

Rozprawa doktorska p.t.:

**Nowe metody bioinformatyczne służące
do przewidywania miejsc wiązania jonów metali
i niskocząsteczkowych ligandów w strukturach RNA**

Anna Philips

Pracownia Bioinformatyki
Instytut Biologii Molekularnej i Biotechnologii
Uniwersytet im. Adama Mickiewicza
Poznań

Promotor: prof. dr hab. Janusz M. Bujnicki

Pracownia Bioinformatyki
Instytut Biologii Molekularnej i Biotechnologii
Uniwersytet im. Adama Mickiewicza
Poznań

Pracownia Bioinformatyki i Inżynierii Białka
Międzynarodowy Instytut Biologii Molekularnej i Komórkowej
Warszawa

Poznań 2013

Podziękowania

Dziękuję mojemu promotorowi prof. dr hab. Januszowi M. Bujnickiemu za kierowanie i opiekę nad moim projektem doktorskim oraz pomoc podczas pisania niniejszej pracy.

Składam serdeczne podziękowania za fachową, koleżeńską pomoc w sprawach naukowych i przy redagowaniu niniejszej rozprawy dr Michałowi Bonieckiemu, dr Grzegorzowi Łachowi, dr Joannie Kasprzak, dr Kristianowi Rotherowi, dr Irinie Tuszyńskiej i mgr Kai Milanowskiej.

Dziękuję również wszystkim koleżankom i kolegom z Wydziału Biologii UAM oraz z Międzynarodowego Instytutu Biologii Molekularnej i Komórkowej w Warszawie, za miłą atmosferę pracy i cenne dyskusje.

Pragnę także podziękować rodzinie za wsparcie, którego mi udziela na wybranej przeze mnie ścieżce kariery naukowej.

Finansowanie

Rozwój oprogramowania MetalionRNA oraz LigandRNA był finansowany przez:

- Narodowe Centrum Nauki (grant 2011/03/N/NZ2/01428 dla A.P.)
- Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu (grant GDWB-04/2011 dla A.P.)
- Poznańskie Centrum Superkomputerowo-Sieciowe (grant nr 143 dla A.P.)
- Fundację na rzecz Nauki Polskiej (grant TEAM/2009-4/2 dla J.M.B)

Publikacje wynikające z pracy

- **Philips A**, Milanowska K, Lach G, Bujnicki JM,
LigandRNA: computational predictor of RNA-ligand interactions.
RNA [artykuł w recenzji]
- **Philips A**, Milanowska K, Lach G, Boniecki M, Rother K, Bujnicki JM,
MetalionRNA: computational predictor of metal-binding sites in RNA structures.
Bioinformatics 2012 Jan 15;28(2):198-205. Epub 2011 Nov 21.

Spis treści

Podziękowania	3
Finansowanie.....	4
Publikacje wynikające z pracy	5
Spis treści	6
Wykaz skrótów.....	8
Streszczenie.....	11
1. Wstęp	12
1.1. Oddziaływania RNA z jonami metali.....	13
1.2. Oddziaływania RNA z ligandami.....	16
1.3. Modelowanie struktury kompleksów RNA z jonami metali i ligandami	27
2. Cel pracy	33
3. Materiały i metody.....	34
3.1. Potencjał statystyczny.....	34
3.2. Przygotowanie zbiorów uczących i testowych.....	37
3.3. Dokowanie ligandów do struktur RNA	44
3.4. Metody i miary użyte do ewaluacji programów	45
3.5. Narzędzia programistyczne	48
3.6. Wizualizacja wyników	48
3.7. Sprzęt komputerowy.....	48
4. Wyniki.....	49
4.1. Anizotropowy potencjał statystyczny.....	49
4.2. Program MetalionRNA.....	54
4.3. Program LigandRNA.....	64
5. Dyskusja.....	73
5.1. Omówienie programu MetalionRNA	73

5.2.	Omówienie programu LigandRNA.....	76
5.3.	Serwery MetalionRNA i LigandRNA.....	81
5.4.	Ograniczenia i perspektywy rozwoju.....	83
5.5.	Modelowanie struktury kompleksów RNA z ligandami a komputerowe projektowanie leków.....	88
6.	Wnioski.....	90
	Załączniki.....	91
	Spis rycin.....	92
	Spis tabel.....	95
	Bibliografia.....	97

Wykaz skrótów

5'-UTR	region niepodlegający translacji na końcu 5' (ang. <i>5'-UnTranslated Region</i>)
AdoCbl	adenozylkobalamina (ang. <i>adenosylcobalamin</i>)
AUC	pole powierzchni pod krzywą ROC (ang. <i>Area Under Curve</i>)
AutoDock	program do dokowania do białek (ang. <i>Automated Docking</i>)
BD	dynamika Browna (ang. <i>Brownian Dynamics</i>)
CASP	konkurs modelowania struktur białek (ang. <i>Critical Assessment of protein Structure Prediction</i>)
c-di-GMP	cykliczny diguanozynomonofosforan (ang. <i>Cyclic diGuanosine Monophosphate</i>)
CHARMM	pole siłowe, jak i pakiet do dynamiki molekularnej (ang. <i>Chemistry at HARvard Molecular Mechanics</i>)
Dock6	program do dokowania do białek i RNA
DNA	kwas deoksyrybonukleinowy (ang. <i>DeoxyriboNucleic Acid</i>)
DrugScore	funkcja oceny dla kompleksów białek z ligandami
DrugScoreRNA	funkcja oceny dla kompleksów RNA z ligandami
FEATURE	program do przewidywania miejsc wiązania jonów metali w białkach i RNA
FMN	mononukleotyd flawinowy (ang. <i>Flavin Mononucleotide</i>)
FPR	odsetek niepoprawnych przewidywań (ang. <i>False Positive Rate</i>)
GLcN6P	glukozamino-6-fosforan (ang. <i>GLucosamiNe-6-Phosphate</i>)
GOLD	program do dokowania do białek (ang. <i>Genetic Optimisation for Ligand Docking</i>)
glmS	rybozym aktywowany gloukozamino-6-fosforanem (ang. <i>glucosamine-6-phosphate activated ribozime</i>)
GPL	licencji wolnego i otwartego oprogramowania (ang. <i>General Public License</i>)
HDV	wirus zapalenia wątroby typu D (ang. <i>Hepatitis Delta Virus</i>)
HIV-1	wirus niedoboru odporności typu 1 (ang. <i>Human Immunodeficiency Virus type 1</i>)
LacZ	gen kodujący białko, β -glaktozydazę

Wykaz skrótów

L-HABA	kwask 2-hydroksy-4-aminomasłowy (ang. <i>2-Hydroxy-4-AminoButyric Acid</i>)
LigandRNA	program do oceny póż liganda w RNA (ang. <i>Ligand-RNA prediction</i>)
MetalionRNA	program do przewidywania miejsc wiązania jonów metali w RNA (ang. <i>Metal ion-RNA prediction</i>)
MORDOR	program do dokowania do RNA (ang. <i>MOlecular Recognition with a Driven dynamics OptimizeR</i>)
mRNA	matrycowy RNA (ang. <i>messenger RNA</i>)
ncRNA	niekodujący RNA (ang. <i>non-coding RNA</i>)
NMR	spektroskopia magnetycznego rezonansu jądrowego (ang. <i>Nuclear Magnetic Resonance</i>)
PB	Poisson-Boltzmann
PC1	2,5,6-triaminopyrimidyn-4-ol (ang. <i>2,5,6-triaminopyrimidine-4-one</i>)
PDB	baza danych zawierająca doświadczalnie rozwiązane struktury białek (ang. <i>Protein Data Bank</i>)
PreQ ₁	pre-kweozyna ₁ (ang. <i>Pre-Queuosine₁</i>)
PTC	centrum peptydylotransferazy (ang. <i>Peptidyl Transferase Center</i>)
PurR	regulon kodujący geny odpowiedzialne za biosyntezę puryn (ang. <i>Purine Regulon</i>)
rDock	patrz RiboDock
RiboDock	program do dokowania do RNA (ang. <i>Ribo-targets Dock</i>)
RMSD	średnie odchylenie standardowe (ang. <i>Root Mean Square Deviation</i>)
RNA	kwask rybonukleinowy (ang. <i>RiboNucleic Acid</i>)
RNaseP	rybonukleaza P (ang. <i>Ribonuclease P</i>)
ROC	krzywe charakterystyki odbiornika (ang. <i>Receiver Operating Characteristic</i>)
rRNA	rybosomowy RNA (ang. <i>ribosomal RNA</i>)
SAH	S-adenozylo-L-homocysteina (ang. <i>S-Adenosyl-L-homocysteine</i>)
SAM	S-Adenozyl-L-metionina (ang. <i>S-adenosyl-L-methionine</i>)
Surflex	program do dokowania do białek (ang. <i>Surface-based Flexible docking</i>)
TPP	pirofosforan tiaminy (ang. <i>Thiamine PyroPhosphate</i>)
TPR	odsetek poprawnych przewidywań (ang. <i>True Positive Rate</i>)

tRNA	transferowy RNA (ang. <i>transfer RNA</i>)
UCSF	Uniwersytet Kalifornijski, San Francisco (ang. <i>University of California, San Francisco</i>)
VS	wirtualna analiza przesiewowa (ang. <i>Virtual Screening</i>)
ZINC	baza danych ligandów do VS

Streszczenie

Głównym celem prezentowanej rozprawy doktorskiej było stworzenie nowych programów komputerowych służących do przewidywania oddziaływań RNA z małymi cząsteczkami (np. potencjalnymi lekami) i jonami metali.

Realizacja niniejszego projektu ma znaczenie dla postępu badań nad biologią RNA. RNA oraz jego funkcje są obecnie jednymi z najintensywniej badanych obszarów biologii molekularnej. Tak szeroki zakres prac wynika z niedawno odkrytego faktu, iż cząsteczki te są nie tylko nośnikami informacji genetycznej, ale biorą też udział w wielu kluczowych dla organizmu procesach, m.in. odkryto, że dzięki właściwościom katalitycznym i/lub regulatorowym mają ogromny wpływ na ekspresję informacji genetycznej.

W celu zrozumienia molekularnych mechanizmów działania RNA niezbędna jest wiedza o jego strukturze i oddziaływaniach z innymi cząsteczkami, gdyż to one determinują jego funkcje. W przypadku, gdy istnieją struktury krystaliczne możliwa jest analiza oddziaływań RNA z innymi związkami – zarówno z jonami, czy ligandami, ale także białkami i kwasami nukleinowymi. W przypadku, gdy dane te nie są dostępne, jedyną alternatywą jest skorzystanie z narzędzi bioinformatycznych. Konieczność podjęcia realizacji prezentowanego projektu jest motywowana brakiem wydajnych i w pełni automatycznych narzędzi tego typu.

Powstałe w realizacji prezentowanej pracy doktorskiej narzędzia umożliwiają modelowanie kompleksów struktur RNA z jonami metali i niskocząsteczkowymi ligandami. Ocena siły wiązania opiera się na potencjałach statystycznych wyprowadzonych z bazy danych znanych struktur. Podejście to zostało wcześniej z powodzeniem zastosowane przez inne grupy badawcze do modelowania kompleksów małych cząsteczek z białkami (np. w metodzie DrugScore). W ramach niniejszego projektu został stworzony potencjał anizotropowy bazujący na wzajemnych odległościach atomów RNA i jonu metalu/atomów liganda, jak i na ich wzajemnej orientacji w przestrzeni. W oparciu o ten potencjał opracowano oprogramowanie komputerowe MetalionRNA, które wyznacza prawdopodobne miejsca wiązania jonów metali w dowolnej cząsteczce RNA o znanej strukturze przestrzennej oraz oprogramowanie LigandRNA, które służy do przewidywania miejsc wiązania i relatywnej siły wiązania RNA z niskocząsteczkowymi ligandami.

1. Wstęp

Jeszcze do niedawna uważano, że nośnikiem informacji genetycznej w komórce jest kwas deoksyrybonukleinowy (DNA), natomiast białka pełnią funkcje enzymów. Kwas rybonukleinowy (RNA) był postrzegany jedynie jako pośrednik między tymi cząsteczkami. Dopiero odkrycie S. Altmana (Guerrier-Takada, Gardiner i wsp. 1983) i T. Cecha (Kruger, Grabowski i wsp. 1982) w latach 80-tych dwudziestego wieku diametralnie zmieniło ten pogląd. Dowiedli oni bowiem, że RNA może posiadać katalityczne właściwości, za co zostali uhonorowani nagrodą Nobla w dziedzinie chemii już w 1989 roku. RNA jest obecnie jedyną cząsteczką zaobserwowaną w naturze, która może zarówno przechowywać informację genetyczną (np. materiał genetyczny wirusów RNA zawarty jest w kwasie rybonukleinowym) jak i posiadać właściwości katalityczne (Westhof i Auffinger 2006).

Dziś już wiadomo, że RNA to nie tylko relikty pozostałe po tzn. hipotetycznym świecie RNA. Świadczą o tym ciągle odkrycia ważnych funkcjonalnie cząsteczek RNA, które mają wpływ na ekspresję informacji genetycznej, właściwości katalityczne czy regulatorowe (Pyle 2002, Grosshans i Filipowicz 2008). Obecnie znanych jest wiele klas cząsteczek RNA, w tym dobrze już poznane mRNA, tRNA i niezwykle istotne rRNA, a także wiele odkrytych niedawno i wciąż słabo poznanych niekodujących białka RNA (ncRNA, ang. *NonCoding RNA*). W celu zrozumienia ról pełnionych przez ncRNA niezbędna jest wiedza o jego strukturze i oddziaływaniach z innymi związkami, gdyż to one bezpośrednio determinują jego funkcje (Tinoco i Bustamante 1999, Eddy 2001). Wiele z poznanych ncRNA posiada dobrze określoną strukturę przestrzenną, stabilizowaną przez jony metali (DeRose 2003) (Rozdział 1.1 „Rola kationów w procesie zwijania się i stabilizacji struktur RNA”) (Holbrook 2008), gdyż przeprowadzenie reakcji enzymatycznych wymaga precyzyjnego ułożenia atomów w przestrzeni (Westhof i Auffinger 2006). To właśnie dzięki poznaniu struktury przestrzennej rybozymów i regulatorowych RNA (Hoogstraten i Sumita 2007) można było wnioskować o mechanizmach ich działania. Badania nad oddziaływaniami RNA z ligandami stwarzają ogromne perspektywy dla współczesnej medycyny (Erdmann, Barciszewski i wsp. 2006, Wheeler, Sobczak i wsp. 2009, Mulhbacher, St-Pierre i wsp. 2010). Obecnie w wielu ośrodkach na świecie prowadzone są prace nad ich wykorzystaniem w terapii chorób bakteryjnych, wirusowych, nowotworowych oraz zwyrodnieniowych. W każdym z wyżej wyniesionych przypadków uzyskanie

pożądanego efektu terapeutycznego będzie możliwe dzięki odpowiednio przygotowanym ligandom, oddziałującym w pożądany sposób z RNA. Dlatego też, tak ważne jest poznanie mechanizmów oddziaływania RNA z innymi cząsteczkami, a co za tym idzie, możliwość przewidywania miejsc ich wiązania. Doświadczalne badanie struktury RNA oraz jego oddziaływań z innymi związkami pozostaje procedurą czasochłonną i kosztowną, uzależnioną od specjalistycznego sprzętu. Alternatywą jest zastosowanie metod komputerowych. Niniejsza praca przedstawia program MetalionRNA, który pozwala na identyfikację miejsc wiązania się jonów magnezu, sodu i potasu w strukturach RNA oraz program LigandRNA służący do oceny i rankingu pól liganda/ów związanych z cząsteczkami RNA.

1.1. Oddziaływania RNA z jonami metali

Rola kationów w procesie zwijania się i stabilizacji struktur RNA

RNA jest liniowym polimerem złożonym z reszt rybonukleotydowych połączonych wiązaniami fosfodiesterowymi. Rybonukleotydy składają się z cukru rybozy, jednej z zasad azotowych (adeniny, guaniny, cytozyny lub uracylu) oraz kwasu fosforowego. Reszty nukleotydowe występujące w RNA tworzą skomplikowaną sieć oddziaływań, powodując, że RNA ma nie tylko strukturę pierwszorzędową (sekwencja), ale również może tworzyć strukturę drugo-, trzecio- i czwartorzędową. Ponieważ łańcuch główny RNA jest ujemnie naładowany neutralizacja elektrostatycznego odpychania się jego elementów poprzez wiązanie dodatnio naładowanych kationów jest niezwykle istotna, gdyż tylko wtedy RNA jest w stanie utworzyć struktury wyższego rzędu (w szczególności trzecio- i czwartorzędowe) (Heilman-Miller, Thirumalai i wsp. 2001, Woodson 2005, Travers, Boyd i wsp. 2007). Badania nad zwijaniem się RNA dowiodły, że stabilność struktur znacząco wzrasta w obecności jedno- (w szczególności Na^+ i K^+) (Holmstrom, Fiore i wsp. 2012) i dwuwartościowych (Mg^{2+}) kationów (Urbanke, Romer i wsp. 1975). Jednocześnie udowodniono, że dwuwartościowe kationy Mg^{2+} stabilizują strukturę RNA silniej niż jony jednowartościowe w tym samym stężeniu (Romer i Hach 1975, Stein i Crothers 1976, Tan i Chen 2010). Im wyższa była koncentracja ujemnego ładunku RNA na jego powierzchni, tym więcej kationów Mg^{2+} znajdowało się w tych miejscach. Zjawisko to wynika zarówno z faktu, że kationy magnezu są dwuwartościowe i ich mniejsza liczba

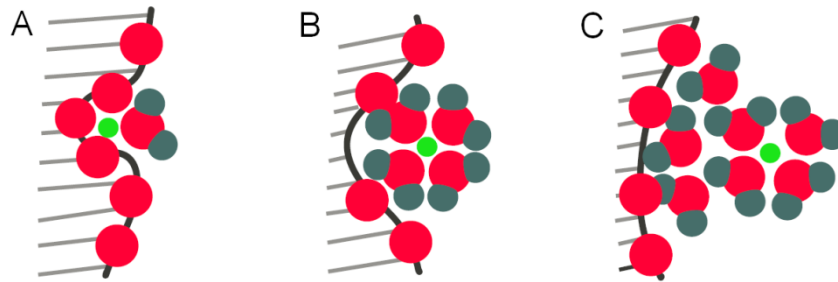
(niż kationów jednowartościowych) wystarcza do zbalansowania ujemnego ładunku RNA, jak i z faktu, że ich promienie są mniejsze (promień jonowy magnezu wynosi $\sim 0,72 \text{ \AA}$, natomiast sodu $\sim 1,02 \text{ \AA}$ i potasu $\sim 1,38 \text{ \AA}$), przez co mieści się ich więcej na ograniczonej powierzchni cząsteczki RNA (Petrov, Bowman i wsp. 2011). Z tych też powodów uważa się, że „silne” oddziaływania kationów Mg^{2+} odpowiedzialne są za stabilizację struktury przestrzennej RNA (Serra, Baird i wsp. 2002, Draper 2004, Draper 2008).

Charakterystyka oddziaływań RNA z Mg^{2+}

Formowanie się wiązania RNA z jonem metalu odbywa się w środowisku wodnym. Energia oddziaływania elektrostatycznego pomiędzy kationem a cząsteczką wody zależy zarówno od ładunku kationu, jak i jego wielkości. Mg^{2+} może utworzyć bezpośrednie wiązanie o osmiokątnej symetrii z sześcioma cząsteczkami wody, na które mogą się później nakładać wiązania z jej kolejnymi warstwami. Obliczenia teoretyczne wraz z analizami doświadczalnymi sugerują, że wartość zmiany entalpii swobodnej przy solwatacji wynosi $-455 \text{ kcal mol}^{-1}$ (Markham, Glusker i wsp. 2002). Dla porównania, wartości tej energii dla większego, jednowartościowego jonu K^+ , który bezpośrednio może związać osiem, bądź dziewięć cząsteczek wody, wynosi jedynie $-80 \text{ kcal mol}^{-1}$ (Draper, Grilley i wsp. 2005).

Wyróżnia się trzy typy wiązania kationów Mg^{2+} do RNA (Draper 2004, Draper, Grilley i wsp. 2005).

- Pierwszy typ to bardzo silne wiązania, gdy kationy są częściowo chelatowane przez ujemnie naładowane atomy tlenu z łańcucha głównego RNA, z którymi bezpośrednio oddziałują (patrz: Rycina 1, Panel A).
- Drugi typ wiązania wyróżnia się, gdy solwatowane kationy stabilnie wiążą się do RNA poprzez jedną warstwę cząsteczek wody (patrz: Rycina 1, Panel B).
- Trzeci typ to wiązanie Mg^{2+} do RNA poprzez więcej niż jedną warstwę wody. Ten typ wiązania jest najslabszy i najmniej stabilny. Kationy zmieniają swoje położenia przy powierzchni RNA, oddziałując z nim jedynie elektrostatycznie, bez bezpośredniego kontaktu i bez zaburzania układu cząstek wody związanych z jonem (patrz: Rycina 1, Panel C).



Rycina 1. Trzy typy wiązania się jonów Mg^{2+} z RNA. Na czerwono zaznaczono tlen, na szaro wodór, natomiast magnez zaznaczony został na zielono. Panel A) Mg^{2+} bezpośrednio związany z atomami RNA. Panel B) Mg^{2+} oddziałujący z RNA poprzez jedną warstwę wody. Panel C) Mg^{2+} oddziałujący z RNA poprzez więcej niż jedną warstwę wody. (Draper, Grilley i wsp. 2005).

Jony a aktywność biologiczna RNA

Jony metali nie tylko wspomagają proces związania się RNA i stabilizują jego strukturę. Udowodniono, że niekiedy obecność jonu jest niezbędna, by cząsteczka RNA mogła stać się katalitycznie aktywna (DeRose 2002, Schnabl i Sigel 2010). Do tego typu RNA katalitycznie aktywnych (rybozymów) należą:

- rybozym typu „głowy młotka” (ang. *hammerhead ribozyme*) (O’Rear, Wang i wsp. 2001);
- rybozym typu „spinki do włosów” (ang. *hairpin ribozyme*) (Walter i Burke 1998);
- rybozym wirusa zapalenia wątroby typu D (rybozym HDV) (Murray, Seyhan i wsp. 1998, Nakano, Chadalavada i wsp. 2000);

które są samo-wycinającymi się cząsteczkami RNA, występującymi głównie w prymitywnych wiroidach i wirusach (Ferre-D’Amare i Scott 2010). Na przykład, podczas replikacji genomu wirusa mozaiki tytoniowej, rybozym typu „głowy młotka” rozcina jego liniowe kopie, oraz prawdopodobnie odpowiada za ich dalszą ligację w formę kolistą (Scott, Martick i wsp. 2009).

Oraz:

- introny grupy I i II, które katalizują reakcję wycinania własnych intronów (Piccirilli, Vyle i wsp. 1993, Shan, Kravchuk i wsp. 2001, Li, Costa i wsp. 2011);
- rybonukleaza P (RNaseP), która bierze udział w dojrzewaniu końca 5’ tRNA u bakterii, archeowców czy eukariontów (Beebe, Kurz i wsp. 1996, Christian, Kaye i wsp. 2000, Reiter, Osterman i wsp. 2012);

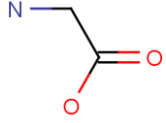
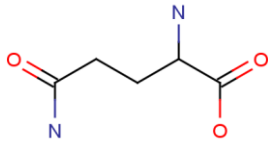
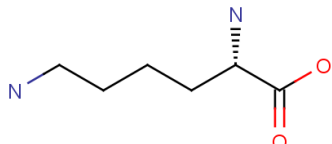
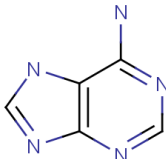
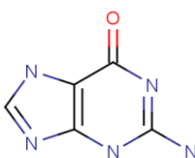
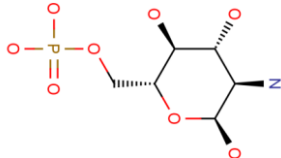
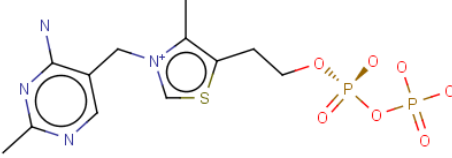
- ryboprzełącznik fluorkowy. Odpowiada on za ekspresję genów: odpowiedzialnych za transport jonów fluorkowych, enzymów wiążących się z fluorkami oraz kilku białek o nieznannej funkcji (Baker, Sudarsan i wsp. 2012). Ryboprzełączniki szerzej zostały opisane w Rozdziale 1.2 „Ryboprzełączniki”.

