosäuresequenzen binden die RNA auf der Oberfläche des β -Faltblattes durch aromatische und basische Seitenketten der Aminosäuren. Andere RRM's, deren Aminosäureabfolge nicht der konservierten RNP1/RNP2-Sequenz entspricht, können strukturell vielfältiger RNA binden und andere Kontaktflächen, darunter Schleifen zwischen Sekundärstrukturelementen (α/β - oder β/β - Schleifen) und α -Helices, nutzen. Die energetischen Grundlagen dieser strukturellen Vielfältigkeit sind bisher jedoch wenig bekannt - ihr Verständnis würde helfen zu erklären, wie die Stabilität von RNA-Protein-Komplexen und die Proteinfunktion reguliert werden. Diese Doktorarbeit analysiert und vergleicht erstmals umfassend die thermodynamischen Merkmale der RNA-Bindung von vier verschiedenen RRM-Typen, welche jeweils unterschiedlich an RNA binden. Die untersuchten RRM-Domänen sind hnRNP A1 RRM1 und hTra2 β 1 RRM, welche die RNA über die β -Faltblätter binden, letztere zusätzlich auch unter Einbezug des N- und C-Terminus, sowie hnRNPF qRRM1 oder qRRM3, welche Schleifen zwischen Sekundärstrukturelementen zur Bindung benutzen, und SRSF1 pseudo-RRM, welches mit α -helicalen Regionen bindet. Die RNA-Sequenzen, welche durch die RRM's erkannt werden, sind ähnlich lang und purinreich: 5'-UAGG-3' wird erkannt durch hnRNP A1 RRM1, 5'-GAA-3' durch hTra2 β 1 RRM, 5'-GGG-3' durch hnRNP F qRRM1/3 und 5'-GGA-3' durch SRSF1 pseudo-RRM. Da die qRRM's von hnRNP F aufeinanderfolgende Guanine (G-tracts) erkennen, welche sich unter moderaten Salzbedingungen spontan zu Strukturen höherer Ordnung zusammenfügen können, sogenannten G-quadruplexen, untersuchten wir die Auswirkungen der G-quadruplex-Stabilität auf die energetischen und kinetischen Merkmale der Komplexbildung zwischen qRRM und einzelsträngiger RNA (ssRNA) mit isothermer Titrationskalorimetrie (ITC), dynamischer Differenzkalorimetrie (DSC), zirkularem Dichroismus Spektroskopie (CD) und Kernresonanzspektroskopie (NMR). Bisher wurden diese Effekte nicht vollständig erforscht, obwohl die Bildung von G-quadruplexen in vitro und in vivo etabliert und ihr Einfluss auf die Regulierung der Telomer-Aufrechterhaltung und des alternativen Spleißens belegt ist. Wir beobachteten das Bindungsgleichgewicht zwischen qRRM3 und 5'-AGGGAU-3' sowie 5'-GGA-(GGGUUA)₄-3' und das Gleichgewicht zwischen ssRNA und G-quadruplexen in Abwesenheit des

Proteins unter Bedingungen, welche den G-quadruplex unterschiedlich stark stabilisieren. Die 5'-AGGGAU-3' RNA-Sequenz wurde zuvor schon für die Bestimmung von NMR-Strukturen der qRRM-ssRNA-Komplexe benutzt, die 5'-GGA-(GGGUUA)₄-3' RNA-Sequenz gleicht Telomer-RNA. Wir modifizierten erstere RNA-Sequenz mit 7-deaza-Guanin (5'-AGG*G*AU-3'; wobei G* 7-deaza-Guanin bedeutet) um das Bindungsgleichgewicht in Abwesenheit von G-quadruplexen zu beobachten. Sowohl 5'-AGG*G*AU-3' als auch 5'-AGGGAU-3' werden genau gleich von qRRMs erkannt. Unsere Ergebnisse zeigen, dass G-tracts in Lösung im Gleichgewicht zwischen ssRNA und G-quadruplexen stehen und nur als ssRNA von qRRM3 gebunden werden können. Dies widerspricht den Ergebnissen anderer Autoren. Die intrinsische Dissoziation des G-quadruplexes ist geschwindigkeitsbestimmend und verlangsamt die Protein-ssRNA-Komplexbildung. qRRM3 bindet G-tract-RNA cotranscriptional und verhindert die G-quadruplex-Bildung.

Die Bindungsgleichgewichte von hnRNP A1 RRM1, hTra2 β 1 RRM, hnRNP F qRRM und SRSF1 pseudo-RRM mit den entsprechenden RNA-Sequenzen wurden mit ITC und DSC untersucht. Dabei wurde hnRNP F qRRM1 und nicht hnRNP F qRRM3 verwendet, da qRRM3 nicht unter denselben Pufferbedingungen wie die anderen Proteine löslich ist. Wir ermittelten die thermodynamischen Parameter, welche die RNA-Bindung bestimmen und ihre Temperaturabhängigkeit. Anschliessend wurde der Beitrag von elektrostatischen Effekten zur gesamten Freien Energie der Bindung (ΔG_{obs}) gemäss der Counterion-Condensation (CC) Theorie ausgewertet. Die Komplexbildung ist im gesamten Bereich physiologischer Temperaturen bedingt durch eine negative Gesamt-Enthalpie während eine negative Gesamt-Entropie entgegenwirkt. Die temperaturbedingten Änderungen in Enthalpie und Entropie unterscheiden sich stark, jedoch kompensieren sie sich, sodass eine beinahe temperaturunabhängige Freie Energie der Bindung die Folge ist. Der Wert von ΔG_{obs} ist interessanter Weise vergleichbar in allen untersuchten Komplexen, was teilweise durch die Dominanz einer ähnlichen Anzahl von zwischenmolekularen Wasserstoff-Brückenbindungen erklärt werden kann. Der nicht-elektrostatische Anteil an ΔG_{obs} beträgt ca. 75%, die elektrostatische Komponente (ΔG_{ele}) ca. 25%. Diese resultiert aus zwei ionischen Bindungen zwischen positiv geladenen Arginin-Seitenketten und negativ geladenen Phosphat-Sauerstoffatomen (ausser im hnRNP F qRRM1-ssRNA-Komplex). Die Arginine befinden sich in der β 2- β 3-Schleife und dem N- oder C-terminalen Bereich. Die Komplexbildung wird begleitet von einer großen negativen Änderung der Wärmekapazität ($\Delta C_{p,obs}$), welche für sequenzspezifische Interaktionen charakteristisch ist. Der Wert von $\Delta C_{p,obs}$ für die verschiedenen RRM-ssRNA-Komplexe ist jedoch unterschiedlich, obwohl eine vergleichbare Anzahl von RNA-Resten erkannt wird. Wir analysierten die einzelnen Bestandteile von $\Delta C_{p,obs}$. Die Kombination von Strukturanalysen mit semi-empirischen ΔC_p -Vorhersagemethoden

und direkter Wärmekapazitätsmessungen führt zu folgenden Ergebnissen: (1) Die Komplexbildung geht mit einer unvollständigen Dehydration der Bindestelle vonstatten, (2) die Seitenketten der RRMs vollführen wesentliche strukturelle Änderungen und (3) die Änderungen der thermischen Bewegung/Dynamik, welche die Reaktionspartner einzeln und im Komplex erfahren, sind ungleich und bewirken eine Gesamtzunahme thermischer Bewegung im Komplex (mit Ausnahme von SRSF1). Letzteres korreliert mit der erwarteten strukturellen Flexibilität der Bindungsstellen.