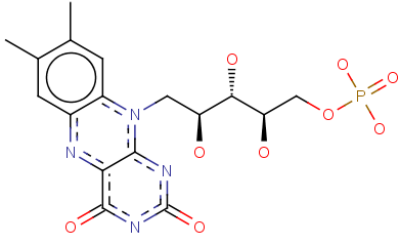
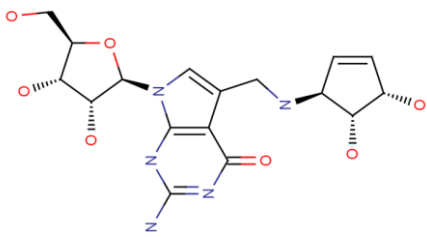
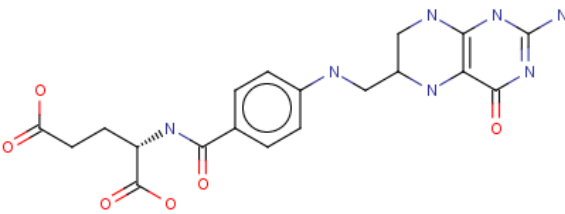
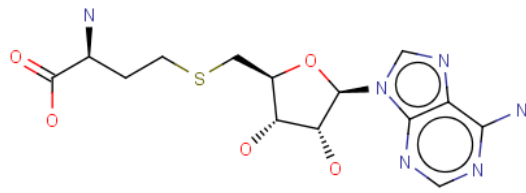
1.2. Oddziaływania RNA z ligandami

Funkcja wielu RNA zależy od ich oddziaływań z innymi cząsteczkami w komórce (Rivas i Eddy 2001, Thomas i Hergenrother 2008, Fulle i Gohlke 2010, Dieterich i Stadler 2013). W szczególności, aktywność biologiczna wielu niekodujących RNA jest regulowana poprzez przyłączanie się do nich małych związków, niskocząsteczkowych ligandów. Poniżej przedstawiono znane cząsteczki ncRNA tego rodzaju.

Ryboprzełączniki

Ryboprzełączniki to elementy regulatorowe występujące najczęściej w regionach niepodlegających translacji na końcu 5' (5'-UTR) bakteryjnego mRNA. Znane są różne mechanizmy działania ryboprzełączników, na przykład poprzez alternatywne składanie, jednak najwięcej z nich reguluje ekspresję informacji genetycznej w odpowiedzi na przyłączenie się liganda (Deigan i Ferre-D'Amare 2011). Ryboprzełączniki tego typu składają się z dwóch domen: aptameru i platformy ekspresyjnej. Aptamer bezpośrednio wiąże się z ligandem, co wpływa na zmianę struktury przestrzennej platformy ekspresyjnej, a przez to na ekspresję zakodowanego genu (Garst, Edwards i wsp. 2011). W większości przypadków przyłączenie się liganda blokuje ekspresję regionu kodującego, ale znane są też przypadki, gdy obecność liganda powoduje jego indukcję (Montange i Batey 2008). Ligandy wiążące się z ryboprzełącznikami to szeroki wachlarz związków, począwszy od wspomnianych wcześniej jonów metali (Rozdział 1.1 „Jony a aktywność biologiczna RNA”) aż po złożone metabolity (Deigan i Ferre-D'Amare 2011). Tabela 1 prezentuje znane ligandy oddziałujące z ryboprzełącznikami oraz ich charakterystykę.

Nazwa i struktura chemiczna liganda	Funkcja ryboprzełącznika
<p>Glicyna (identyfikator PDB: GLY)</p> 	<p>Ryboprzełącznik glicynowy reguluje syntezę i transport glicyny (Mandal, Lee i wsp. 2004).</p>
<p>Glutamina (identyfikator PDB: GLN)</p> 	<p>Ryboprzełącznik glutaminowy reguluje metabolizm glutaminy i azotu oraz krótkich peptydów o nieznannej funkcji (Ames i Breaker 2011).</p>
<p>Lizyna (identyfikator PDB: LYS)</p> 	<p>Ryboprzełącznik lizynowy kontroluje syntezę i transport lizyny (Grundy, Lehman i wsp. 2003, Rodionov, Vitreschak i wsp. 2003, Sudarsan, Wickiser i wsp. 2003).</p>
<p>Adenina (identyfikator PDB: ADE)</p> 	<p>Ryboprzełącznik adeninowy reguluje syntezę i transport puryn (kontroluje ekspresję regulonu <i>PurR</i>) (Johansen, Nygaard i wsp. 2003, Mandal i Breaker 2004).</p>
<p>Guanina (identyfikator PDB: GUN)</p> 	<p>Ryboprzełącznik guaninowy reguluje syntezę i transport puryn (kontroluje ekspresję regulonu <i>PurR</i>) (Johansen, Nygaard i wsp. 2003, Mandal, Boese i wsp. 2003).</p>
<p>Glukozamino-6-fosforan (GLcN6P, identyfikator PDB: GLP)</p> 	<p>Ryboprzełącznik <i>glmS</i> reguluje syntezę GLcN6P, który jest podstawowym składnikiem ściany komórkowej bakterii (Hampel i Tinsley 2006, Klein i Ferre-D'Amare 2006, Cochrane, Lipchock i wsp. 2007).</p>
<p>Pirofosforan tiaminy (identyfikator PDB: TPP)</p> 	<p>Ryboprzełącznik TPP reguluje syntezę i transport tiaminy, jak i podobnych metabolitów (Mironov, Gusarov i wsp. 2002, Rodionov, Vitreschak i wsp. 2002).</p>

<p>Mononukleotyd flawinowy (identyfikator PDB: FMN)</p> 	<p>Ryboprzełącznik FMN odpowiedzialny za ekspresję genu <i>LacZ</i>, reguluje syntezę i transport ryboflawiny (wit. B₂)(Mironov, Gusarov i wsp. 2002, Winkler, Cohen-Chalamish i wsp. 2002).</p>
<p>Pre-kweozyna₁ (PreQ₁, identyfikator PDB: PRF)</p> 	<p>Ryboprzełącznik PreQ₁ reguluje syntezę i transport pre-kweozyny₁ (Roth, Winkler i wsp. 2007).</p>
<p>Tetrahydrofolat (identyfikator PDB: THL)</p> 	<p>Ryboprzełącznik Tetrahydrofolat reguluje syntezę i transport tetrahydrofolatu (Tausch, Ceres i wsp. 2011).</p>
<p>S-adenozyl-L-homocysteina (identyfikator PDB: SAH)</p> 	<p>Ryboprzełącznik SAH reguluje metabolizm S-adenozyl-L-homocysteiny (Edwards, Reyes i wsp. 2010, Batey 2011).</p>

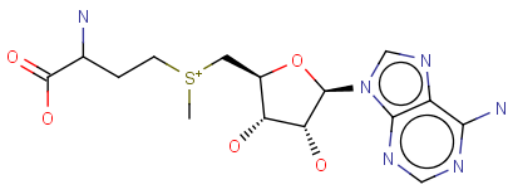
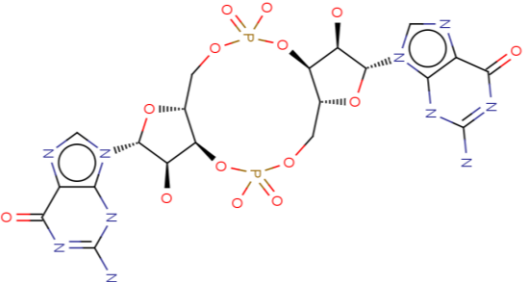
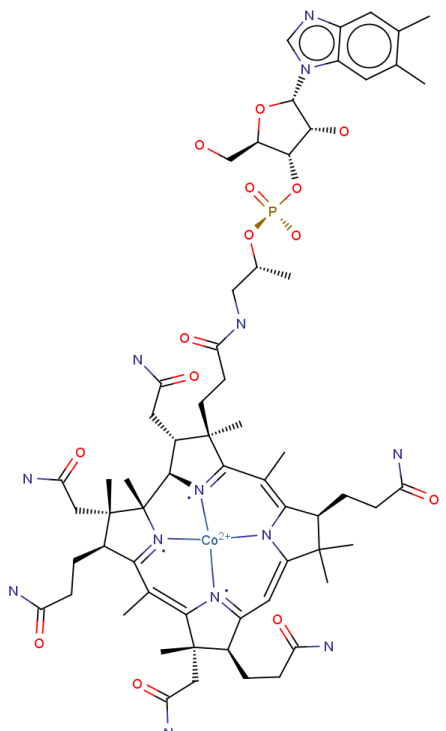
<p>S-Adenozylo-L-metionina (identyfikator PDB: SAM)</p> 	<p>Ryboprzełącznik SAM reguluje syntezę i transport metioniny (Epshtein, Mironov i wsp. 2003, Winkler, Nahvi i wsp. 2003, Corbino, Barrick i wsp. 2005, Fuchs, Grundy i wsp. 2006)</p>
<p>Cykliczny diguanozynomonofosforan (c-di-GMP, identyfikator PDB: C2E)</p> 	<p>Ryboprzełącznik c-di-GMP reguluje ekspresję genów kontrolowanych przez sygnałną cząsteczkę c-di-GMP (Zhou, Sayre i wsp. 2012)</p>
<p>Adenozylokobalamina (AdoCbl, identyfikator PDB: B1Z)</p> 	<p>Ryboprzełącznik AdoCbl kontroluje metabolizm kobalaminy (wit. B₁₂) (Nahvi, Sudarsan i wsp. 2002, Borovok, Gorovitz i wsp. 2006, Warner, Savvi i wsp. 2007)</p>

Tabela 1. Znane ligandy oddziałujące z ryboprzełącznikami oraz ich charakterystyka.

Rybo przelaczniki wystepuja znacznie czesciej u bakterii niz w komorkach eukariotycznych (Auslander, Ketzer i wsp. 2010). Jest to rezultatem czestego narazenia bakterii na zmiany srodowiska, w wyniku czego musialy one wyksztalcić mechanizmy umożliwiające im szybkie reagowanie na nowe warunki (Gripenland, Netterling i wsp. 2010). Za ich wysoką zdolność adaptacji odpowiadają między innymi właśnie rybo przelaczniki, które, niczym sensory, sterują ekspresją odpowiednich genów w odpowiedzi na zmiany w otoczeniu (pojawienie się bądź brak danego metabolitu). Jest to szczególnie istotne dla szczepów patogennych, które stają się wirulentne w odpowiedzi na sygnał gospodarza, aby obronić się przed jego systemem odporności (Geissmann, Possedko i wsp. 2006). Przykładowo rybo przelacznik guaninowy jest kluczowy dla zjadliwości gronkowca złocistego (*S. aureus*) w komórkach ssaków (Mulhbacher, St-Pierre i wsp. 2010). Powyższe fakty sprawiają, iż rybo przelaczniki są potencjalnym celem działania leków antybakteryjnych, w szczególności na szczepy bakterii opornych na antybiotyki. Obecnie prowadzone są badania nad wykorzystaniem rybo przelacznika lizynowego, guaninowego, TPP czy FMN (Mulhbacher, Brouillette i wsp. 2010) w takiej terapii.

Rybosomowy RNA

Rybosomowy RNA (rRNA) nie występuje i nie funkcjonuje niezależnie, ale wraz z ponad pięćdziesięcioma różnymi białkami stanowi integralną część rybosomu. Rybosom, nazywany też komórkową maszyną syntezy białek, jest największym i najbardziej złożonym enzymem o całkowitej masie ponad dwóch milionów Daltonów (Poehlsaard i Douthwaite 2005). To za jego pośrednictwem odbywa się proces translacji genów na białka, który jest kluczowy dla wszystkich żyjących komórek. Rybosomy występują zarówno w komórkach eukariotycznych, jak i prokariotycznych, jednak obserwuje się różnice w ich budowie. To właśnie dzięki nim istnieje możliwość zastosowania antybiotyków blokujących translację u bakterii bez ryzyka, że zablokują one translację w rybosomie gospodarza (Poehlsaard i Douthwaite 2005).

Rybosom u prokariotów składa się z dwóch podjednostek: 30S oraz 50S, o różnych współczynnikach sedymentacji. Mniejsza podjednostka (30S) złożona jest z 16S rRNA oraz około 20 białek (oznaczanych S1-S20). Z kolei duża podjednostka (50S) składa się z dwóch cząsteczek rRNA: 5S oraz 23S, a także z ponad 30 białek (L1-L36). Rybosomowe białka w znacznej mierze pełnią jedynie role strukturalne.

Wstęp

Za funkcje rybosomu, a przez to oddziaływania z innymi cząsteczkami odpowiadają komponenty zbudowane z RNA (Douthwaite, Fourmy i wsp. 2005). To RNA tworzy struktury: centrum peptydylotransferazy (ang. *Peptidyl Transferase Center*, PTC) w dużej podjednostce czy centrum dekodującego (ang. *decoding center*) w małej podjednostce, które ze względu na swoje kluczowe role są celem wielu antybiotyków (Bottger 2006, Foloppe, Matassova i wsp. 2006, Dunkle, Xiong i wsp. 2010). Poniżej przedstawiono znane klasy antybiotyków wiążące się z rRNA.

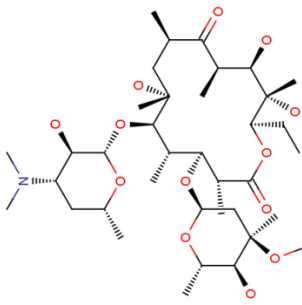
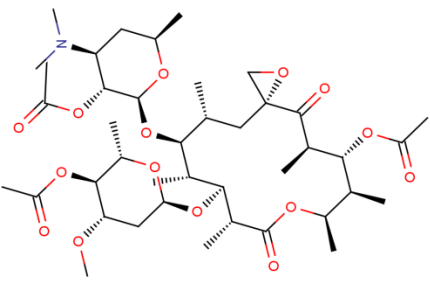
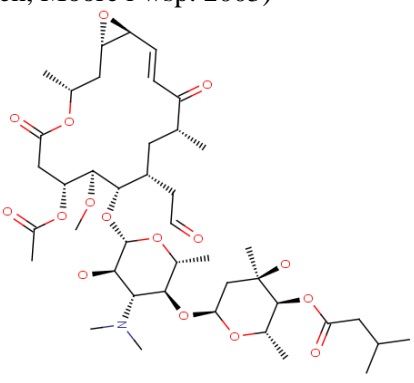
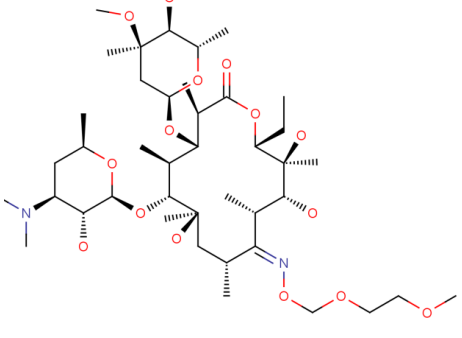
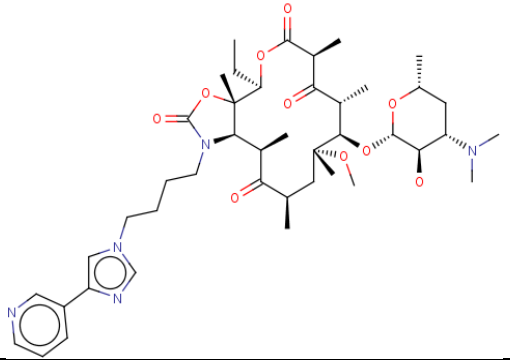
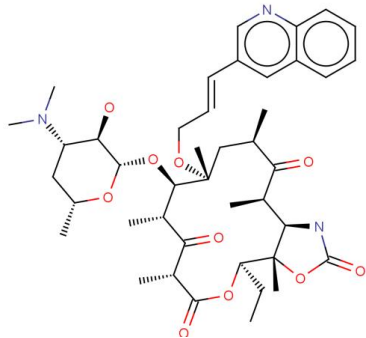
W PTC w dużej podjednostce rybosomu.

- Makrolidy (erytromycyna, azytromycyna (Hansen, Ippolito i wsp. 2002), klarytromycyna (Schlunzen, Zarivach i wsp. 2001), tylozyna (Ban, Nissen i wsp. 1999), spiramycyna (Poulsen, Kofoed i wsp. 2000), karbomycyna A, roksytromycyna, troleandomycyna) – antybiotyki te powodują przedwczesne zakończenie syntezy łańcucha peptydowego (zahamowanie translokacji peptydylotransferazy) poprzez blokadę przemieszczania się powstającego peptydu. W konsekwencji prowadzi to do odłączenia się peptydylotRNA od rybosomu (Ban, Nissen i wsp. 1999, Poulsen, Kofoed i wsp. 2000, Hermann 2005).
- Ketolidy (telytromycyna, cetrymocylna) – to semi-syntetyczna grupa antybiotyków, stworzona na podstawie makrolidu – erytromycyny. Zamiast bocznego pierścienia aromatycznego zawierają one grupę ketonową przy węglu C3, a pozycje 11 oraz 12 połączone zostały w pierścień heterocykliczny. Leki te oddziałują w podobny sposób jak ich pierwowzory, lecz wykazują znacznie większą skuteczność, szczególnie pod względem zwalczania opornych szczepów bakterii (Wilson, Harms i wsp. 2005).
- Linkozamidy (klindamycyna) – wiążą się do miejsca PTC na dnie tunelu wzdłuż jego osi, uniemożliwiając w ten sposób związanie się substratu (Douthwaite 1992).
- Streptogramina A i B (A: wirginiamycyna M, dalfoprystyna; B: chinuprystyna), antybiotyki te blokują formowanie się wiązania peptydowego, wiążąc się na początku tunelu PTC (Porse i Garrett 1999).
- Pleuromutyliny (tiamulina, walnemulina) – blokują powstawanie wiązania peptydowego bezpośrednio przez oddziaływanie ze substratem, wiążącym się

zarówno do miejsca akceptorowego jak i donorowego (Schlunzen, Pyetan i wsp. 2004).

- Inne: anizomycyna, chloramfenikol, blastycydyna S, sparsomycyna – blokują powstawanie wiązania peptydowego (Hansen, Moore i wsp. 2003).

Tabela 2 przedstawia wybrane struktury antybiotyków z wyżej wymienionych klas. Rycina 2 przedstawia przykładowy antybiotyk – erytromycynę związany z rRNA w PTC dużej podjednostki rybosomu.

Makrolidy	
<p>Erytromycyna (identyfikator PDB: ERY) (Schlunzen, Zarivach i wsp. 2001)</p> 	<p>Troleandomycyna (identyfikator PDB: TAO) (Berisio, Schlunzen i wsp. 2003)</p> 
<p>Karbomycyna A (identyfikator PDB: CAI) (Hansen, Moore i wsp. 2003)</p> 	<p>Roksytromycyna (identyfikator DPB: ROX) (Schlunzen, Zarivach i wsp. 2001)</p> 
Ketolidy	
<p>Telytromycyna (identyfikator PDB: TEL) (Berisio, Harms i wsp. 2003)</p> 	<p>Cetromycyna (identyfikator PDB: 773) (Schlunzen, Harms i wsp. 2003)</p> 

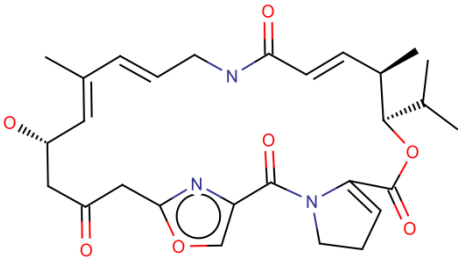
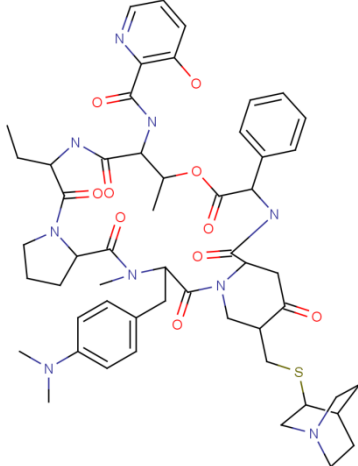
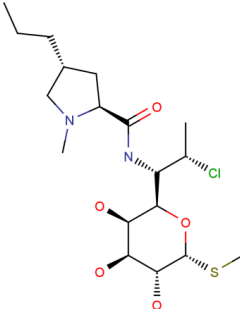
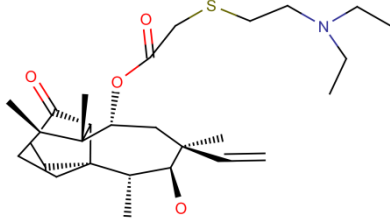
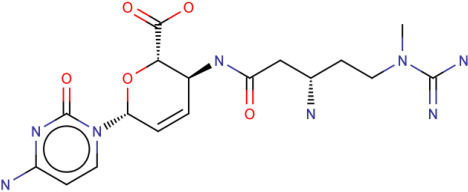
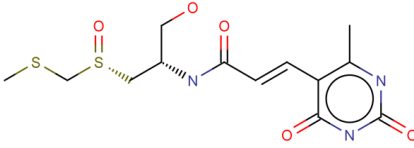
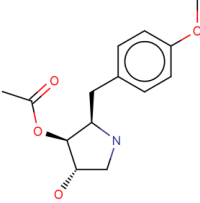
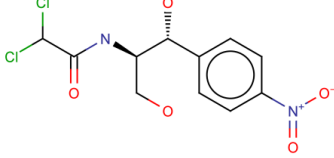
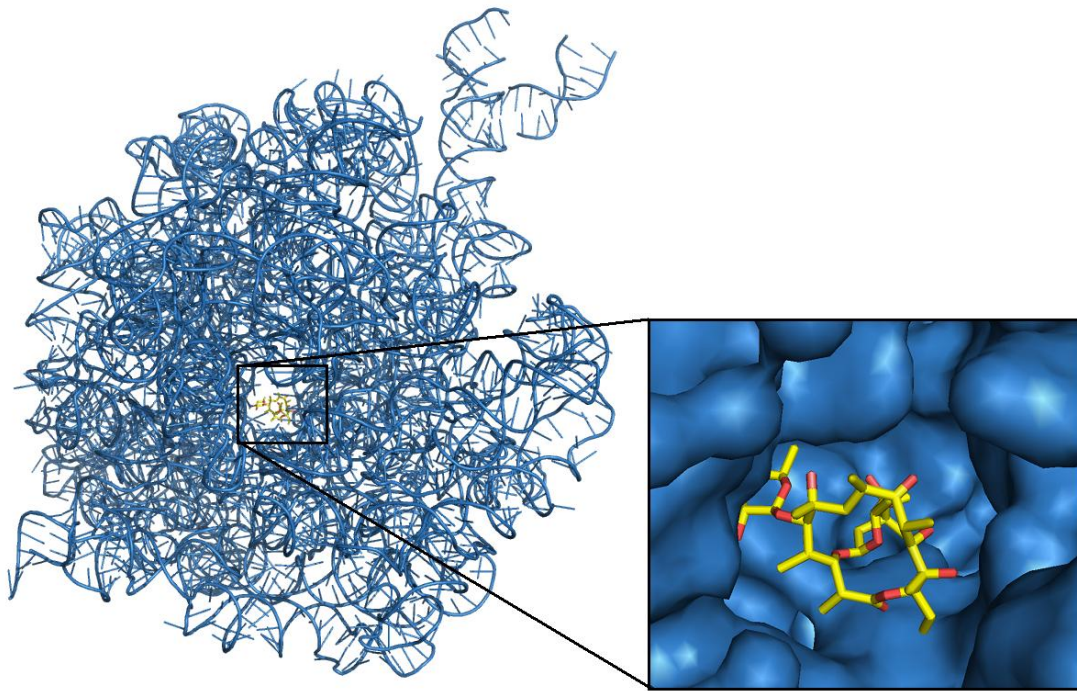
Streptogramina A	Streptogramina B
<p>Wirginiamycyna M (identyfikator PDB: VIR) (Hansen, Moore i wsp. 2003, Tu, Blaha i wsp. 2005)</p> 	<p>Chinuprystyna (identyfikator PDB: ---) (Harms, Schlunzen i wsp. 2004, Tu, Blaha i wsp. 2005)</p> 
Linkozamid	Pleuromutylna
<p>Klindamycyna (identyfikator PDB: CLY) (Douthwaite 1992, Porse i Garrett)</p> 	<p>Tiamulina (identyfikator PDB: MUL) (Schlunzen, Pyetan i wsp. 2004)</p> 
Inne (Hansen, Moore i wsp. 2003)	
<p>Blastycydyna S (identyfikator PDB: BLS)</p> 	<p>Sparsomycyna (identyfikator PDB: SPS)</p> 
<p>Anizomycyna (identyfikator PDB: ANM)</p> 	<p>Chloramfenikol (identyfikator PDB: CLM)</p> 

Tabela 2. Wybrane antybiotyki wiążące się z rRNA w dużej podjednostce rybosomu.



Rycina 2. Erytromycyna związana w dużej podjednostce rybosomu *Deinococcus* (kod PDB: 1JZY).

W małej podjednostce rybosomu.

- Aminoglikozydy (streptomycyna, higromycyna B (Brodersen, Clemons i wsp. 2000), paromycyna (Vicens i Westhof 2001), tobramycyna (Vicens i Westhof 2002), genetycyna (Vicens i Westhof 2003), apramycyna (Han, Zhao i wsp. 2005)) – oddziałują z centrum dekodującym zajmując miejsce A (akceptor aminoacylowy) i w ten sposób zakłócając interakcję kodonu (w mRNA) z antykodonom. Prowadzi to do zaburzenia odczytu informacji genetycznej i zahamowania syntezy białek (Carter, Clemons i wsp. 2000, Poehlsgaard i Douthwaite 2005, Wilson, Harms i wsp. 2005).
- Tetracykliny – blokują wiązanie się aminoacylo-tRNA do kompleksu mRNA z rybosomem (Brodersen, Clemons i wsp. 2000, Pioletti, Schlunzen i wsp. 2001).
- Oksazolidynony (linezolid) – blokują translację poprzez inhibicję tworzenia się kompleksu inicjującego. To pierwsze chemicznie zsyntetyzowane antybiotyki anty-rybosomowe (Patel, Yan i wsp. 2001).
- Inne: spektynomycyna (Carter, Clemons i wsp. 2000); paktamycyna (Brodersen, Clemons i wsp. 2000); edeina (Pioletti, Schlunzen i wsp. 2001).

Wstęp

Tabela 3 przedstawia wybrane struktury antybiotyków z wyżej wymienionych klas. Rycina 3 przedstawia przykładowy antybiotyk – hygromycynę B związany z rRNA centrum dekodującego w małej podjednostce rybosomu.

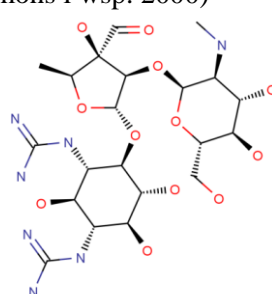
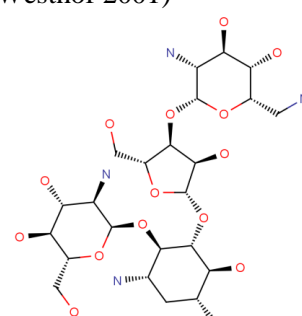
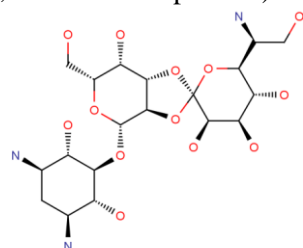
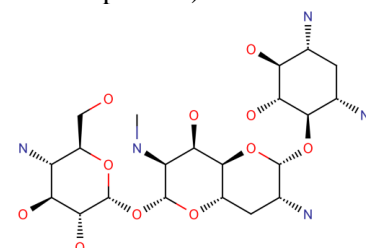
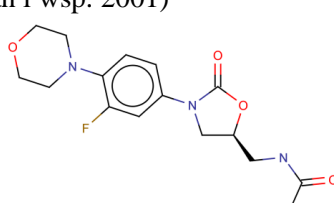
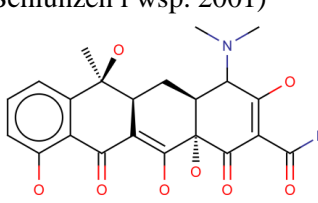
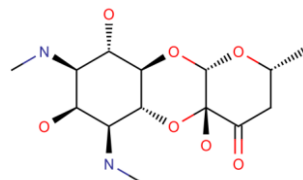
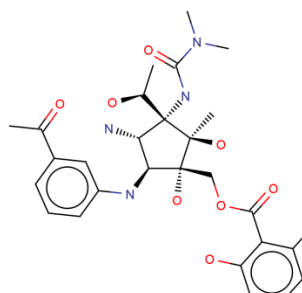
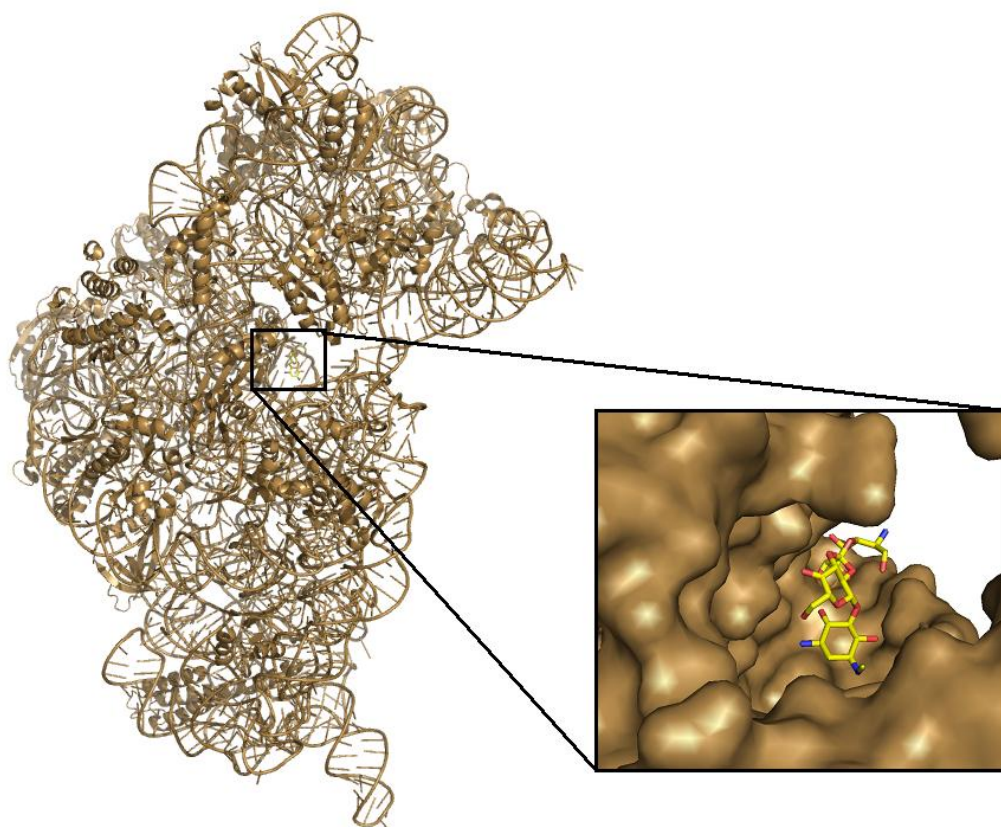
Aminoglikozydy	
<p>Streptomycyna (identyfikator PDB: SRY) (Carter, Clemons i wsp. 2000)</p> 	<p>Paromycyna (identyfikator PDB: PAR) (Vicens i Westhof 2001)</p> 
<p>Higromycyna B (identyfikator PDB: HYG) (Brodersen, Clemons i wsp. 2000)</p> 	<p>Apramycyna (identyfikator PDB: AM2) (Han, Zhao i wsp. 2005)</p> 
Oksazolidynon	Tetracyklina
<p>Linezolid (identyfikator PDB: ZLD) (Patel, Yan i wsp. 2001)</p> 	<p>Tetracyklina (identyfikator PDB: TAC) (Pioletti, Schlunzen i wsp. 2001)</p> 
Inne	
<p>Spektynomycyna (identyfikator PDB: SCM) (Carter, Clemons i wsp. 2000)</p> 	<p>Paktamycyna (identyfikator PDB: PCY) (Brodersen, Clemons i wsp. 2000)</p> 

Tabela 3. Wybrane antybiotyki wiążące się z rRNA w małej podjednostce rybosomu.



Rycina 3. Hygromycyna B związana w małej podjednostce rybosomu *Thermus* (kod PDB: 1HNZ).

Rybozomy

Rybosom nie jest jedynym celem aminoglikozydów. Związki te blokują: miejsce rozpoczęcie się dimeryzacji HIV-1 RNA (Ennifar, Paillart i wsp. 2006), wycinanie intronów grupy I (Park, Kim i wsp. 2000), rybozomy HDV (Chen, Wu i wsp. 1997) czy rybozomy typu „głowy młotka” z kilku wiroidów atakujących rośliny (Borda i Sigurdsson 2004). Więcej informacji na temat rybozymów znajduje się w Rozdziale 1.1 „Jony a aktywność biologiczna RNA”.

Istnieje również wiele opublikowanych prac o potencjale terapeutycznym katalitycznych RNA. Najczęściej rybozym typu „głowy młotka”, który jest najmniejszym i najlepiej poznanym rybozymem, wykorzystywany jest jako podstawa do projektowania modyfikowanych rybozymów o działaniu terapeutycznym, czułych na obecność specyficznego liganda (Soukup i Breaker 1999, Auslander, Ketzer i wsp. 2010, Mastroiannopoulos, Uney i wsp. 2010).

Sily fizyczne uczestniczące w oddziaływaniach RNA z ligandami

Ligandy wiążą się do cząsteczek RNA niekowalencyjnie, co oznacza, że siła tych oddziaływań jest wypadkową wielu rodzajów oddziaływań składających się na nie atomów. Ponieważ cząsteczki RNA są silnie, ujemnie naładowane, ligandy wiążące się do nich, są z reguły naładowane dodatnio. Znane są jednak przypadki wiązania się anionów do RNA (Kieft, Chase i wsp. 2010). Oddziaływania składające się na siłę wiązania kompleksu RNA z ligandem to przede wszystkim klasyczne oddziaływania elektrostatyczne, bezpośrednie oddziaływania polarne, wodorowe, hydrofobowe i hydrofilowe czy van der Waalsa. Należy jednak podkreślić, że obecność jonów znajdujących się zarówno w rozpuszczalniku jak i związanych z RNA ma ogromny wpływ na oddziaływania RNA z ligandami i je moduluje.

1.3. Modelowanie struktury kompleksów RNA z jonami metali i ligandami

W celu zrozumienia ról pełnionych przez RNA niezbędna jest wiedza o jego strukturze i oddziaływaniach z innymi cząsteczkami, gdyż to one bezpośrednio determinują jego funkcje (Eddy 2001, DeRose 2003). Istnieją doświadczalne metody rozwiązywania struktur kompleksów RNA z jonami metali i ligandami (a także z białkami i innymi kwasami nukleinowymi), które dają wiele informacji na temat tych oddziaływań. Do najpopularniejszych z nich należą:

- krystalografia rentgenowska,
- spektroskopia magnetycznego rezonansu jądrowego (ang. *Nuclear Magnetic Resonance*, NMR),
- mikroskopia elektronowa.

Doświadczalne metody wyznaczania struktury RNA i jego kompleksów są czasochłonne i drogie, a niekiedy niemożliwe jest ich zastosowanie. (Scott i Hennig 2008). W takiej sytuacji alternatywą pozostaje skorzystanie z narzędzi bioinformatycznych, które wspierają badania doświadczalne i stanowią ich uzupełnienie.

Komputerowe przewidywanie miejsc wiązania RNA z jonami metali

Poprawne zidentyfikowanie miejsc wiązania i dodanie jonów metali do struktur RNA (a także białek), których niekiedy w nich brakuje (np. struktury rozwiązywane za pomocą NMR), a niekiedy są błędnie oznaczone (np. jako cząsteczka wody) jest kluczowe dla przeprowadzenia dynamiki molekularnej. W związku z tym istnieje wiele funkcji i programów do tego stworzonych, które najczęściej zawarte są w oprogramowaniu towarzyszącym metodom do dynamiki (np. CHARMM (Brooks, Brooks i wsp. 2009)). Obecnie również wiele przeglądark do struktur oferuje funkcję dodawania jonów (np. Chimera (Pettersen, Goddard i wsp. 2004)). Wszystkie te metody wyznaczają elektrostatyczne pole sił dla zadanej struktury i w najbardziej korzystnych energetycznie miejscach umieszczają jony. Metody te dobrze sprawdzają się dla białek, które to posiadają zróżnicowany ładunek na powierzchni. Jednak nie zawsze dają dobry rezultat dla kwasów nukleinowych, gdyż te są całkowicie ujemnie naładowane. Powstało więc kilka alternatywnych podejść do tego zagadnienia. Większość z nich modeluje właściwości fizyko-chemiczne zadanego systemu. Hermann i Westhof (Hermann i Westhof 1998) zastosowali dynamikę brownowską (ang. *Brownian Dynamics*, BD) do przeprowadzenia symulacji zachowania się kationów w polu elektrostatycznym, a w konsekwencji do przewidzenia miejsc ich wiązań w strukturach RNA. Misra i Draper (Misra i Draper 2000) stworzyli analityczny model, stosując nieliniowe równanie Poissona-Boltzmann (PB) do opisu zależności energii i stechiometrii pomiędzy wiązaniem jonu Mg^{2+} a związaniem się RNA. Tan i Chen (Tan i Chen 2005, Tan i Chen 2010) także zbudowali swój model oddziaływań opierając się na teorii PB. Umożliwił on modelowane oddziaływań elektrostatycznych i sterycznych jonów bezpośrednio związanych z RNA, a do przewidywania oddziaływań jonów słabo związanych użyli modelu pola średniego. Zaletą wszystkich wymienionych metod jest fakt, że umożliwiają one badanie zadanego systemu przy dowolnych parametrach zdefiniowanych przez użytkownika (np. temperatura, stężenie jonów, obecność innych cząsteczek itd.). Niemniej jednak, metody te są bardzo kosztowne obliczeniowo i wymagają eksperckiej wiedzy zarówno do ich zastosowania jak i interpretacji wyników. Niemożliwa jest też automatyzacja procesu ich użycia, stąd nie zostały one uwzględnione w porównaniu skuteczności metod w prezentowanej pracy.

Program FEATURE (Liang, Banatao i wsp. 2003) natomiast stanowi przykład zautomatyzowanego oprogramowania, które jest dostępne poprzez serwer internetowy, gdzie możliwe jest przeprowadzenie przewidywań „online” dla zadanej cząsteczki (WebFEATURE). FEATURE został wytrenowany do przewidywania jonów magnezu w strukturach RNA poprzez uczenie nadzorowane. Jego zbiór uczący składał się zarówno z pozytywnych, jak i negatywnych przykładów miejsc wiązania kationów Mg^{2+} . Mg^{2+} natomiast zostały podzielone na dwie grupy – te „silnie” związane bezpośrednio z RNA (ang. *site-bound*) oraz te „słabo” związane, fluktuujące wokół powierzchni RNA (ang. *diffused*). W sumie w stworzonym modelu statystycznym uwzględniono 126 fizyko-chemicznych i strukturalnych czynników, które uczestniczą bądź wpływają na powstanie kompleksu RNA z jonami Mg^{2+} .

Komputerowe przewidywanie oddziaływań RNA z ligandami

Do przewidywania oddziaływań RNA z ligandami stosuje się dokowanie molekularne (ang. *molecular docking*). Jest to metoda, która na drodze obliczeniowej pozwala na dopasowanie małych cząsteczek (ligandów) do innego związku lub jego fragmentu, zwanego receptorem. Ponieważ funkcjonalne znaczenie cząsteczek RNA zostało odkryte stosunkowo niedawno dotychczasowe badania oddziaływań koncentrowały się głównie na strukturach białkowych. Istnieje więc szereg metod umożliwiających analizy oddziaływań białek z białkami, bądź białek z ligandami (programy do dokowania AutoDock, GOLD, Surflex itd.), których jakość przewidywań jest bardzo wysoka (Kellenberger, Rodrigo i wsp. 2004, Bottegoni 2011). Wzrost zainteresowania oddziaływaniami RNA z ligandami w ostatniej dekadzie, spowodował rozwój metod do dokowania i powstanie funkcji oceny dla RNA, możliwych do zaimplementowania w istniejących już programach dla białek. Wynika to z faktu, iż głównymi problemami związanymi z dokowaniem molekularnym są: wygenerowanie pól liganda bliskich natywnej oraz wyszukanie tej pozy w licznych rezultatach dokowania. O ile istniejące programy do dokowania dla białek doskonale radzą sobie z pierwszym zadaniem również w przypadku RNA, o tyle nie sprawdzają się one w tym drugim (Detering i Varani 2004, Li, Shen i wsp. 2010). Dlatego też wiele grup zajmujących się przewidywaniem oddziaływań RNA z ligandami (również w przypadku niniejszego projektu) zdecydowało się na użycie już istniejących programów do dokowania w celu wygenerowania alternatywnych pól liganda

w strukturze RNA. Skupiły się one natomiast na stworzeniu funkcji oceny wygenerowanych kompleksów, by efektywnie wybierać te bliskie struktury natywnej.

Morley i Afshar jako pierwsi zaproponowali specyficzną funkcję oceny dla kompleksów RNA z ligandem. Rozszerzyli oni funkcjonalność swojego programu do dokowania dodając empiryczną funkcję RiboDock (lub rDock) do oceny pozycji ligandów w strukturach RNA (Morley i Afshar 2004). Ich funkcja uwzględnia bardzo szeroki zakres oddziaływań międzymolekularnych, typowych dla wiązania ligandów w strukturach RNA. Są to wiązania wodorowe, oddziaływania lipofilowe, steryczne, dodatnio naładowane oddziaływania węgiel-akceptor, oddziaływania warstwowe pierścieni aromatycznych (ang. *stacking interactions*), odpychanie par donor-donor i akceptor-akceptor oraz zmiana entropii w procesie wiązania. Funkcja ta została sparametryzowana i testowana na zbiorze jedynie 10 kompleksów.

Moitessier i wsp. (Moitessier, Westhof i wsp. 2006) stworzyli funkcję oceny ograniczoną jedynie do oddziaływań RNA z aminoglikozydami. Funkcja ta została zaimplementowana w programie do dokowania białek AutoDock (Morris, Huey i wsp. 2009). Opiera się ona na wyliczeniu energii oddziaływań międzycząsteczkowych, na którą składa się między innymi energia oddziaływań dynamicznie związanych cząsteczek wody. Bardzo istotną cechą tej metodologii jest fakt, że autorzy dodatkowo wprowadzili możliwość dokowania elastycznego.

DrugScoreRNA jest empiryczną funkcją zaproponowaną przez Gohlke i wsp. (Pfeffer i Gohlke 2007). Funkcja ta opiera się na izotropowym potencjale statystycznym (zależnym tylko od odległości), który został wyprowadzony ze znanych struktur kompleksów RNA z ligandami. Metodologia ta została z powodzeniem zastosowana już wcześniej, przez tą samą grupę badawczą, do stworzenia funkcji ocen dla kompleksów białek z ligandami (Gohlke, Hendlich i wsp. 2000). To podejście zakłada, że siłę oddziaływania między atomem liganda i atomem RNA, znajdujących się w danej odległości od siebie, można oszacować na podstawie znormalizowanej funkcji rozkładu par tych atomów. Funkcja rozkładu par atomów z kolei została wyprowadzona ze znanych kompleksów w postaci statystyk kontaktów międzyatomowych. Oryginalny potencjał DrugScoreRNA otrzymano z 670 krystalograficznie rozwiązanych kompleksów kwasów nukleinowych z ligandami i z białkami. Funkcja ta również została zaimplementowana w programie AutoDock.

Dock6 to zbiór programów i skryptów do przeprowadzenia dokowania ligandów do struktur białek, jednak ostatnio jego funkcjonalność została rozszerzona również do dokowania ligandów do struktur RNA (Lang, Brozell i wsp. 2009). Procedura ta składa się z następujących kroków: wygenerowanie sfer reprezentujących powierzchnię receptora oraz wybór tych sfer, które definiują miejsce/a aktywne. Możliwy jest wybór największego klastra sfer, gdy nie ma żadnej informacji o miejscu wiązania się liganda do cząsteczki receptora (największy klaster teoretycznie znajduje się w najbardziej korzystnym z punktu widzenia energetycznego miejscu receptora, stąd dokowanie do tego miejsca może dać poprawne rezultaty). Możliwy jest też wybór sfer w określonym promieniu od dowolnego miejsca zdefiniowanego przez użytkownika, w przypadku, gdy znane jest w przybliżeniu miejsce wiązania liganda. Następny krok to wyznaczenie tak zwanej siatki (ang. *grid*). Na tym etapie tworzy się mapa energetyczna układu na podstawie położenia sfer i powierzchni receptora. Kolejny krok stanowi właściwe dokowanie liganda. Polega ono na dopasowywaniu zadanej cząsteczki do receptora. Program Dock6 daje możliwość przeprowadzenia dokowania sztywnego (ang. *rigid body docking*) oraz elastycznego (ang. *flexible docking*). Dokowanie sztywne polega na dopasowywaniu liganda do receptora w sposób sztywny, bez jakichkolwiek modyfikacji tej cząsteczki. Dokowanie elastyczne natomiast daje możliwość dokowania ligandów do takich receptorów, do których zdają się one nie pasować bez zmiany konformacji. Dock6 najpierw identyfikuje sztywne fragmenty liganda (np. pierścienie aromatyczne), a następnie tworzy wiele alternatywnych konformacji, np. przez obroty sztywnych fragmentów wokół wiązań pojedynczych. Następnie każda wygenerowana poza jest poddawana ocenie. Program Dock6 oferuje aż dziewięć funkcji oceny, różniących się złożonością obliczeniową oraz podstawami teoretycznymi (Kuntz, Blaney i wsp. 1982). Program Dock6 został omówiony szerzej w Rozdziale 3.3 „Dokowanie ligandów do struktur RNA”.

Najnowszym z programów do dokowania ligandów do struktur RNA jest MORDOR (ang. *MOlecular Recognition with a Driven dynamics OptimizeR*) stworzony przez Guilberta i Jamesa (Guilbert i James 2008). Bazuje on na metodologii podobnej do Dock6 i podobnie jak on wykorzystuje klasyczne pole siłowe mechaniki molekularnej w trakcie procesu dokowania. Jego zdolność identyfikacji pozycji liganda bliskich strukturze natywnej jest porównywalna z rezultatami Dock6, lecz Dock6 jest 3~10 razy szybszy (Lang, Brozell i wsp. 2009).

Ocena faktycznej jakości modeli

Zarówno komputerowe wyznaczanie miejsc wiązania jonów w strukturach RNA, jak i dokowanie do niego ligandów, jest rodzajem teoretycznego przewidywania. Jak w przypadku każdego przewidywania, nigdy nie daje ono stu procentowej pewności, że jego wynik jest poprawny. Nawet najdokładniejsza funkcja oceny może wytypować niepoprawne miejsce wiązania jonu czy konformację liganda, więc każdy otrzymany wynik powinien zostać poddany dodatkowej weryfikacji. Jeśli istnieje rozwiązana doświadczalnie struktura modelowanego kompleksu najpewniejszą metodą jest porównanie z nią otrzymanego rezultatu. Jednakże komputerowe przewidywanie oddziaływań między cząsteczkami stosuje się najczęściej w przypadkach, gdy struktura natywna nie jest znana, bądź poszukuje się alternatywnych inhibitorów znanego receptora. W takim przypadku zaleca się zastosowanie poniższych technik, w zależności od posiadanych danych, np.:

- porównanie wiązań (otrzymanych doświadczalnie z mutagenyzy czy sieciowania molekularnego);
- sprawdzenie podobieństwa struktur generowanych kompleksów (np. przy pomocy klasteryzacji);
- analiza konfliktów sterycznych;
- analiza rozkładu potencjału elektrostatycznego na powierzchni cząsteczek;
- analiza stereochemiczna;
- wizualna ocena wyników.

2. Cel pracy

Celem prezentowanej pracy doktorskiej było stworzenie nowych programów komputerowych służących do przewidywania oddziaływań RNA z jonami metali (program MetalionRNA) i niskocząsteczkowymi ligandami (program LigandRNA). Podejście zastosowane w obu programach jest takie samo i opiera się na potencjałach statystycznych wyprowadzonych z bazy danych znanych struktur kompleksów RNA z jonami metali i ligandami. Ponadto celem pracy było przetestowanie powstałych programów na zbiorze znanych kompleksów RNA z jonami metali (MetalionRNA) oraz z ligandami (LigandRNA).

Dodatkowym zadaniem praktycznym było użycie programu LigandRNA do przeprowadzenia wirtualnej analizy przesiewowej (ang. *Virtual Screening*, VS) dla struktury centrum peptydylotransferazy bakteryjnych rybosomów, która jest celem działania wielu antybiotyków. W najbliższej przyszłości (poza ramami projektów badawczych opisanych w niniejszej rozprawie) planowana jest weryfikacja doświadczalna przewidzianych potencjalnych nowych inhibitorów rybosomu bakteryjnego, będzie ona przeprowadzona w laboratorium prof. Bujnickiego w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej w Warszawie.

3. Materiały i metody

3.1. Potencjał statystyczny

Potencjały statystyczne mogą posłużyć do opisu oddziaływań między cząsteczkami, a w przypadku niniejszego projektu, do opisu oddziaływań między RNA a jonami metali i ligandami. Należą do grupy potencjałów empirycznych, zbudowanych na podstawie zbioru rzeczywistych, rozwiązanych eksperymentalnie, struktur.

Oddziaływania między biomolekułami można opisać przy pomocy energii swobodnej Gibbsa, której zmiana wiąże się ze stałą wiązania K_b (ang. *binding constant*). (Sippl 1993, Sippl 1995):

$$\Delta G^0 = -RT \ln K_b$$

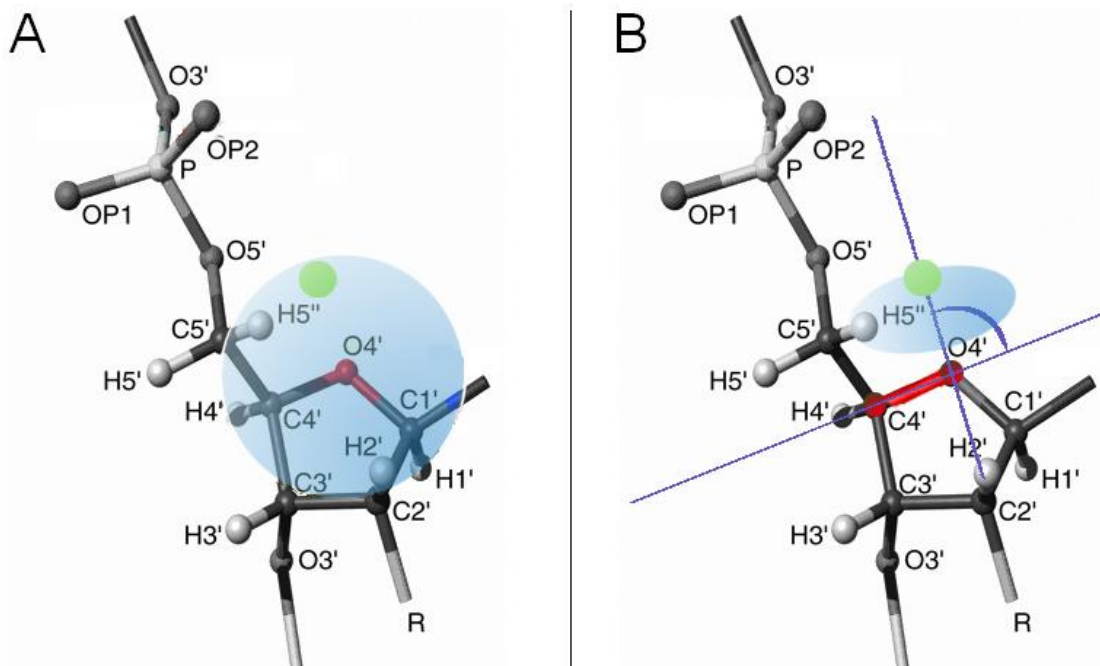
gdzie R oznacza stałą gazową, T temperaturę bezwzględną układu, a K_b wyraża się wzorem:

$$K_b = \frac{[RNA \cdots ligand]}{[RNA][ligand]}$$

gdzie $[RNA \cdots ligand]$ to stężenie kompleksów RNA z ligandem (jony metali są szczególnym przypadkiem jednoatomowych ligandów), natomiast $[RNA]$ i $[ligand]$ to stężenia niezwiązanych składników tego kompleksu. Potencjały statystyczne bazują na założeniu, że wartość energii swobodnej Gibbsa może zostać oszacowana na podstawie statystyk występowania kontaktów pomiędzy poszczególnymi atomami oddziałujących ze sobą molekuł. W przypadku niniejszego projektu są to wspomniane oddziaływania między RNA a jonami metali, bądź atomami liganda. Statystyki kontaktów międzyatomowych wyprowadzone zostały ze struktur znanych kompleksów (zdeponowanych w bazie PDB (ang. *Protein Data Bank*)). Podejście to zostało już wcześniej z powodzeniem zastosowane przez inne grupy badawcze do opisu oddziaływań białek i RNA z ligandami (Gohlke, Hendlich i wsp. 2000, Pfeffer i Gohlke 2007) (patrz: Rozdział 1.3 „Modelowanie struktury kompleksów RNA z jonami metali i ligandami”). Autorzy wymienionych metod stworzyli potencjały izotropowe, bazujące jedynie na odległościach między atomami. W niniejszej pracy stworzony potencjał jest anizotropowy, co oznacza, że poza odległościami pomiędzy

Materiały i metody

poszczególnymi atomami, uwzględnione zostały także tworzone przez nie kąty. Rycina 4 prezentuje różnice w obu podejściach. Wprowadzenie parametru kąta znacząco zmniejszyło przedziały opisujące przestrzeń wokół atomów RNA, co w rezultacie udokładniło potencjał i pozwoliło na bardziej szczegółowe ujęcie specyficzności oddziaływań.



Rycina 4. Panel A) idea potencjału izotropowego, zależnego tylko od odległości, w tym przypadku między atomem RNA O4' a Mg^{2+} (zielona sfera). Panel B) idea potencjału anizotropowego zależnego zarówno od odległości między parą atomów RNA O4' a Mg^{2+} jak i od kąta tworzonego przez atomy pary [C4', O4'] i Mg^{2+} .

Omawiany potencjał statystyczny został wyprowadzony w oparciu o rozkład Boltzmann. Polegało to na przełożeniu n -elementowej funkcji korelacji $g^{(n)}(d_1, \alpha_1; \dots, d_n, \alpha_n)$, na potencjał statystyczny $W^{(n)}(d_1, \alpha_1; \dots, d_n, \alpha_n)$, który reprezentuje energię swobodną Gibbsa, przy użyciu następującej formuły:

$$W^{(n)}(d_1, \alpha_1; \dots, d_n, \alpha_n) = -RT \ln g^{(n)}(d_1, \alpha_1; \dots, d_n, \alpha_n)$$

gdzie $g^{(n)}$ oznacza częstość występowania kontaktów pomiędzy jonem (bądź atomem liganda) c a wszystkimi zdefiniowanymi parami atomów RNA $[a, b]$

(patrz: Rozdział 4.1 „Anizotropowy potencjał statystyczny”); d oznacza odległość pomiędzy atomem c a atomem b ; α oznacza kąt pomiędzy atomami (a, b, c). W niniejszym projekcie funkcja korelacji $g^{(n)}$ została obliczona ze wzoru:

$$g^{(n)}(d_1, \alpha_1; \dots d_n, \alpha_n) = \frac{g^{(n)}_{\text{obserwowane}}(d_1, \alpha_1; \dots d_n, \alpha_n)}{g^{(n)}_{\text{oczekiwane}}(d_1, \alpha_1; \dots d_n, \alpha_n)}$$

gdzie $g^{(n)}_{\text{obserwowane}}$ to znormalizowana liczba oddziaływań między parą atomów RNA [a, b] a kationem/atomem liganda c , w przedziale odległości/kąta (d, α), która została zaobserwowana w przygotowanym wcześniej w celu wyprowadzenia potencjałów zestawie kompleksów o znanej strukturze. Ponieważ przedziały przestrzeni opisywane przez wartości odległości d i kąta α mają różną objętość (im większa wartość d tym większa objętość dla tego samego kąta α), konieczne było wprowadzenie normalizacji. Polegało ono na podzieleniu liczby kontaktów w odpowiednich przedziałach odległości i kątów (d, α), przez wartości objętości odpowiadających im przedziałów V :

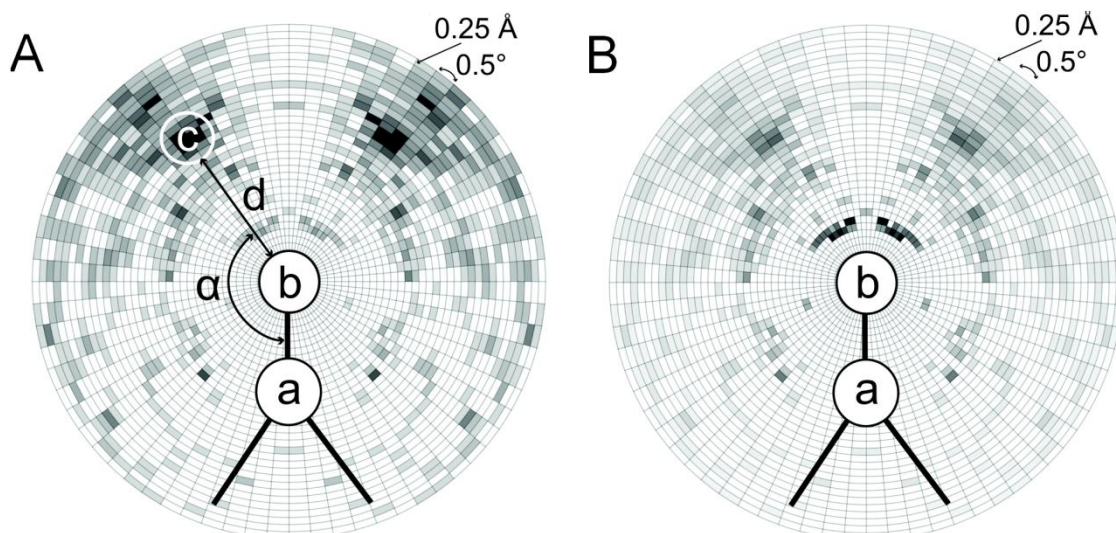
$$V(d, \alpha) = \frac{2}{3} \pi [(d + \Delta d)^3 - d^3] |\cos(\alpha + \Delta \alpha) - \cos \alpha|$$

Natomiast $g^{(n)}_{\text{oczekiwane}}$ reprezentuje liczbę oczekiwanych kontaktów pomiędzy parą atomów RNA [a, b] a kationem/atomem liganda c , w przedziale odległości/kąta (d, α). W przypadku prezentowanego potencjału wartość ta została wyznaczona na podstawie wartości ułamków molowych atomów danego typu:

$$g^{(n)}_{\text{oczekiwane}}(d_1, \alpha_1; \dots d_n, \alpha_n) = X_{[a,b]} * Y_c * g^{(n)}_{\text{obs_całkowite}}(d_1, \alpha_1; \dots d_n, \alpha_n)$$

gdzie $X_{[a,b]}$ oraz Y_c reprezentują odpowiednio ułamki molowe: par atomów RNA typu [a, b] oraz kationów/atomów liganda typu c , natomiast $g^{(n)}_{\text{obs_całkowite}}$ reprezentuje całkowitą liczbę obserwowanych kontaktów w przedziale odległości/kąta (d, α).

Rycina 5 przedstawia przykład potencjału wyprowadzonego dla oddziaływań pary atomów RNA [P, OP2] oraz Mg^{2+} . Panel A) przedstawia funkcję $g^{(n)}_{\text{obserwowane}}$ przed normalizacją. Panel B) przedstawia potencjał statystyczny $W^{(n)}$.



Rycina 5. Panel A) Potencjał statystyczny otrzymany dla pary RNA [P, OP2] (na rycinie [a, b]) oraz jonów Mg^{2+} (na rycinie przykładowe położenie tego jonu c). Panel B) Znormalizowany potencjał statystyczny, przedstawiony na Panelu A. Potencjał ukazany jest w skali szarości (im ciemniejszy kolor tym mniejsza wartość potencjału).

3.2. Przygotowanie zbiorów uczących i testowych

Zbiory MetalionRNA

Do stworzenia zbioru uczącego i testowego wykorzystane zostało 113 struktur pobranych z bazy danych PDB (Rose, Beran i wsp. 2011) (patrz: Tabela 4), które spełniały następujące kryteria:

- zawierały RNA oraz jeden, bądź więcej jonów (magnezu, sodu lub potasu) z nim związanych;
- zostały rozwiązane przy użyciu krystalografii rentgenowskiej, a ich rozdzielczość nie przekraczała 2 Å dla struktur RNA z Mg^{2+} i 3 Å dla struktur z K^+ i Na^+ . Kryterium to wynika bezpośrednio z faktu, iż rozdzielczość struktur rozwiązanych przy pomocy tej techniki jest kluczowym czynnikiem dla właściwego rozpoznania typu i pozycji kationu, a jedynie struktury o najwyższej jakości mogły posłużyć do stworzenia wiarygodnego potencjału statystycznego. Przyjęcie wyższego progu (3 Å) dla struktur z K^+ i Na^+ wynika z faktu, iż tylko dla takiego zakresu udało się pozyskać liczbę struktur wystarczającą do stworzenia rzetelnych statystyk;
- miały najlepszą rozdzielczość spośród struktur, w których identyczność sekwencji RNA przekraczała 90%.

Ponadto, by zapewnić, że kationy, które posłużyły do wyprowadzenia potencjału, oddziałują jedynie z cząsteczką RNA, ze struktur zostały usunięte kationy, które znajdowały się w odległości bliższej niż 9 Å do atomu innego niż atom RNA, wody lub do innego kationu.

Do przeprowadzenia dodatkowego testu programu MetalionRNA posłużyło 116 struktur z bazy PDB (patrz: Tabela 5):

- zawierających DNA i co najmniej jeden Mg^{2+} ;
- rozwiązanych przy pomocy krystalografii, o rozdzielczości lepszej niż 3 Å;
- o najlepszej rozdzielczości spośród struktur, w których identyczność sekwencji DNA przekraczała 90%.

Dodatkowo, do wyprowadzenia potencjału MetalionRNA ze zbioru uczącego programu FEATURE posłużyły struktury, których identyfikatory PDB zawiera Tabela 6.

Kompleksy RNA z Mg^{2+}	Kompleksy RNA z Na^+	Kompleksy RNA z K^+
1D4R, 1DFU, 1DUL, 1EHZ, 1EVV, 1F27, 1FUF, 1HQ1, 1I7J, 1IK5, 1J1U, 1JID, 1JZV, 1L2X, 1LNT, 1NUJ, 1NUV, 1PJO, 1XPE, 2A43, 2ANR, 2G3S, 2G91, 2OE5, 2OIY, 2Q1R, 2QEK, 2R1S, 2R20, 2R8S, 2VAL, 2Z75, 2ZY6, 310D, 354D, 3CGS, 3D2V, 3DIL, 3GAO, 3GVN, 3JXQ, 3LA5, 3MEI, 3NKB, 3OIN, 3P4A, 437D, 468D, 469D, 470D	1F7Y, 1J6S, 1J8G, 1L2X, 1L3Z, 1MDG, 1NTA, 1NTB, 1Q93, 1VBY, 1XPE, 1Y6T, 2DQO, 2GDI, 2GXB, 2OJ0, 2P7D, 2R20, 3BNT, 3DIL, 3L25, 420D, 434D, 437D	1DUL, 1FFY, 1HC8, 1JBR, 1JBS, 1L2X, 1P79, 1QU2, 1SDS, 1Y39, 1YRJ, 1ZCI, 2ANR, 2AWE, 2FCX, 2FD0, 2GDI, 2GRB, 2OIJ, 2OIY, 2QBZ, 2QEK, 3BNP, 3BNQ, 3BNS, 3CUL, 3DIL, 3DIX, 3DVV, 3F2Q, 3FAR, 3FTF, 3FTM, 3G78, 3GX5, 3IBK, 3L0U, 3LQX

Tabela 4. Lista kodów z bazy danych PDB struktur RNA z jonami magnezu, sodu i potasu, użytych do wyprowadzenia potencjału statystycznego oraz przeprowadzenia walidacji krzyżowej.

Kompleksy DNA z Mg²⁺
101D, 109D, 119D, 150D, 159D, 184D, 1BDN, 1D23, 1D33, 1D35, 1D38, 1D41, 1D45, 1D49, 1D57, 1D60, 1D63, 1D8X, 1DA1, 1DA3, 1DCG, 1DJ6, 1DNF, 1DNZ, 1DPL, 1EDR, 1EI4, 1EM0, 1EN3, 1EN9, 1ENN, 1F69, 1FD5, 1FMS, 1FQ2, 1FTD, 1G8N, 1I3T, 1I47, 1ICK, 1IH1, 1IKK, 1IXJ, 1JGR, 1LJX, 1LP7, 1M6F, 1N1O, 1NGT, 1NR8, 1P1Y, 1PUY, 1QP5, 1R3G, 1RQY, 1S45, 1UE2, 1UHX, 1VAQ, 1VRO, 1VZK, 1XVK, 1XVR, 1Y8L, 1ZPI, 208D, 253D, 267D, 291D, 292D, 2B0K, 2B1B, 2B1D, 2B3E, 2D25, 2D34, 2DCG, 2DES, 2DP7, 2DZ7, 2ELG, 2FII, 2GVR, 2H9S, 2JJA, 2PIS, 2WNA, 303D, 306D, 308D, 313D, 335D, 336D, 355D, 360D, 370D, 382D, 383D, 392D, 3DNB, 3EY0, 3GDD, 3GGK, 3IFF, 3LPV, 3LTR, 403D, 412D, 423D, 428D, 431D, 442D, 455D, 5ANA, 5DNB, 8BNA

Tabela 5. Lista kodów z bazy danych PDB struktur DNA z jonami magnezu, użytych do dodatkowych testów.

Kompleksy ze zbioru uczącego programu FEATURE
1CX0, 1DRZ, 1DUH, 1EHZ, 1EVV, 1F27, 1FIR, 1GID, 1HR2, 1I9V, 1JJ2, 301D, 354D, 357D, 364D, 429D, 430D, 462D

Tabela 6. Lista kodów z bazy danych PDB struktur ze zbioru uczącego programu FEATURE.

Zbiory LigandRNA

Do stworzenia zbioru uczącego i testowego wykorzystanych zostało 251 struktur pobranych z bazy danych PDB (patrz: Tabela 7), które spełniały następujące kryteria:

- zawierały RNA oraz jeden, bądź więcej ligandów z nim związanych;
- miały najlepszą rozdzielczość spośród kompleksów z tym samym ligandem, w których identyczność sekwencji RNA przekraczała 90%;
- ligandy były organiczne (np. nie były to jony).

Ponadto, by zapewnić, że ligandy, które posłużyły do wyprowadzenia potencjału, oddziałują jedynie z cząsteczką RNA, ze struktur zostały usunięte ligandy, które znajdowały się w odległości bliższej niż 6 Å do atomu innego niż atom RNA, wody lub do jonu.

Kompleksy RNA z ligandami

1AJU, 1AKX, 1AM0, 1ARJ, 1BYJ, 1EHT, 1EI2, 1ET4, 1EVV, 1F1T, 1F27, 1F7U, 1FJG, 1FMN, 1FUF, 1FYP, 1G4Q, 1HNW, 1HNX, 1HNZ, 1I97, 1I9V, 1IBL, 1J5A, 1J7T, 1J8G, 1JZX, 1JZY, 1JZZ, 1K01, 1K73, 1K8A, 1K9M, 1KC8, 1KD1, 1KOC, 1KOD, 1KQS, 1LC4, 1LVJ, 1M1K, 1M90, 1MWL, 1N32, 1N33, 1N8R, 1NBK, 1NEM, 1NJM, 1NJJ, 1NTA, 1NTB, 1NWX, 1O15, 1O9M, 1OND, 1P9X, 1PBR, 1Q81, 1Q8N, 1QD3, 1RAW, 1SM1, 1TN2, 1TOB, 1U8D, 1U8T, 1UUD, 1UUI, 1VQ6, 1VQ8, 1VQL, 1VQM, 1VQN, 1VS5, 1XBP, 1XMQ, 1XPF, 1Y26, 1YHQ, 1YI2, 1YIJ, 1YIT, 1YJW, 1YRJ, 1YYK, 2A04, 2AU4, 2B57, 2BE0, 2BEE, 2CKY, 2DQQ, 2EES, 2EEU, 2ESI, 2ESJ, 2ET3, 2ET4, 2ET5, 2ET8, 2F4S, 2F4T, 2F4U, 2FCX, 2FCY, 2FCZ, 2FD0, 2G32, 2G5K, 2G5Q, 2G9C, 2GCV, 2GDI, 2GIS, 2GQ4, 2GUN, 2H0Z, 2HHH, 2HO6, 2HOJ, 2HOM, 2J00, 2JUK, 2KXM, 2NPY, 2NZ4, 2O3V, 2O3W, 2O3X, 2O3Y, 2OE5, 2OE8, 2OGN, 2OGO, 2OIJ, 2P7E, 2PWT, 2QAM, 2QB9, 2QBA, 2QEX, 2QOU, 2QOW, 2QOY, 2QOZ, 2QP0, 2QP1, 2QWY, 2TOB, 2TRA, 2UUB, 2UUC, 2UXB, 2UXD, 2VQE, 2W89, 2WDG, 2WRN, 2XQD, 2Z4L, 2Z74, 2Z75, 3B4A, 3B4B, 3B5S, 3BNQ, 3C44, 3C5D, 3C7R, 3CC4, 3CMA, 3CPW, 3CQS, 3D0U, 3D2G, 3D2V, 3D2X, 3DF1, 3DF3, 3DIG, 3DIL, 3DIO, 3DIQ, 3DIR, 3DIY, 3DJ0, 3DJ2, 3DLL, 3DS7, 3DVV, 3DW4, 3E5C, 3E5E, 3E5F, 3EGZ, 3F2Q, 3FO6, 3G4M, 3G4S, 3G6E, 3G71, 3G9C, 3GAO, 3GCA, 3GER, 3GLP, 3GOT, 3GS1, 3GS5, 3GX2, 3GX3, 3GX5, 3I2S, 3IQR, 3J0U, 3L3C, 3LA5, 3MXH, 3NPB, 3NPQ, 3OAT, 3OFC, 3OFR, 3OFZ, 3OHJ, 3OHZ, 3OI3, 3OI5, 3OWW, 3SD1, 3SD3, 3SFS, 3SKI, 3SKL, 3SKR, 3SKZ, 3SLM, 3SLQ, 3SUH, 3SUX, 3T1Y, 3TD1, 3TZR, 3UXQ, 3UXR, 3UZ3, 3UZ4, 3UZL, 3UZM, 3V7E

Tabela 7. Lista kodów z bazy danych PDB struktur kompleksów RNA z ligandami, użytych do wprowadzenia potencjału statystycznego oraz przeprowadzenia walidacji krzyżowej.

LigandRNA – Wybór kompleksów do zbioru testowego

Wszystkie 251 kompleksów RNA z ligandami zostało manualnie podzielonych na klastry według struktury chemicznej ligandów. W sumie otrzymano 62 klastry. Tabela 8 prezentuje identyfikatory PDB ligandów, jakie znalazły się w poszczególnych klastrach. Następnie przeprowadzono serię dokowań wybranych kompleksów (przy użyciu programu Dock6) w celu wygenerowania pól liganda o różnym RMSD. Wybór kompleksów był prowadzony w taki sposób, żeby uzyskać zbiór testowy o możliwie największej różnorodności ligandów. Preferowane były też te struktury, które zostały użyte wcześniej przez inne grupy badawcze do testowania ich metod.

Ostatecznie do zbioru testowego włączono 42 kompleksy RNA z ligandami (patrz: Tabela 9), dla których program Dock6 wygenerował od 69 do 1000 pól liganda (średnio 798 pól na kompleks) i przynajmniej jedna poza miała RMSD o wartości $<2 \text{ \AA}$ w stosunku do struktury referencyjnej. W omawianym zbiorze testowym znalazło się

Materialy i metody

wszystkie 10 struktur ze zbioru testowego programu RiboDock (Morley i Afshar 2004), sześć z 11 z zestawu testowego funkcji oceny dla aminoglikozydów (Moitessier, Westhof i wsp. 2006), 21 z 36 użytych do testów DrugScoreRNA (Pfeffer i Gohlke 2007) oraz 30 z 50 struktur ze zbioru testowego programu Dock6 (Lang, Brozell i wsp. 2009). Ponadto zbiór testowy LigandRNA pokrywał się ze zbiorami testowymi do oceny działania programów do dokowania białek użytymi dla RNA w przypadku 20 struktur z 36 (Detering i Varani 2004) i 33 z 60 (Li, Shen i wsp. 2010).

Nazwa klastra	Identyfikatory PDB ligandów
mały związek organiczny	D2X, PYI, ACE, IPA, GOL, TSE, MPD, TRS, CMY, OLZ, SLZ, SPD, CIR, S9L, SPM, NEG, ACA
puryna	ADE, HPA, 2BP, A2F, XAN, 6AP, 6GU, 6GO, TEP, PQ0
6-pyr	XXX, KSG
aminokwas	PHA, ARG, GND, HRG, LYS
podobny do kanamycyny	LLL, TOY, KAN, GET
podobny do neomycyny	NMY, PAR
podobny do neomycyny – aromatyczny łańcuch boczny	JS5, LHA
śruba molekularna	G80, MUL, G34
nukleozyd purynowy	3AD, GNG, GMP, DGP, 5GP, AMP, SAH, EEM, GTP
piranoza	G6P, GLP
poходna pirymidyny	5AZ, 3AY
podobny do spiramycyny	ERY, CTY, ZIT, CAI, SPR, TYK, ROX, TAO
tetracyklina	TAC, CTC
podobny do kwasu tetra-hydrofoliowego	FFO, THF, FOZ
tiamina	TPS, TPP
pozostałe ligandy należące do jednoelementowych klastrów	TEL, ANM, P14, SRY, P12, PA2, AB9, N30, AB6, P13, PCY, AM2, ZLD, MGR, IEL, ROS, VIR, CLM, CLY, FMN, WIN, T8B, RIO, DOL, AKN, SYB, LIV, JS4, G0B, CNC, SIN, PPU, SAM, SFG, SS0, BLS, N33, D2X, PYI, 1PE, SPS, HYG, MES, SCM, PMZ, BTN, HMT

Tabela 8. Klastry ligandów i ich identyfikatory PDB.

	kod PDB	Chemiczna nazwa liganda	identyfikator PDB liganda	Klaster liganda	Najmniejsza wartość RMSD [Å] pozy
1	1q8n	zieleń malachitowa	MGR	666	1,1
2	2f4u	(2r)-4-amino...*	AB6	6. piranoza.r.r	0,8
3	1uts	rbt550	P13	65-6-n6	1,8
4	2oe5	apramycyna	AM2	66. piranoza. piranoza	0,3
5	1uud	rbt203	P14	6-i-amidy	1,6
6	1uui	rbt205	P12	6-n6	1,2
7	2et8	neomycyna A	XXX	6-pyr	0,7
8	2o3x	NB33	N30	6-pyr-fur	0,5
9	1nbk	kwasy 2-amino-5-guanidynowalerianowy	GND	aminokwas	1,3
10	1koc	arginina	ARG	aminokwas	1,3
11	1aju	arginina	ARG	aminokwas	1,5
12	1f1t	N,N'- tetrametylorozamina	ROS	anthracen	1,1
13	1fmn	flawina	FMN	flawina	1,4
14	2fcz	rybostamycyna	RIO	fur-6-pyr	0,6
15	1byj	gentamycyna C1A	LLL	podobny do kanamycyny	0,3
16	1mwl	genetycyna	GET	podobny do kanamycyny	0,4
17	1tob	tobramycyna	TOY	podobny do kanamycyny	0,7
18	2tob	tobramycyna	TOY	podobny do kanamycyny	1,0
19	2g5q	amikacyna	AKN	podobny do kanamycyny - długi ł. boczny	0,5
20	1pbr	paromycyna	PAR	podobny do neomycyny	0,5
21	1fyp	paromycyna	PAR	podobny do neomycyny	0,6
22	1nem	paromycyna	PAR	podobny do neomycyny	0,6
23	1j7t	paromycyna	PAR	podobny do neomycyny	1,1
24	1ei2	neomycyna	NMY	podobny do	1,8

Materiały i metody

25	2be0	2''-o-[n**	JS5	neomycyny podobny do neomycyny - aromatyczny ł. boczny	1,1
26	2pwt	paromycyna PM-II-162	LHA	podobny do neomycyny - aromatyczny ł. boczny	1,6
27	2fd0	liwidomycyna A	LIV	podobny do neomycyny-pyr	1,4
28	2bee	2''-O-[N-(3-(aminopropylo) 2-aminoetylo]paromycyna	JS4	podobny do neomycyny - z ł. bocznym	1,7
29	2ogn	pleuromutylicyna	G80	śruba molekularna	1,0
30	1xbp	tiamulina	MUL	śruba molekularna	1,0
31	1eht	teofilina	TEP	puryna	0,2
32	1y26	adenina	ADE	puryna	0,4
33	1am0	adenozynomonofosforan	AMP	nukleozyd purynowy	0,9
34	3gx2	sinefungina	SFG	analog nukleozydu purynowego	0,5
35	3d2x	3-[(4-hydroksy...***	D2X	nukleotyd (pirymidyna)	1,4
36	1xpf	spermina	SPM	mały związek organiczny	1,3
37	1kod	cytrulina	CIR	mały związek organiczny	1,7
38	1fjg	spektynomycyna	SCM	spektynomycyna	0,3
39	1hnw	tetracyklina	TAC	tetracyklina	0,6
40	3sux	tetrahydrofuran	THF	podobny do kwasu tetra- hydrofoliowego	0,4
41	2gdi	pirofosforan tiaminy	TPP	tiamina	1,3
42	1f27	biotyna	BTN	poskręcany	0,8

Tabela 9. Zbiór testowy programu LigandRNA.

* (2R)-4-amino-N-((1R,2S,3R,4R,5S)-5-amino-4-[(2-amino-2-deoksy-alfa-D-glukopyranozylo)oksy]-2-{2-[(3aminopropylo)amino]etoksy}-3-hydroksycykloheksylo)-2-hydroksybutanoamid

** 2''-O-[N-(3-(aminometylo)-pyridina)-2-aminoetylo]paromycyna

*** 3-[(4-hydroksy-2-metylopyrimidyna-5-yl)metylo]-

5-(2{[(R)hydroksy(fosfonooksy)fosforylo]oksy}etylo)-4-metylo-1.3-tiazolo-3-ium

3.3. Dokowanie ligandów do struktur RNA

Dock6

W niniejszej pracy pakiet Dock6 został wykorzystany do wygenerowania alternatywnych pól ligandów (o różnym RMSD do struktur natywnych) dla kompleksów ze zbioru testowego programu LigandRNA. Ponadto otrzymane z niego wyniki zostały uwzględnione w porównaniu metod do dokowania dla RNA, ponieważ program ten posiada taką funkcjonalność (Lang, Brozell i wsp. 2009). Więcej informacji na temat zbioru testowego znajduje się w Rozdziale 3.2 „Przygotowanie zbiorów uczących i testowych”, a na temat programu Dock6 w Rozdziale 1.3 „Komputerowe przewidywanie oddziaływań RNA z ligandami”.

Przed procesem dokowania zarówno struktura receptora, jak i liganda muszą zostać uprzednio przygotowane. We wszystkich przypadkach do przygotowania cząsteczek został użyty program UCSF Chimera (Pettersen, Goddard i wsp. 2004). Posiada on wbudowaną funkcję „Dock Prep”, umożliwiającą automatyczne przygotowanie struktur z dodawaniem wodorów oraz ładunków elektrostatycznych włącznie (do wszystkich receptorów i ligandów zostało zastosowane pole sił CHARMM27 (MacKerell, Banavali i wsp. 2000)). Program UCSF Chimera został również użyty do generowania powierzchni molekularnej cząsteczek RNA.

Generowanie sfer opisujących kieszeń katalityczną zostało wykonane programem z pakietu Dock6 (Sphgen). Minimalny, jak i maksymalny rozmiar generowanych sfer, pozostał taki, jak w ustawieniach domyślnych, tj. odpowiednio 1,4 Å i 4,0 Å. Zakres ten pozwolił zarówno na umieszczanie sfer w wąskich przestrzeniach receptora, jak i w szerokich. Do selekcji sfer znajdujących się w miejscu aktywnym posłużył program Sphere_selector (z pakietu Dock6). Umożliwia on wybór sfer znajdujących się w zadanej odległości od liganda, jeśli jego położenie jest znane. Opcja ta stosowana jest najczęściej podczas dokowania liganda w znanym miejscu wiązania się innego związku. Jednak w przypadku niniejszego projektu posłużyła ona do wygenerowania pól o różnym RMSD dla znanych ligandów ze zbioru testowego kompleksów RNA z ligandami.

Generowanie siatki odbywało się przy pomocy programu Grid (pakiet Dock6) z wykorzystaniem ustawień domyślnych. Dokowanie (program Dock6) również odbyło się z użyciem ustawień domyślnych, by zwiększyć liczbę generowanych pól,

Materialy i metody

zmodyfikowane zostały jedynie parametry: `max_orientations = 1000` (w celu zwiększenia liczby generowanych konformacji) oraz `min_anchor_size = 1` (by program mógł w dowolny sposób poruszać ligandem. Im mniejszy jest sztywny rdzeń liganda, tym większe modyfikacje w strukturze liganda są możliwe).

Surflex

W niniejszej pracy użyto programu do dokowania do białek – Surflex (Jain 2003) (z użyciem parametrów domyślnych) do przeprowadzenia wirtualnej analizy przesiewowej dla struktury centrum peptydylotransferazy bakteryjnych rybosomów. Do przygotowania struktury receptora, tak jak w przypadku dokowania przy użyciu Dock6, posłużył program UCSF Chimera. Jako źródło potencjalnych inhibitorów został wykorzystany zbiór związków lekopodobnych (ang. *drug-like*) z bazy ligandów ZINC (Irwin, Sterling i wsp. 2012). Dla każdego liganda wygenerowane zostało 10 póż. Podczas działania program Surflex tworzy prototyp liganda, tzw. „protomol”, będący dużą, sztuczną cząsteczką pasującą do kieszeni katalitycznej. Jeśli ligand pasuje do struktury „protomola” (bądź jej części) oceniany jest wysoko. W przypadku niniejszego projektu Surflex został użyty jedynie do wygenerowania póż ligandów, a do ich oceny został wykorzystany program LigandRNA.

3.4. Metody i miary użyte do ewaluacji programów

RMSD

Średnia odległość pomiędzy odpowiadającymi sobie atomami (ang. *Root Mean Square Deviation*, RMSD) jest miarą wykorzystywaną do porównywania struktur. Wyraża jak bardzo dwie struktury są do siebie podobne. RMSD jest liczone przy pomocy następującego wzoru:

$$RMSD(X, Y) = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}$$

gdzie X i Y to porównywane struktury, $(x_i - y_i)$ to dystans pomiędzy kolejnymi, odpowiadającymi sobie atomami w tych strukturach, a n to liczba atomów. Typowo wartość RMSD wyraża się w Angstrmach [Å].

W niniejszej pracy miara RMSD została wykorzystana do porównania przewidzianych jonów metali oraz pól ligandów ze strukturami referencyjnymi podczas fazy testowania programów.

Sprawdzian krzyżowy

Sprawdzian krzyżowy (ang. *cross validation*) jest metodą statystyczną, polegającą na podziale zbioru testowego na podzbiory, a następnie wykorzystaniu części z tych podzbiorów jako zbioru uczącego, na którym zostaje wytrenowana metoda, a pozostałej części zbiorów jako zbioru testowego dla tej metody.

Podczas testowania programu MetalionRNA zastosowana została 5-krotna walidacja, co oznacza, że zbiór struktur został podzielony na pięć równych podzbiorów. Następnie kolejno każdy z nich był brany jako zbiór testowy, a pozostałe cztery tworzyły zbiór uczący do wyprowadzenia potencjału statystycznego.

Program LigandRNA natomiast był testowany w taki sposób, że zbiór kompleksów RNA z ligandami został podzielony na podzbiory według struktury chemicznej ligandów. Następnie wybrany klaster był usuwany ze zbioru struktur, a struktury z pozostałych klastrów tworzyły zbiór uczący. W kolejnym kroku program LigandRNA oceniał pozycje ligandów wygenerowane przy pomocy programu Dock6 (patrz: Rozdział 3.3 „Dock6”) dla wybranych struktur z usuniętego wcześniej klastra. Procedura ta miała na celu zapewnienie, że te same ligandy (bądź podobne) nie znajdują się jednocześnie w zbiorze uczącym i zbiorze testowym.

Krzywa ROC

Krzywe ROC (krzywe charakterystyki odbiornika; ang. *Receiver Operating Characteristic*) w niniejszej pracy posłużyły do oceny jakości wyników sprawdzianu krzyżowego dla programu MetalionRNA.

Program MetalionRNA, na podstawie potencjału statystycznego obliczonego dla całej przestrzeni otaczającej strukturę RNA, przewiduje potencjalne miejsca wiązania się kationów. Na potrzeby omawianego testu program MetalionRNA wygenerował dla każdej testowanej struktury maksymalną liczbę przewidywań tak, by wypełniły one całą przestrzeń wokół niej. Przewidziane jony opisane są oceną - wyliczoną dla nich wartością potencjału statystycznego. Ponadto, poprzez porównywanie przewidywań

Materialy i metody

MetalionRNA z danymi referencyjnymi, można sklasyfikować przewidziane kationy do czterech wymienionych poniżej kategorii:

- wynik prawdziwie dodatni (ang. *True Positive*), oznaczony skrótem TP – jon metalu poprawnie przewidziany jako oddziałujący z RNA;
- wynik fałszywie dodatni (ang. *False Positive*), oznaczony skrótem FP – jon metalu, który w strukturze referencyjnej nie występuje, jednak został przewidziany;
- wynik fałszywie ujemny (ang. *False Negative*), oznaczony skrótem FN – jon metalu, który występuje w strukturze referencyjnej, natomiast nie został przewidziany;
- wynik prawdziwie ujemny (ang. *True Negative*), oznaczony skrótem TN – jon metalu słusznie nieprzewidziany w danym miejscu.

W celu wykreślenia krzywej ROC oblicza się wartość dwóch parametrów:

odsetek poprawnych przewidywań (ang. *True Positive Rate*, TPR)

$$TPR = TP/P$$

oraz odsetek niepoprawnych przewidywań (ang. *False Positive Rate*, FPR)

$$FPR = FP/N$$

dla różnej liczby przewidzianych jonów. Przykładowo, zaczynając od przewidywania o najlepszej ocenie, za każdym razem zwiększa się liczba rozpatrywanych przewidywań o jedno (kolejne co do wartości oceny). Przewidywania znajdujące się w obrębie zadanej odległości (ang. *cut-off distance*) od prawdziwego jonu zostaną oznaczone jako TP, a te które są zbyt daleko jako FP. Liczba jonów znajdujących się w strukturze referencyjnej to wartość pozytywów P (ang. *Positives*), natomiast wartość negatywów N (ang. *Negatives*) to liczba wszystkich wygenerowanych przewidywań, dla danej struktury, pomniejszona o wartość P. W celu liczbowego wyrażenia jakości przewidywania wylicza się też wartość pola powierzchni pod krzywą ROC (ang. *Area Under Curve*, AUC).

3.5. Narzędzia programistyczne

Narzędziem programistycznym wykorzystywanym w badaniach opisanych w niniejszej rozprawie doktorskiej jest język programowania Python (<http://www.python.org/>). Jest to język skryptowy oparty o programowanie obiektowe. Jego ważną cechą jest niezależność od systemu operacyjnego – napisany w nim program będzie działał na każdej platformie, dla której dostępne są interpretery języka Python. Ponadto, istnieje wiele bibliotek rozszerzających jego podstawową funkcjonalność. W realizacji niniejszego projektu zastosowano specjalistyczny pakiet bioinformatyczny – BioPython (Cock, Antao i wsp. 2009), ułatwiający pracę z danymi biologicznymi na poziomie struktury przestrzennej.

3.6. Wizualizacja wyników

Ryciny struktur kompleksów RNA z jonami metali i ligandami zostały wykonane przy użyciu programu: PyMol (DeLano 2002) (<http://www.pymol.org/>).

Wszystkie grafy i wykresy wykonane zostały przy użyciu programu do rysowania wykresów Gnuplot (<http://www.gnuplot.info/>) oraz programu do obróbki graficznej Gimp (<http://www.gimp.org/>).

3.7. Sprzęt komputerowy

Tabela 10 przedstawia sprzęt komputerowy wykorzystany do stworzenia i testowania programów MetalionRNA oraz LigandRNA.

Komputer PC	32 bit, 4 rdzenie: Intel® Core™ 2.83GHz, 8G RAM, Ubuntu 12.10, i386.
Klaster komputerowy Reef	64 bit, 6176 węzłów obliczeniowych, sumaryczna moc obliczeniowa: 54.27 TFLOPS, wielkość pamięci systemowej: 10.208 TB, Scientific Linux CERN SLC release 5.6 (Boron).

Tabela 10. Sprzęt komputerowy.

4. Wyniki

4.1. Anizotropowy potencjał statystyczny

W ramach niniejszego projektu zostały stworzone potencjały statystyczne opisujące oddziaływania RNA z jonami metali i ligandami. Obydwa potencjały powstały na tej samej bazie teoretycznej opisanej szerzej w Rozdziale 3.1 „Potencjał statystyczny”, różnią się jedynie pod względem zdefiniowanych typów: par atomów RNA i atomów liganda.

Oddziaływania RNA z jonami metali

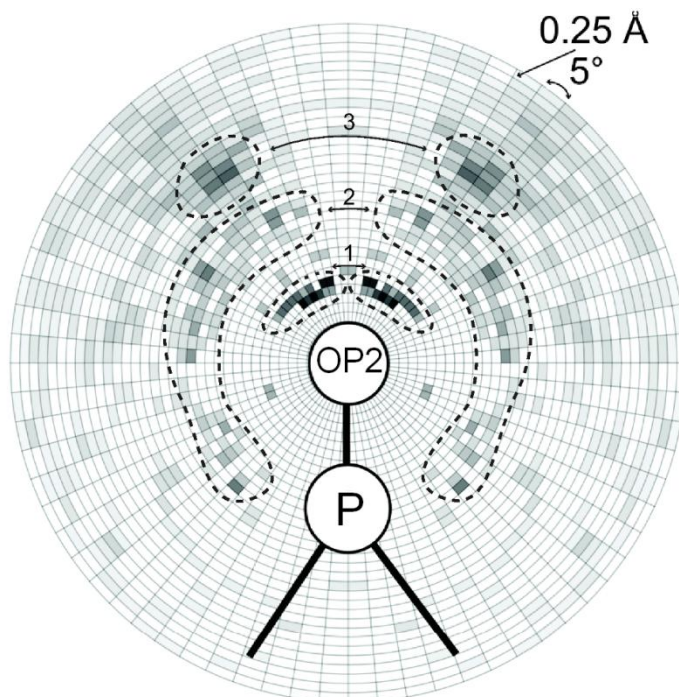
W przypadku oddziaływań RNA z jonami metali kluczową rolę odgrywają atomy tlenu i azotu, dlatego też to one zostały zdefiniowane jako atomy „oddziałujące” b w parze atomów RNA $[a, b]$. Zdefiniowane pary prezentuje Tabela 11. Atomy a z pary atomów RNA są z kolei kowalencyjnie związane z b i służą do wyznaczania kąta (a, b, c), gdzie c to jon metalu: Mg^{2+} , K^+ lub Na^+ .

Łańcuch główny	Łańcuch boczny adeniny	Łańcuch boczny guaniny	Łańcuch boczny cytozyny	Łańcuch boczny uracylu
P, OP1	C2, N1	C2, N1	C2, N3	C2, N3
P, OP2	C2, N3	C2, N2	C2, O2	C2, O2
P, O5'	C4, N3	C2, N3	C4, N3	C4, N3
C1', O4'	C5, N7	C4, N3	C4, N4	C4, O4
C2', O2'	C6, N1	C5, N7		
C3', O3'	C6, N6	C6, N1		
C4', O4'	C8, N7	C6, O6		
C5', O5'		C8, N7		

Tabela 11. Pary atomów RNA zdefiniowane do wyprowadzenia potencjału RNA z jonami metali.

Przykładowy rozkład potencjału wokół pary atomów RNA [P, OP2] dla oddziaływań z Mg^{2+} został już zaprezentowany w Rozdziale 3.1 „Potencjał statystyczny” (Rycina 5, Panel B). Poniżej jeszcze raz został przedstawiony ten sam potencjał (Rycina 6), lecz zaznaczono na nim trzy obszary o najniższej wartości potencjału (po dwa razy każdy, ponieważ rozkład jest symetryczny). Warto zauważyć, że obszary te reprezentują trzy możliwe stany przyjmowane przez Mg^{2+} związany

z RNA (Draper 2004), (Rozdział 1.1 „Charakterystyka oddziaływań RNA z Mg^{2+} ”). (1) Pierwszy obszar odpowiada chelatowanemu kationowi Mg^{2+} (2) Drugi wyróżniony obszar odpowiada stanowi wiązania, w którym pośredniczy jedna warstwa cząsteczek wody, natomiast (3) obszar reprezentuje sytuację, w której kation Mg^{2+} pozostaje uwodniony i oddziałuje z RNA poprzez więcej niż jedną warstwę cząsteczek wody.



Rycina 6. Potencjał statystyczny dla pary atomów RNA [P, OP2] i kationu Mg^{2+} . Trzy zaznaczone obszary (po dwa razy każdy, ponieważ wykres jest symetryczny) odpowiadają trzem możliwym stanom które przyjmuje Mg^{2+} związany z RNA (patrz: Rozdział 1.1 „Charakterystyka oddziaływań RNA z Mg^{2+} ”(Draper 2004). Źródło: (Philips, Milanowska i wsp. 2012).

Potencjał statystyczny, wyprowadzony dla pozostałych par atomów RNA (Tabela 11) został dołączony do niniejszej rozprawy na płycie CD w postaci plików tekstowych (patrz: „Załączniki”). Przyjęto następującą konwencję nazewnictwa plików: łańcuch główny/reszta nukleotydu (B – łańcuch główny (ang. *Backbone*), A – adenina, G - guanina, C – cytozyna, U – uracyl); typ pary atomów RNA (atom *a*, atom *b*); typ kation (MG, K, NA). Np. B_C1',_O4'_NA.txt. Format pliku z potencjałem uzyskanym dla pary atomów RNA [C1', O4'] i Na^+ demonstruje Rycina 7. W pierwszej i drugiej kolumnie znajdują się kolejno wartości odległości *d* i kąta *α* opisujące przedziały przestrzeni, natomiast w trzeciej kolumnie znajdują się odpowiadające

Wyniki

im wartości potencjału. Zarówno nazewnictwo plików jaki i ich format sprawiają, że potencjał jest łatwy w implementacji w programach do dokowania oraz możliwa jest jego wizualizacja przy użyciu dowolnego narzędzia do generowania grafów czy wykresów (Np. Mathematica, MatLab czy Gnuplot).

```
----- B_C4'_O4'_NA.txt -----  
  
2.875  95.0  0.0176736547563  
3.125  50.0  0.0150604973256  
3.125  65.0  0.0150604973256  
3.375  55.0  0.0259734373767  
3.375  60.0  0.0129867186883  
3.625  65.0  0.0113134946589  
3.875  60.0  0.0198878939627  
3.875  100.0  0.00994394698134  
4.125  5.0  0.00880880630546
```

Rycina 7. Przykładowy fragment pliku z potencjałem statystycznym wyprowadzonym dla pary atomów RNA [C1', O4'] z łańcucha głównego oraz jonu Na⁺. W pierwszej kolumnie znajduje się wartość odległości wyrażona w Angstromach, w drugiej wartość kąta wyrażona w stopniach, natomiast trzecia kolumna zawiera wartość potencjału dla tego przedziału przestrzeni.

Oddziaływania RNA z ligandami

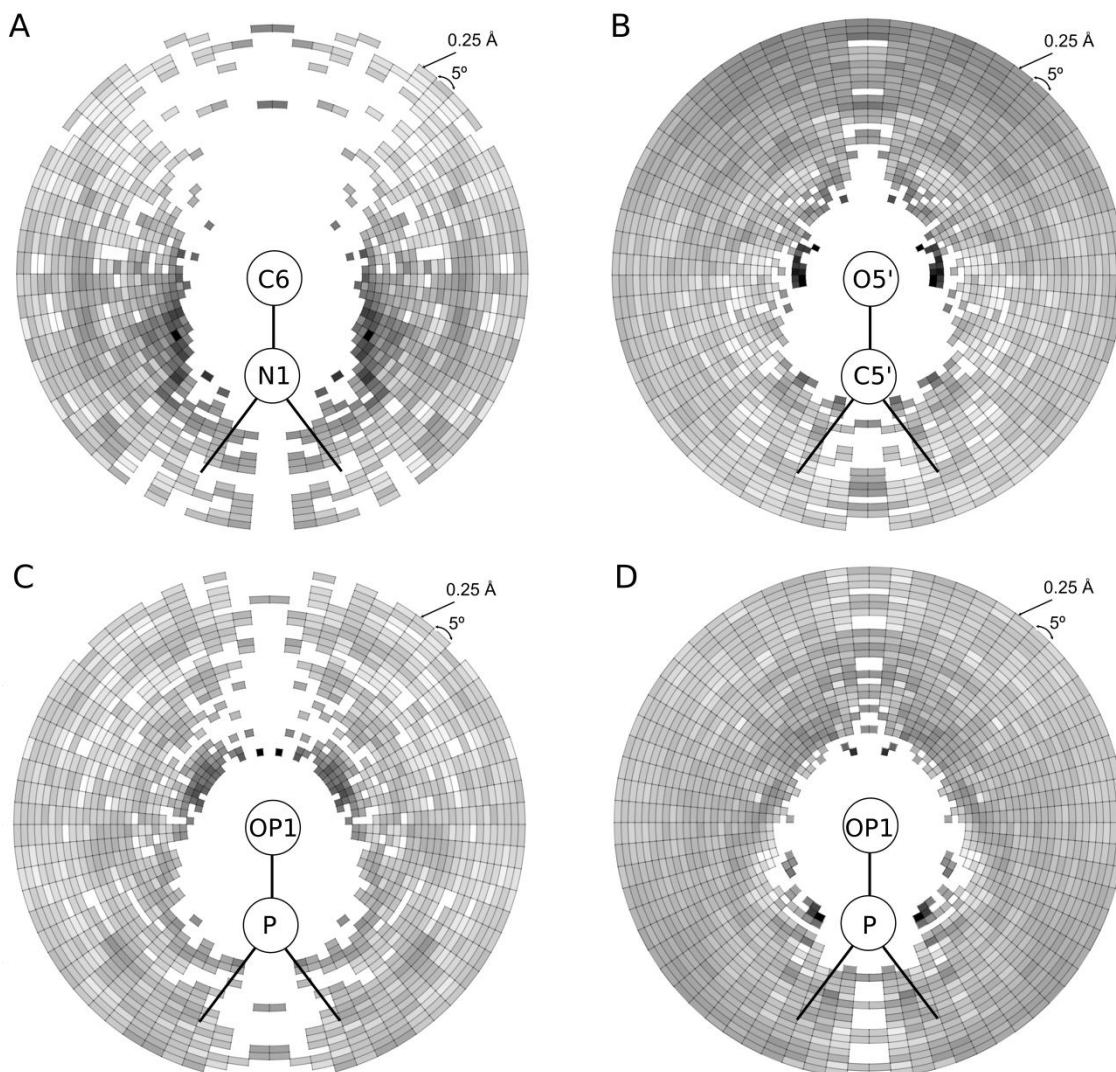
W przypadku oddziaływań RNA z ligandami wszystkie atomy RNA są istotne, ze względu na różnice w strukturach chemicznych poszczególnych ligandów z nim oddziałujących. Dlatego też lista par RNA (przez co, atomów *b* zdefiniowanych jako „oddziałujące”) została rozszerzona (patrz: Tabela 12). Atomy ligandów *c* sklasyfikowano według następującego podziału: węgiel sp (C.1), węgiel sp² (C.2), węgiel sp³ (C.3), węgiel w pierścieniach aromatycznych (C.ar), węgiel w ugrupowaniach amidynowych i guanidynowych (C.cat), azot sp (N.1), azot sp² (N.2), azot sp³ (N.3), naładowany dodatnio azot sp³ (N.4), azot w pierścieniach aromatycznych (N.ar), azot w wiązaniach amidowych (N.am), azot w ugrupowaniach amidynowych i guanidynowych (N.pl3), tlen sp² (O.2), tlen sp³ (O.3), tlen w grupach karboksylowych i fosforanowych (O.co2), siarka sp² (S.2), siarka sp³ (S.3), sulfony (S.O2), fosfor sp³ (P.3), fluor (F), chlor (Cl). Podział ten został dokonany według notacji typów atomów Tripos (SYBYL Molecular Modeling Software, 7.3; Tripos Inc.: St. Louis, MO, 2006).

Łańcuch główny	Łańcuch boczny adeniny	Łańcuch boczny guaniny	Łańcuch boczny cytozyny	Łańcuch boczny uracylu
C2', C1'	N3, C2	N3, C2	N3, C2	N3, C2
C3', C2'	C4, N3	C4, N3	C4, N3	C4, N3
C4', C3'	C5, C4	C5, C4	C5, C4	C5, C4
C5', C4'	C6, C5	C6, C5	C6, C5	C6, C5
O5', C5'	N1, C6	N1, C6	N1, C6	N1, C6
O5', P	C2, N1	C2, N1	C2, N1	C2, N1
P, OP2	C6, N6	C6, O6	C2, O2	C2, O2
P, OP2	C8, N7	C8, N7	C4, N4	C4, O4
C5', O5'	N9, C8	N9, C8		
C4', O4'	C4, N9	C4, N9		
C3', O3'		C2, N2		
C2', O2'				

Tabela 12. Pary atomów RNA zdefiniowane do wyprowadzenia potencjału RNA z ligandami.

Rycina 8 prezentuje rozkład wartości potencjału statystycznego dla wybranych typów oddziaływań. Panel A przedstawia preferowane geometrie wiązania pary atomów RNA [N1, C6] i atomu liganda typu C.ar. W przypadku tych oddziaływań, występują głównie oddziaływania dalekie, zależne od orientacji przestrzennej. Diagram ukazuje, że węgle w pierścieniach aromatycznych liganda najczęściej znajdują się w odległości 3,5 Å i pod kątem 90-180° od pary atomów RNA [N1, C6]. Białe pole otaczające atomy RNA wskazuje na brak występowania oddziaływań tego typu w zbiorze uczącym potencjału. Rozkład potencjału dla pary atomów RNA [C5', O5'] (Panel B) posiada dobrze zdefiniowaną parę minimów – najciemniejsze pola w odległości od atomu RNA O5' ~2,5 Å i pod kątem 60~90° – odpowiadającą preferowanym pozycjom zajmowanym przez atom liganda typu O.3. Warto też zauważyć, że większość powierzchni wokół atomów RNA zaznaczona jest ciemnoszarym kolorem, co wskazuje na wysokie powinowactwo atomów tego typu do tworzenia oddziaływań polarnych. Panel C prezentuje rozkład potencjału dla pary atomów RNA [P, OP1] i atomu liganda typu N.3. W przypadku tej pary, minimum występuje w bliskiej odległości od atomów RNA, tj. ~2,5 Å i pod kątem 20~60°. Panel 4 natomiast przedstawia rozkład potencjału pary [P, OP1] i C.3. Podobnie jak w przypadku rozkładu potencjału dla pary [C5', O5'] i O.3 (Panel B) wskazuje on na silne oddziaływania występujące między atomami tej pary, ze słabą zależnością przestrzenną (prawie całe pole wokół atomów liganda zaznaczone jest na ciemnoszary).

Wyniki



Rycina 8. Rozkład potencjału dla par: Panel A) [N1, C6] i C.ar. Panel B) [C5', O5'] i O.3. Panel C) [P, OP1] i N.3. Panel D) [P, OP1] i C.3.

Potencjał statystyczny został dołączony do niniejszej rozprawy na płycie CD w postaci plików tekstowych (patrz: „Załączniki”). Konwencja nazewnictwa plików jest taka sama jak w przypadku potencjału stworzonego dla oddziaływań RNA z kationami tzn.: łańcuch główny/reszta nukleotydu (B – łańcuch główny, A – adenina, G - guanina, C - cytozyna, U – uracyl); typ pary atomów RNA (atom *a*, atom *b*); typ atomu liganda (np. N.3). Np. B_C1',_O4'_N.3.txt. Format pliku z potencjałem również nie uległ zmianie, w pierwszej i drugiej kolumnie znajdują się odpowiednio wartości odległości *d* i kąta α opisujące przedziały przestrzeni, natomiast w trzeciej kolumnie znajdują się odpowiadające im wartości potencjału.

4.2. Program MetalionRNA

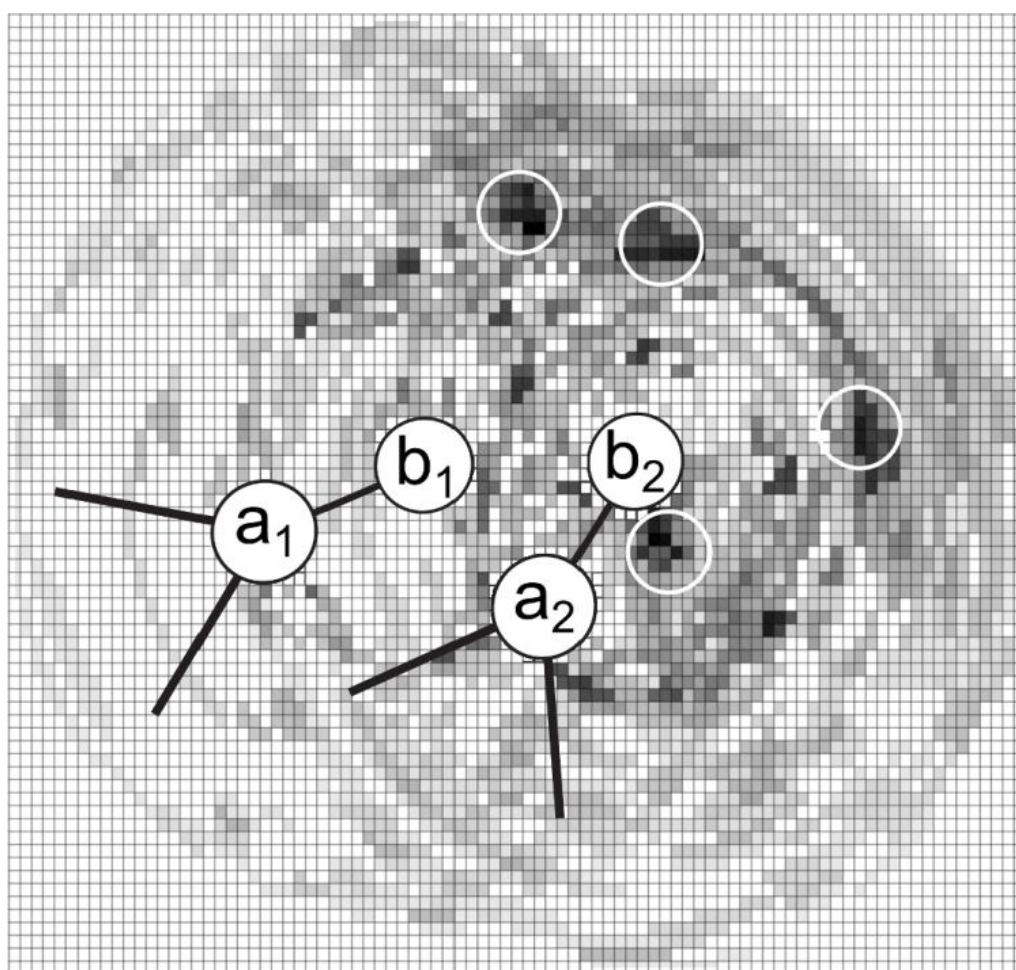
Charakterystyka programu MetalionRNA

Program MetalionRNA służy do przewidywania miejsc wiązania się jonów metali w strukturach przestrzennych RNA. Możliwe jest przewidywanie trzech typów jonów: magnezu, sodu i potasu.

MetalionRNA jako dane wejściowe przyjmuje plik ze strukturą RNA (w formacie PDB lub MOL2), następnie rozpina sieć sześcienną wokół tej cząsteczki. Zastosowanie sieci do dyskretyzacji przestrzeni (ang. *grid base approach*) pozwala na mapowanie danych statystycznych na dobrze zdefiniowane punkty w przestrzeni. Podejście to było już z sukcesem zastosowane w przewidywaniu oddziaływań białek z ligandami, na przykład w programie Dock6 czy AutoDock. Odległość między węzłami tej sieci wynosi 0,25 lub 0,5 Å. Wartości te zostały tak dobrane, by minimalna odległość pomiędzy atomem RNA a jonem metalu była przynajmniej jednokrotnie większa. By uniknąć niepotrzebnych obliczeń, sieć jest rozpinana jedynie w odległości 9 Å od każdej zdefiniowanej pary atomów RNA [*a*, *b*] znalezionej w zadanej strukturze. Jeżeli w strukturze występują modyfikowane reszty nukleotydów, to MetalionRNA najpierw sprawdza jakiej reszty jest to modyfikacja, a następnie analizuje wszystkie pary atomów [*a*, *b*] zdefiniowane dla niemodyfikowanej reszty tego typu. Jeśli dana para atomów nie występuje w modyfikowanej reszcie, jest ona pomijana. Dodatkowo, węzły sieci, które „zachodzą” na atomy RNA są wykluczane z dalszych obliczeń. Następnie, program przypisuje węzłom sieci wartości potencjału statystycznego $W^{(n)}$. Potencjał dla węzłów sieci w odległości <9 Å od więcej niż jednej pary atomów RNA jest sumą potencjałów na niego się składających (sytuację taką obrazuje Rycina 9). Następnie, po zakończeniu tego procesu, MetalionRNA bierze węzeł o najlepszym potencjale statystycznym (czyli o najniższej wartości) i do wartości jego potencjału dodaje wartości potencjału węzłów znajdujących się w promieniu rzeczywistego kationu (np. w przypadku Mg^{2+} promień ten wynosi 0,75 Å). Suma ta jest jednocześnie oceną przewidywania. Jeśli promień przewidzianego jonu metalu nie zachodzi na wcześniejsze przewidywanie, o lepszej ocenie, MetalionRNA umieszcza w centralnym węźle środek nowo przewidzianego kationu i usuwa wszystkie węzły sieci, które posłużyły do jego wyznaczenia. Proces ten jest powtarzany aż do uzyskania 100 przewidywań. Następnie wybierane są przewidywania najlepiej ocenione, w liczbie zdefiniowanej przez użytkownika lub wyznaczonej wartości domyślnej. Zapewnia to,

Wyniki

że zostaną wybrane przewidywania o najlepszych ocenach całkowitych (najniższych sumach potencjału), a nie o najniższych wartościach potencjału węzła centralnego. Liczba domyślna przewidywań jest obliczana na podstawie liczby reszt nukleotydowych w danej strukturze RNA. Średnia liczba kationów danego typu przypadająca na resztę została wyznaczona przy użyciu zbioru 113 kompleksów RNA z kationami, użytych wcześniej do wytrenowania potencjału statystycznego (Rozdział 3.2 „Przygotowanie zbiorów uczących i testowych”). Wartość domyślna to średnia liczba jonów metalu przypadająca na daną strukturę plus jeden.



Rycina 9. Schematyczna ilustracja sposobu obliczania potencjału statystycznego dla węzłów sieci w odległości $< 9 \text{ \AA}$ od dwóch par atomów RNA [P, OP2] i identyfikacja najbardziej prawdopodobnych miejsc wiązania jonu metalu. Źródło: (Philips, Milanowska i wsp. 2012).

Dodatkową funkcjonalnością programu jest możliwość obliczenia (i wyświetlenia w przeglądarce struktur) powierzchniowego potencjału statystycznego dla cząsteczki RNA. Dla każdego ciężkiego atomu RNA MetalionRNA zlicza wartości potencjałów z poszczególnych węzłów sieci znajdujących się w odległości do 2,5 Å od niego i przypisuje mu wartość tej sumy – powierzchniowego potencjału statystycznego. W wynikowym pliku PDB ze strukturą wartości czynnika temperaturowego (ang. *b-factor*) MetalionRNA zastępuje wartościami powierzchniowego potencjału statystycznego, wyliczonymi wcześniej dla każdego ciężkiego atomu. Dzięki temu użytkownik ma możliwość wyświetlenia powierzchniowego potencjału statystycznego poprzez użycie opcji wyświetlenia wartości „*b-factor*” w dowolnej przeglądarce struktur.

Ocena działania programu – sprawdzian krzyżowy, krzywe ROC

W celu oceny działania programu MetalionRNA został przeprowadzony 5-krotny sprawdzian krzyżowy, a otrzymane wyniki zostały zilustrowane w postaci krzywych ROC. Do testów zostały użyte trzy zestawy struktur RNA (zawierające jeden z trzech najczęściej oddziałujących z nimi kationów: magnezu, sodu i potasu). Dodatkowo walidacja przeprowadzona została na strukturach DNA oddziałujących z kationami magnezu przy użyciu potencjału wyprowadzonego dla oddziaływań RNA z tymi kationami. Wszystkie obliczenia MetalionRNA były prowadzone dwukrotnie – dla siatki o rozdzielczości 0,25 Å i 0,5 Å. Promień atomu magnezu ustalony został na 0,75 Å, atomu sodu na 1,0 Å a potasu na 1,5 Å.

Rycina 10 przedstawia krzywe ROC wykreślone dla przewidywań oddziaływań RNA z Mg^{2+} . Każda krzywa reprezentuje inną wartość progu odcięcia (ang. *cut-off*), tzn. maksymalnego dystansu dzielącego przewidywanie od prawdziwego kationu, dla którego to przewidywanie jest uznawane za wynik poprawny (prawdziwie dodani TP). Wartości tych odległości zostały ustalone odpowiednio na 0,72 Å (promień jonowy magnezu – linia ciągła, czerwona), 0,75; 1,0; 1,5; 2,0 i 3,0 Å. Obliczone pola powierzchni pod krzywymi (AUC) otrzymanymi dla nich wynoszą: 50%, 56%, 81%, 93%, 95%, 96% dla siatki o odległościach między węzłami równymi 0,25 Å; i 62%, 62%, 81%, 95%, 96%, 97% dla siatki o odległościach między węzłami równymi 0,5 Å.

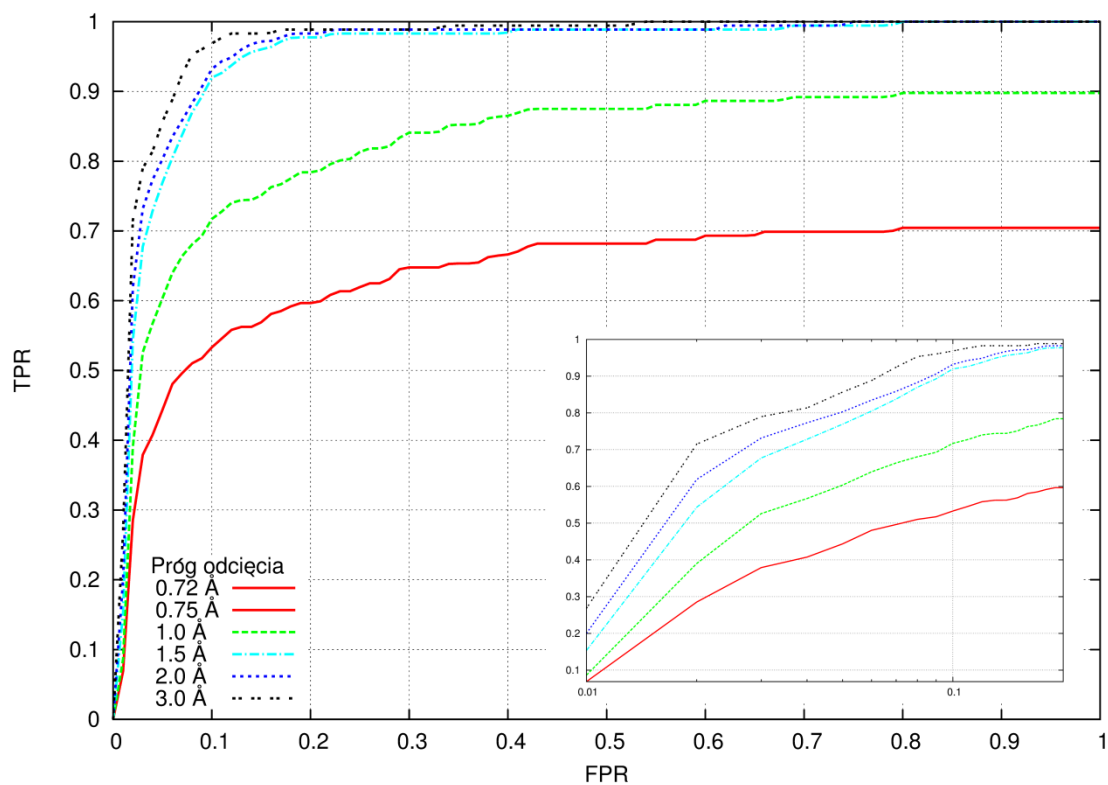
Rycina 11 przedstawia krzywe ROC wykreślone dla przewidywań oddziaływań RNA z sodem. Wartości maksymalnego dystansu dzielącego przewidywanie

Wyniki

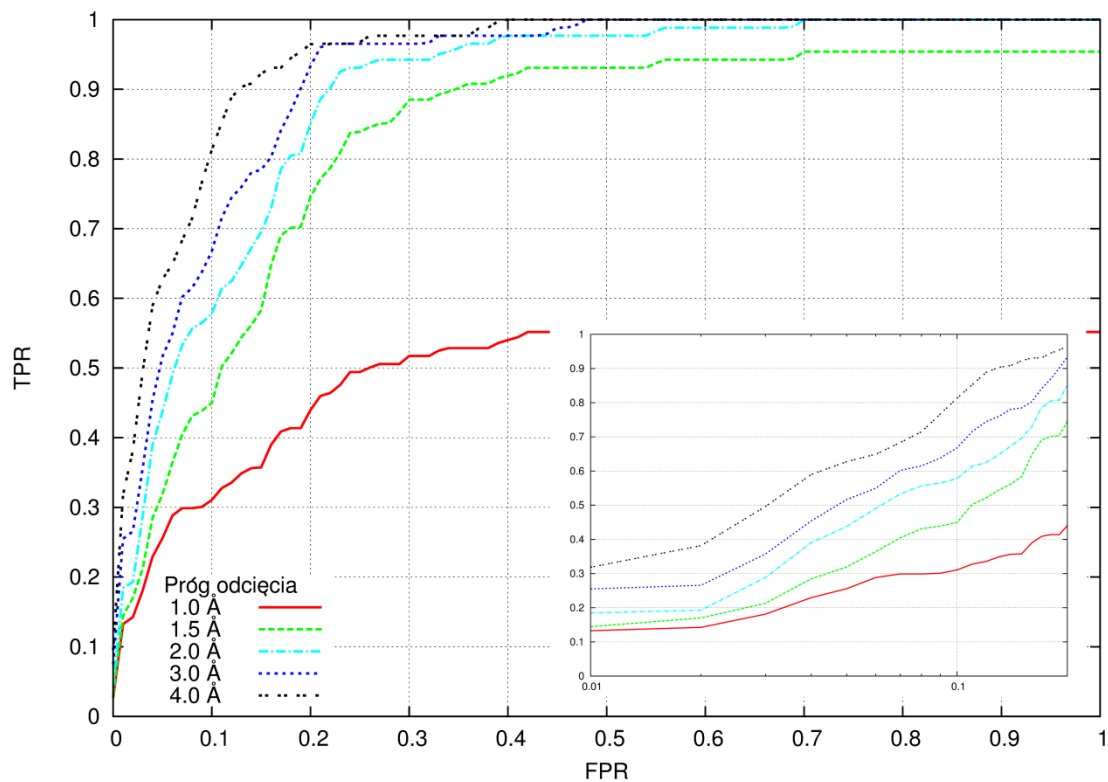
od prawdziwego kationu, dla której to przewidywanie jest uznawane za wynik poprawny wynoszą: 1,0 (promień jonowy sodu – linia ciągła, czerwona); 1,5; 2,0; 3,0 i 4,0 Å, a odpowiadające im wartości AUC: 43%, 82%, 87%, 91%, 93% (siatka 0,25 Å) i 47%, 78%, 85%, 88%, 91% (siatka 0,5 Å).

Krzywe ROC dla oddziaływań RNA z jonami potasu przedstawia Rycina 12. Otrzymane wartości AUC pól powierzchni pod krzywymi dla progów odcięcia: 1,38; 1,5; 2,0; 3,0 i 4,0 Å wynoszą odpowiednio: 54%, 61%, 84%, 96%, 97% dla siatki o rozdzielczości 0,25 Å i 54%, 61%, 81%, 97%, 98%; dla siatki o rozdzielczości 0,5 Å. Promień jonowy potasu wynosi 1,38 Å.

Przewidywania miejsc wiązania jonów magnezu w strukturach DNA, przy pomocy potencjału wyprowadzanego z kompleksów RNA z Mg^{2+} , podsumowuje Rycina 13. Powierzchnia pola pod krzywymi ROC dla progów odcięcia: 0,72, 0,75; 1,0; 1,5; 2,0 i 3,0 Å wynosi odpowiednio: 44%, 49%, 74%, 90%, 91%, 93% (siatka 0,25 Å) i 56%, 56%, 72%, 88%, 91%, 93% (siatka 0,5 Å). Tym samym test ten dowiódł, że MetalionRNA generuje poprawne wyniki dla struktur DNA, które nie były zawarte w zbiorze uczącym potencjału. Krzywe wygenerowane dla zestawu struktur DNA są jedynie nieznacznie gorsze od tych otrzymanych dla przewidywań Mg^{2+} w strukturach RNA.

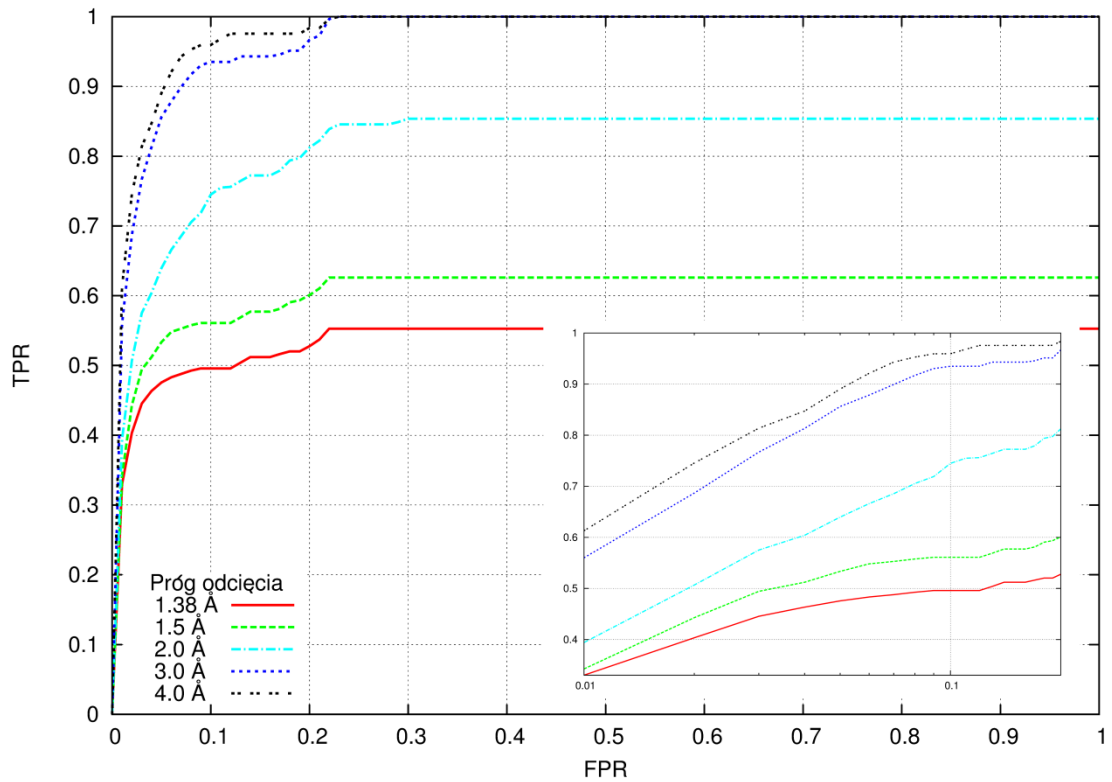


Rycina 10. Krzywe ROC dla zestawu struktur RNA z Mg^{2+} . Mały wykres przedstawia zakres od 0 do 0,2 Å całego wykresu w skali logarytmicznej.

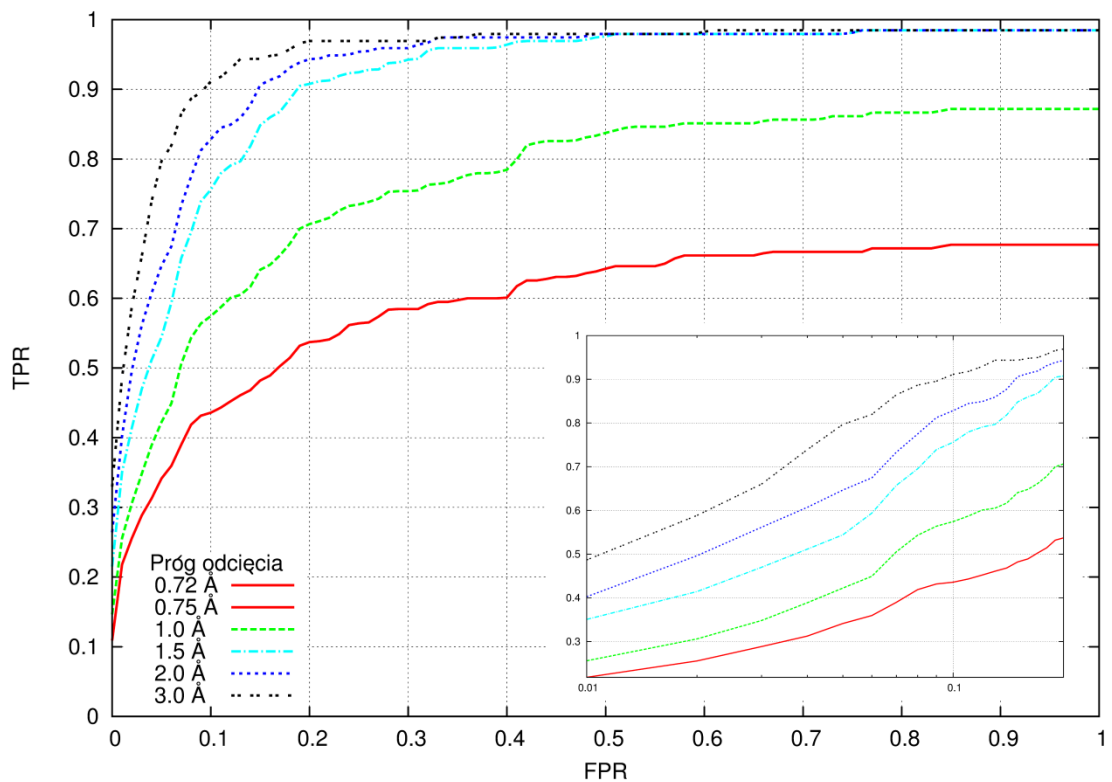


Rycina 11. Krzywe ROC dla zestawu struktur RNA z Na^{+} . Mały wykres przedstawia zakres od 0 do 0,2 Å całego wykresu w skali logarytmicznej.

Wyniki



Rycina 12. Krzywe ROC dla zestawu struktur RNA z K^+ . Mały wykres przedstawia zakres od 0 do 0,2 Å całego wykresu w skali logarytmicznej.



Rycina 13. Krzywe ROC dla zestawu struktur DNA z Mg^{2+} . Mały wykres przedstawia zakres od 0 do 0,2 Å całego wykresu w skali logarytmicznej.

Ocena działania programu – porównanie z programem FEATURE

Program FEATURE (Rozdział 1.3 „Komputerowe przewidywanie miejsc wiązania RNA z jonami metali”) służy m.in. do przewidywania jonów magnezu w strukturach RNA. Został przetestowany przez swoich autorów na 58-nukleotydowym fragmencie rRNA (podjednostka 23S) pochodzącym z *Bacillus stearotherophilus* (kod PDB: 1HC8) (Liang, Banatao i wsp. 2003), z którym związanych jest siedem jonów Mg^{2+} i jeden K^+ . By porównać wyniki wygenerowane przez FEATURE z wynikami otrzymanymi z programu MetalionRNA, zostały przeprowadzone, programem MetalionRNA, przewidywania Mg^{2+} dla tej struktury, z użyciem:

- oryginalnego potencjału MetalionRNA, który został wytrenowany w ramach niniejszego projektu na 113 kompleksach RNA z jonami metali (Tabela 4),
- potencjału wytrenowanego na zestawie uczącym FEATURE (Tabela 6).

Otrzymane wyniki podsumowuje Tabela 13 i Rycina 14. Program MetalionRNA obliczył, że dla struktury RNA tej wielkości (58 nt.), rozwiązanej w warunkach średniej koncentracji jonów, liczba jonów magnezu z nią związanych powinna wynosić sześć. Stąd też pierwsze sześć przewidywań programu uznane zostały za znaczące. Kolejne przewidywania mogą odpowiadać alternatywnym miejscom wiązania się jonów, potencjalnie zajmowanym np. w warunkach wyższej koncentracji Mg^{2+} . Pośród pierwszych sześciu najlepszych przewidywań przy użyciu oryginalnego potencjału MetalionRNA znalazły się cztery trafne przewidywania. Odpowiadają one pozycjom jonów magnezu znajdujących się w strukturze referencyjnej (kod PDB: 1HC8). Jony te zostały zidentyfikowane z dokładnością od 0,6 do 1,9 Å. Pozostałe trzy jony ze struktury referencyjnej zostały przewidziane na pozycjach: 10, 13 i 29. Używając potencjału wyprowadzonego ze zbioru uczącego programu FEATURE, program MetalionRNA poprawnie przewidział jedynie dwa jony (z dokładnością 0,8 i 0,6 Å) z siedmiu znajdujących się w strukturze referencyjnej. Pozostałym pięciu jonom odpowiadały przewidywania na pozycjach: 8, 9, 21, 29 i 33.

Pośród siedmiu jonów magnezu związanych ze strukturą 1HC8, według definicji zaproponowanej przez twórców FEATURE, dwa z nich związanych jest „silnie”, natomiast pozostałe pięć „słabo”. Program FEATURE posiada osobne funkcje do przewidywania jonów obu rodzajów (patrz: 1.3 „Komputerowe przewidywanie miejsc wiązania RNA z jonami metali”). Dla jonów „silnie” związanych FEATURE uzyskał

Wyniki

dziewięć przewidywań o „liczącej się ocenie”, tzn. powyżej zdefiniowanego progu (Liang, Banatao i wsp. 2003), natomiast dla jonów „słabo” związanych przewidywań było aż 224. FEATURE poprawnie zidentyfikował pozycje dwóch „silnie” związanych jonów magnezu spośród pierwszych sześciu najlepszych przewidywań dla tego rodzaju jonów (pozycje w rankingu: 2 i 5, z dokładnością 3,6 i 1,5 Å). Jony „słabo” związane zostały przewidziane na znacznie późniejszych pozycjach w rankingu jonów tego rodzaju (8, 15, 88, 102, 202).

Program MetalionRNA, przy użyciu obu wariantów potencjału, przewidział cztery spośród pięciu „słabo” związanych Mg^{2+} ze znacznie lepszą dokładnością niż FEATURE. Jedynie w przypadku jonu nr 1160 program FEATURE osiągnął lepszy wynik, jednak dopiero na 17 pozycji w rankingu podczas, gdy program MetalionRNA już na szóstej pozycji w swoim rankingu zidentyfikował ten jon z akceptowalną dokładnością (1,9 Å). Przewidywania pozostałych dwóch „słabo” związanych Mg^{2+} (1161 i 1172) były relatywnie słabe w przypadku obu metod.

Program MetalionRNA przewidział także jeden z dwóch „silnie” związanych Mg^{2+} (*1163) z bardzo dobrą dokładnością oraz na wysokiej pozycji w rankingu. Używając oryginalnego potencjału: dokładność 0,6 Å, pozycja 3; używając potencjału wyprowadzonego ze zbioru uczącego FEATURE: dokładność 0,6 Å, pozycja 4. Dla tego jonu FEATURE również uzyskał dobry wynik: dokładność 1,5 Å, pozycja 5 w osobnym rankingu przewidywań dla jonów „silnie” związanych. Drugi „silnie” związany jon (*1167) został przewidziany przez program MetalionRNA z dokładnością 3,8 Å (pozycja 10) i 1,1 Å (pozycja 8) przy użyciu dwóch wymienionych wcześniej potencjałów. FEATURE przewidział ten jon z dokładnością 3,6 Å w pozycji 2.

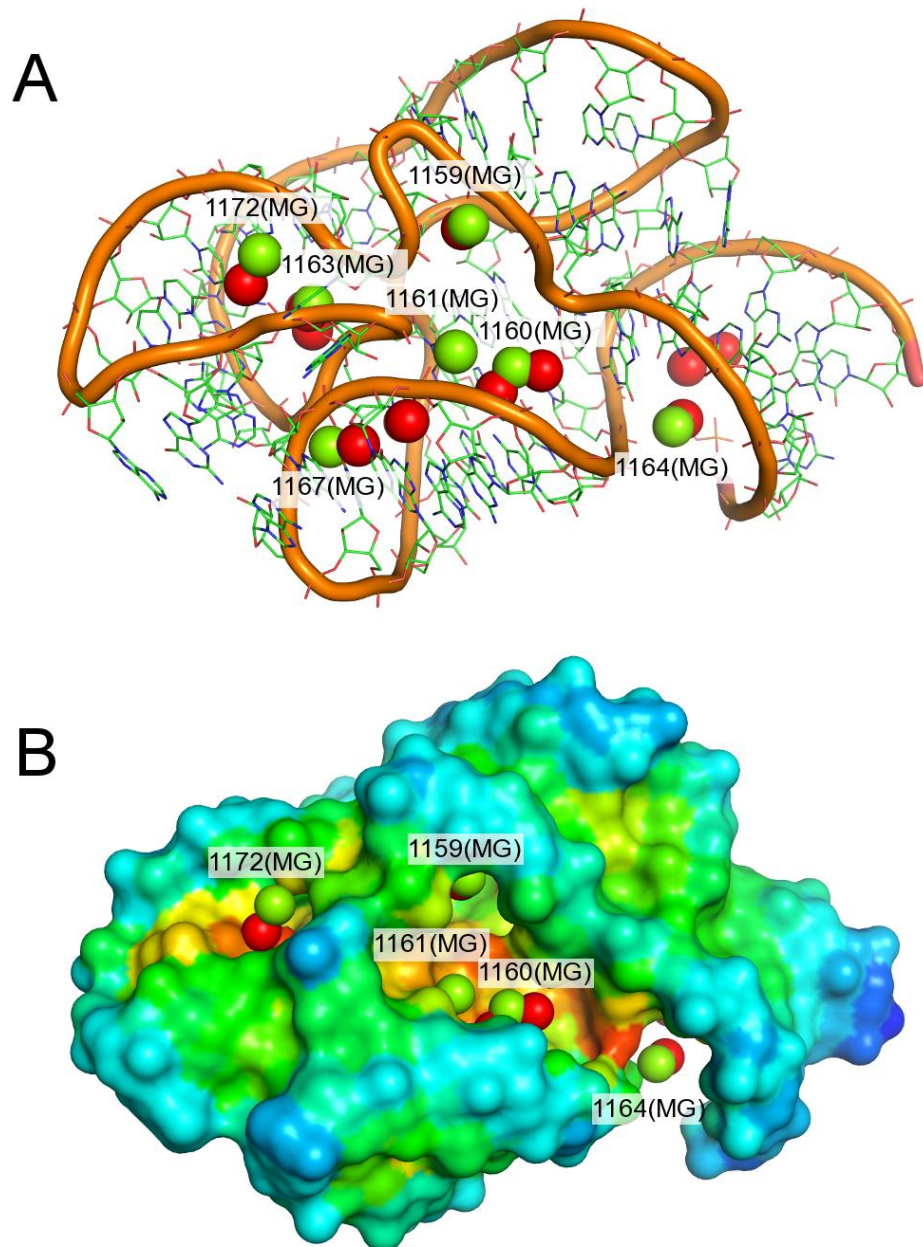
Podsumowując, spośród pierwszych sześciu przewidywań MetalionRNA poprawnie zidentyfikował cztery z siedmiu jonów Mg^{2+} znajdujących się w strukturze referencyjnej. Jedynie dwa przewidywania, okazały się nieprawdziwe. Podczas, gdy FEATURE zidentyfikował te jony ze znacznie większą liczbą niepoprawnych przewidywań.

Dodatkową ciekawostką jest fakt, że najlepiej ocenione przewidywanie jonu Mg^{2+} z rankingu FEATURE (dokładność 1,4 Å) odpowiada miejscu wiązania się K^+ w tej strukturze (kod PDB: 1HC8, numer atomu w strukturze: *1162). Program

MetalionRNA zidentyfikował ten jon na pozycji czwartej (dokładność 2,2 Å) w rankingu przewidywań przeprowadzonych dla jonów potasu, natomiast trzy pierwsze przewidywania pokrywały się z miejscami wiązania się Mg^{2+} . Jeśli wziąć pod uwagę ranking MetalionRNA dla przewidywań jonów magnezu, miejsce wiązania się K^+ odpowiada przewidywaniu na pozycji 18 (dokładność 1,7 Å).

Mg²⁺ (nr atomu)	MetalionRNA		FEATURE
	Zbiór uczący MetalionRNA	Zbiór uczący FEATURE	
1159	0,8 Å (1)	0,8 Å (1)	1,6 Å (15), 0,4 Å (58)
1160	1,9 Å (6)	1,7 Å (33)	0,6 Å (17), 2,0 Å (8)
1161	2,9 Å (29)	3,7 Å (29)	1,5 Å (102)
*1163	0,6 Å (3)	0,6 Å (4)	1,5 Å (*5)
1164	1,4 Å (2)	1,1 Å (21)	0,7 Å (181), 1,6 Å (88)
*1167	3,8 Å (10)	1,1 Å (8)	3,6 Å (*2)
1172	3,2 Å (13)	3,4 Å (9)	2,5 Å (202)

Tabela 13. Porównanie wyników działania programu MetalionRNA (z oryginalnym potencjałem i potencjałem wyprowadzonym ze zbioru uczącego programu FEATURE) z wynikami działania programu FEATURE. Przewidywania były prowadzone dla struktury rRNA z podjednostki 23S (kod PDB: 1HC8). Jony „silnie” związane zostały oznaczone asteryskiem.



Rycina 14. Panel A) Struktura rRNA (kod PDB: 1HC8) z doświadczalnie wyznaczonymi pozycjami jonów Mg²⁺ (zielone sfery z etykietami). Najlepsze przewidywania dla Mg²⁺ programu MetalionRNA ukazane zostały jako czerwone sfery. Panel B) przedstawia powierzchniowy anizotropowy potencjał statystyczny obliczony dla tej struktury. Im kolor bardziej czerwony tym mniejszy potencjał i większe prawdopodobieństwo wiązania się jonu w tym miejscu.

4.3. Program LigandRNA

Charakterystyka programu LigandRNA

Program LigandRNA służy do oceny siły wiązania różnych pólz liganda/ów, w strukturach RNA, wygenerowanych wcześniej przy użyciu zewnętrznego programu do dokowania (np. Dock6). LigandRNA jako dane wejściowe przyjmuje plik z przestrzenną strukturą RNA (w formacie PDB lub MOL2) oraz plik z pozami liganda/ów (MOL2). Następnie rozpina sieć wokół struktury RNA i w identyczny sposób jak MetalionRNA wyznacza wartości potencjału statystycznego dla poszczególnych węzłów tej sieci. Jedyna różnica polega na tym, że program LigandRNA oblicza dla każdego węzła kilka wartości potencjału, które odpowiadają różnym typom atomów liganda. By ograniczyć ilość obliczeń LigandRNA najpierw sprawdza, jakie typy atomów zawiera ligand (lub ligandy) i tylko dla nich przeprowadza obliczenia. Następnie program przystępuje do oceny zadanych pólz liganda/ów. Ocena pozy liganda to suma ocen tworzących go atomów. Ocena atomu z kolei to suma potencjałów węzłów sieci znajdujących się w promieniu tego atomu.

Podobnie jak program MetalionRNA, program LigandRNA posiada możliwość obliczenia (i wyświetlenia w odpowiedniej przeglądarce struktur) powierzchniowego potencjału statystycznego cząsteczki RNA. Program LigandRNA oblicza wartości sumy tego potencjału dla wszystkich typów atomów liganda, oraz dla następujących typów atomów osobno: O, C i N.

Ocena działania programu LigandRNA

W celu oceny działania programu LigandRNA kompleksy RNA z ligandami zostały sklasteryzowane według struktury chemicznej liganda, a następnie został przeprowadzony sprawdzian krzyżowy. Więcej na temat przygotowania zestawu uczącego i testowego oraz wspomnianej procedury znajduje się w Rozdziałach 3.2 „Przygotowanie zbiorów uczących i testowych” i 3.4 „Metody i miary użyte do ewaluacji programów”.

Ponadto, by rzetelnie porównać skuteczność potencjałów LigandRNA z DrugScoreRNA, taka sama procedura została zastosowana również do przetestowania tego potencjału. DrugScoreRNA został wpieryw zaimplementowany przez autorkę niniejszej rozprawy doktorskiej. Następnie potencjał ten został wytrenowany na tych

Wyniki

samych zestawach uczących, co LigandRNA i użyty do oceny póż ligandów w kompleksach z zestawu testowego. W teście uwzględniono także wyniki programu Dock6, który został użyty do wygenerowania póż ligandów dla kompleksów z zestawu testowego, ponieważ podczas tego procesu przypisał on każdej pozie swoją ocenę. Ponadto, pozy ligandów z zestawu testowego były również oceniane przy pomocy kombinacji LigandRNA i Dock6.

Tabela 16 przedstawia szczegółowe wyniki przeprowadzonego testu – wartości RMSD póż zidentyfikowanych jako najlepsze przez dane metody w zestawie testowym 42 kompleksów. Program LigandRNA znalazł pozę o najlepszej wartości RMSD do struktury referencyjnej ($i < 3 \text{ \AA}$) spośród wszystkich metod w 13 przypadkach. DrugScoreRNA, z użyciem potencjału wytrenowanego na zbiorze uczącym LigandRNA, w 5 przypadkach, Dock6 w 8, a kombinacja metod LigandRNA-Dock6 znalazła najlepsze rozwiązanie aż w 16 przypadkach.

Ponieważ zbiór testowy LigandRNA pokrywa się w dużym stopniu ze zbiorami testowymi innych metod opisanych w Rozdziale 1.3 „Komputerowe przewidywanie oddziaływań RNA z ligandami” (RiboDock, funkcja oceny dla aminoglikozydów, oryginalny DrugScoreRNA i MORDOR) było możliwe ich porównanie. Tabela 16 prezentuje również wyniki, opublikowane przez autorów tych metod, dla podzbiorów struktur wspólnych dla zbiorów testowych LigandRNA i danej metody. Program RiboDock znalazł najlepsze rozwiązanie tylko w 1 z 10 analizowanych przypadków, funkcja oceny dla aminoglikozydów zwróciła najlepsze rozwiązania w 2 przypadkach z 6; oryginalny DrugScoreRNA w 4 z 21 a MORDOR w 10 z 32 przypadków.

Tabela 14 i Tabela 15 prezentują procent poprawnie zidentyfikowanych póż przez LigandRNA, DrugScoreRNA, Dock6 i kombinację LigandRNA-Dock6. Poprawnie zidentyfikowana poza oznacza, że jej wartość RMSD do struktury natywnej wynosi odpowiednio nie więcej niż: 1,0, 1,5 lub 2 \AA . Test został przeprowadzony dla dwóch grup kompleksów:

- dla których Dock6 wygenerował przynajmniej jedną pozę z RMSD $< 2 \text{ \AA}$ – zestaw ten obejmuje 42 struktury, a wyniki testu przedstawia Tabela 14;
- dla których Dock6 wygenerował przynajmniej jedną pozę z RMSD $< 1 \text{ \AA}$ – zestaw ten obejmuje 23 struktury, a wyniki testu przedstawia Tabela 15.

Przeprowadzony sprawdzian wykazał, że wszystkie testowane metody odnajdują pozę bliską natywnej (o wartości RMSD odpowiednio $<1,0$; $<1,5$ lub $<2 \text{ \AA}$) w podobnym procencie przypadków. Pośród 42 kompleksów RNA z ligandami, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD $<2 \text{ \AA}$, najwyżej oceniona poza przez LigandRNA, DrugScoreRNA i Dock6 miała RMSD $<2 \text{ \AA}$ do struktury referencyjnej w odpowiednio 35,7%, 31,0% i 35,7% przypadków, a RMSD $<1 \text{ \AA}$ w 21,4%, 14,3% i 16,7% przypadków. Gdy brane były pod uwagę tylko kompleksy, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD $<1 \text{ \AA}$ do struktury natywnej, procent ten wynosił odpowiednio: 42,9%, 28,6% i 33,3% (gdy RMSD $<1 \text{ \AA}$) oraz 42,9%, 42,9% i 47,6% (gdy RMSD $<2 \text{ \AA}$).

Zastosowanie kombinacji ocen LigandRNA-Dock6 znacząco podniosło procent poprawnie zidentyfikowanych póż. Wspólna ocena jest wypadkową ocen składowych tych dwóch metod, które jak wykazano powyżej, z osobna mają podobną skuteczność. Z tego też powodu wagi przypisane obu oceną są równe i wynoszą 0,5. Łączona ocena dała 47,6% póż o wartości RMSD $<2 \text{ \AA}$ w zbiorze 42 struktur oraz 61,9%, gdy rozpatrywane są tylko pozy, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD $<1 \text{ \AA}$. Rycina 15 przedstawia przykład ośmiu kompleksów, dla których kombinacja LigandRNA-Dock6 zidentyfikowała pozę podobną do struktury natywnej (RMSD $<2,5 \text{ \AA}$). Z wykresów wynika, że korelacja RMSD-ocena pozy ulega poprawie przy zastosowaniu kombinacji metod LigandRNA-Dock6.

RMSD	LigandRNA	DrugScoreRNA	Dock6	LigandRNA+Dock6
$<1 \text{ \AA}$	21,4%	14,3%	16,7%	21,4%
$<1,5 \text{ \AA}$	28,6%	21,4%	23,8%	35,7%
$<2 \text{ \AA}$	35,7%	31,0%	35,7%	47,6%

Tabela 14. Procent najlepiej ocenionych póż o RMSD do struktury natywnej odpowiednio mniejszym niż 1, 1,5; 2 \AA , uzyskany dla zbioru 42 kompleksów, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD $<2 \text{ \AA}$.

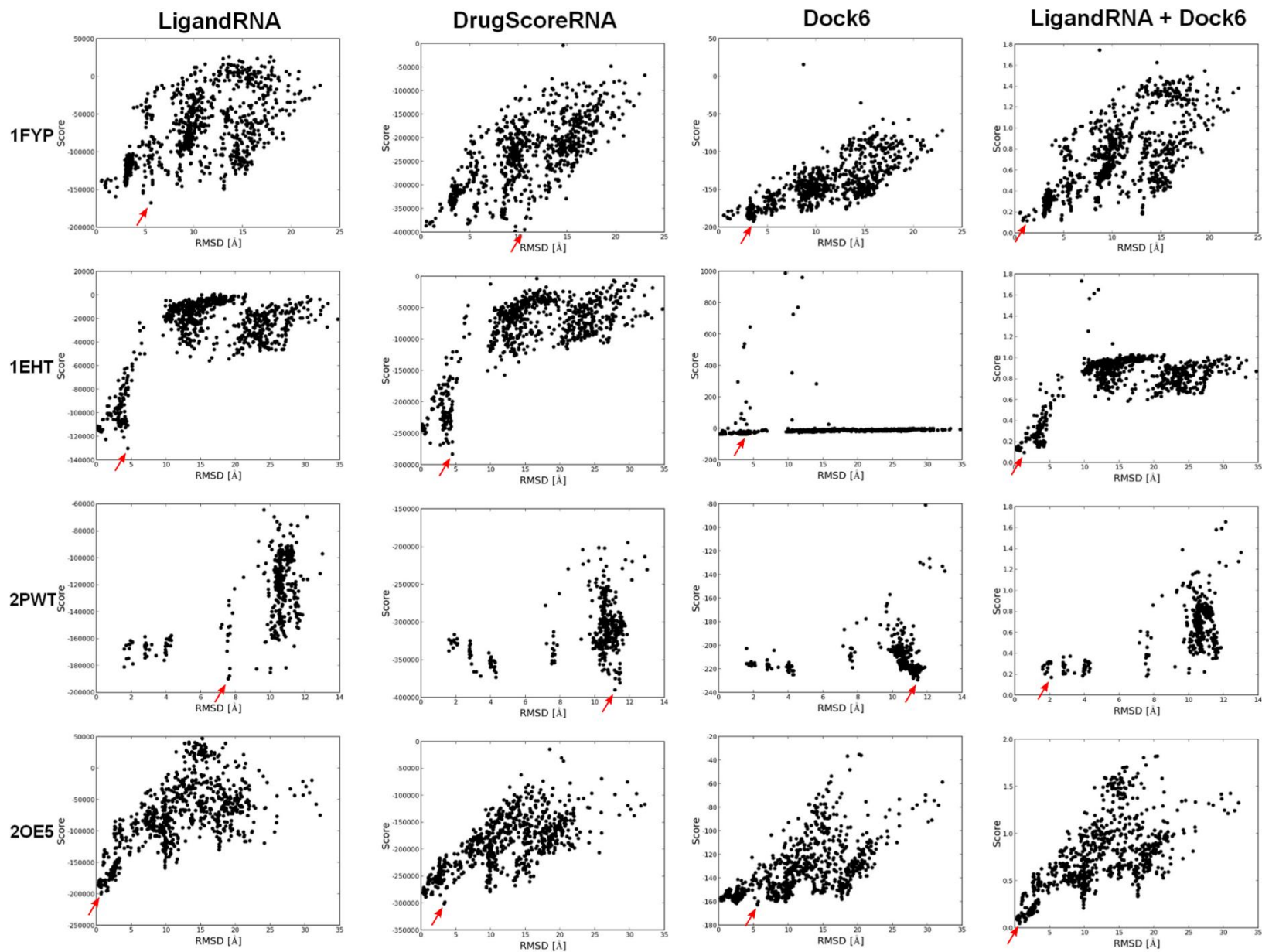
RMSD	LigandRNA	DrugScoreRNA	Dock6	LigandRNA+Dock6
$<1 \text{ \AA}$	42,9%	28,6%	33,3%	42,9%
$<1,5 \text{ \AA}$	42,9%	33,3%	38,1%	57,1%
$<2 \text{ \AA}$	42,9%	42,9%	47,6%	61,9%

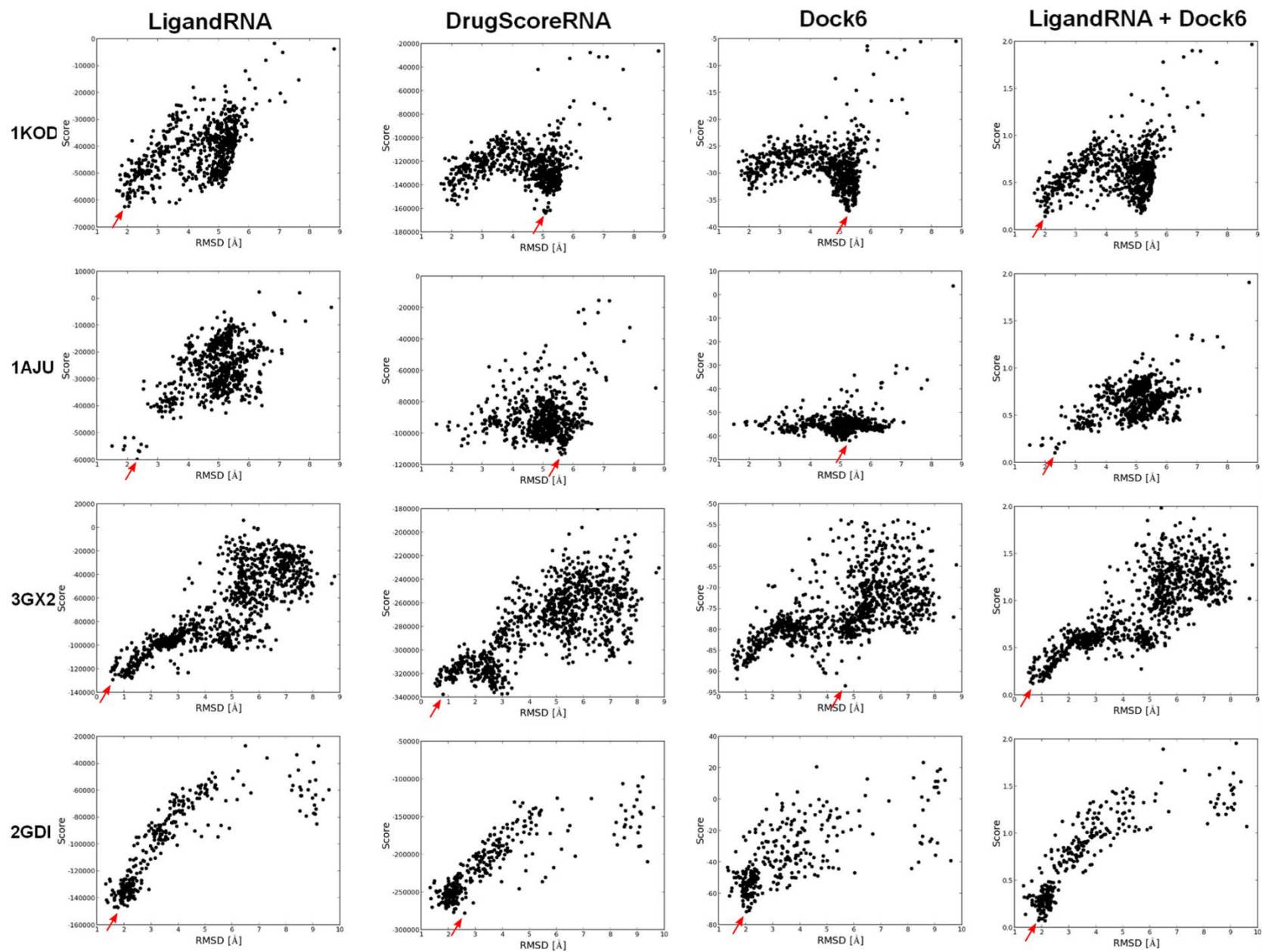
Tabela 15. Procent najlepiej ocenionych póż o RMSD do struktury natywnej odpowiednio mniejszym niż 1, 1,5; 2 \AA , uzyskany dla podzbioru kompleksów, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD $<1 \text{ \AA}$.

Wyniki

kod PDB	Ligand RNA	DrugScore RNA	Dock6	Ligand RNA + Dock6	RiboDock	Funkcja dla aminoglikozydów	Oryginalny DrugScore RNA	MORDOR
1q8n	1,3	1,7	7,7	1,3	-	-	4,3	1,9
2f4u	1,0	3,0	1,9	1,0	-	-	-	2,0
1uts	4,4	8,1	4,2	4,4	-	-	5,6	4,9
2oe5	0,7	3,3	5,6	0,7	-	-	-	1,8
1uud	6,4	5,4	5,4	6,4	-	-	3,6	8,9
1uui	5,6	5,9	1,5	1,5	-	-	4,5	8,4
2et8	0,8	4,6	0,7	0,8	-	0,9	-	1,6
2o3x	0,7	0,7	2,0	0,7	-	-	-	-
1nbk	3,0	6,1	2,6	3,0	-	-	12,0	3,6
1koc	2,3	2,3	2,6	2,3	2,7	-	4,5	1,6
1aju	2,3	5,6	5,3	2,3	-	-	3,3	4,5
1fl1	1,1	1,1	1,2	1,1	-	-	-	0,3
1fmm	2,9	2,2	2,9	2,9	0,8	-	3,3	1,4
2fcz	0,6	7,2	0,9	0,6	-	-	-	1,7
1byj	6,8	1,1	7,4	6,8	1,8	-	0,7	5,4
1mwl	6,8	0,4	9,2	6,8	-	0,7	3,4	1,0
1tob	2,6	5,1	4,8	2,6	9,8	-	3,3	2,6
2tob	1,0	1,0	1,5	1,0	1,5	-	0,9	0,9
2g5q	0,7	0,7	1,6	0,7	-	1,5	-	-
1pbr	6,2	8,3	8,5	6,2	1,8	-	1,1	1,6
1fyp	5,1	5,1	3,2	1,2	-	-	2,8	2,5
1nem	9,2	1,6	0,6	1,1	8,7	-	2,3	1,0
1j7t	8,6	2,6	1,1	8,6	-	1,9	3,2	1,1
1ei2	6,9	6,9	5,9	6,9	-	-	3,1	3,2
2be0	10,7	3,5	3,5	10,7	-	2,0	-	2,0
2pwt	7,6	11,2	11,4	2,1	-	-	-	-
2fd0	4,6	3,5	3,5	4,6	-	-	-	1,0
2bee	9,3	9,3	9,3	9,3	-	1,5	-	2,0
2ogn	8,6	8,0	8,5	9,0	-	-	-	-
1xbp	8,1	1,7	8,9	8,1	-	-	-	-
1eht	4,5	4,5	0,3	1,4	3,6	-	7,5	1,3
1y26	0,5	3,8	0,5	0,5	-	-	-	0,5
1am0	4,6	1,8	1,4	1,6	1,8	-	4,6	0,9
3gx2	0,6	0,8	4,7	0,6	-	-	-	-
3d2x	1,6	2,2	1,9	1,6	-	-	-	-
1xpf	3,4	6,6	6,6	3,4	-	-	-	5,3
1kod	1,9	5,1	5,3	1,9	3,2	-	1,7	2,0
1fjg	6,3	0,3	0,3	6,3	-	-	-	-
1hnw	8,3	5,0	5,0	8,3	-	-	-	-
3sux	0,4	0,5	0,5	0,4	-	-	-	-
2gdi	1,8	2,6	2,0	1,8	-	-	-	2,0
1f27	6,7	6,4	6,9	6,7	-	-	3,6	1,0

Tabela 16. Wyniki testu programu przeprowadzonego na wybranych 42 kompleksach RNA z ligandami. Pierwsza kolumna zawiera kody PDB kompleksów, następane wartości RMSD (w Angstrmach) najlepiej ocenionych póż przez analizowane programy.





Rycina 15. Przykłady zależności oceny pozycji i jej wartości RMSD do struktury referencyjnej. Strzałką zostały zaznaczone pozycje zidentyfikowane przez metody jako najlepsze.

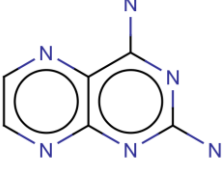
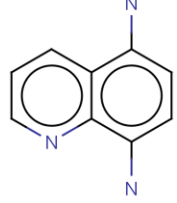
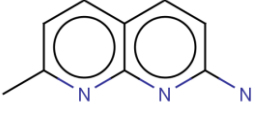
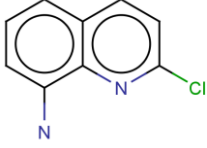
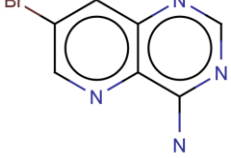
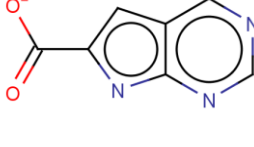
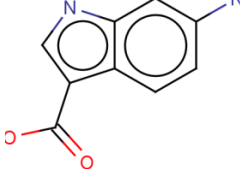
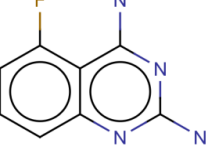
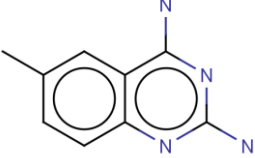
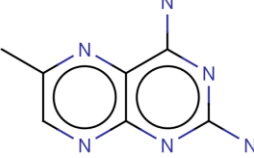
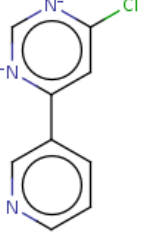
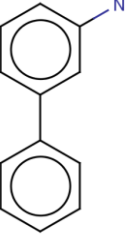
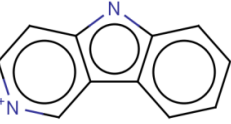
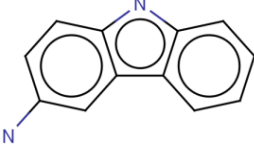
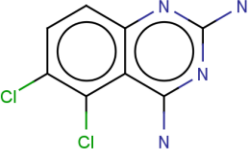
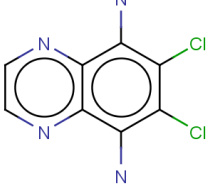
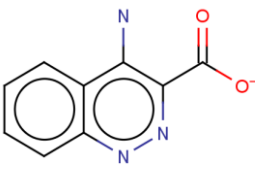
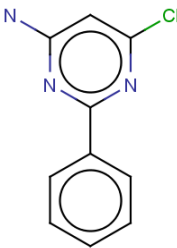
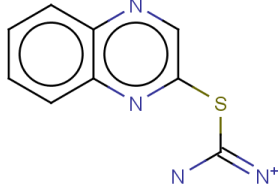
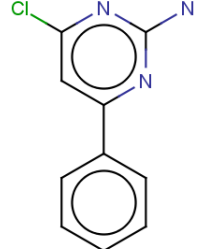
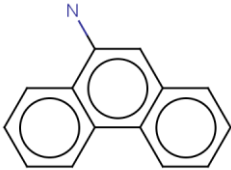
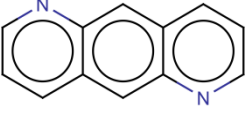
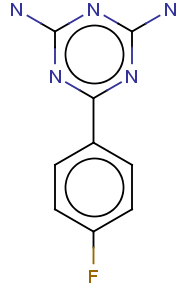
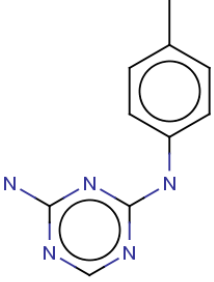
Wirtualna analiza przesiewowa

Dodatkowym testem programu LigandRNA było użycie go do wytypowania potencjalnych inhibitorów centrum peptydylotransferazy bakteryjnych rybosomów. Najpierw, przy pomocy programu do dokowania Surflex, przeprowadzono wirtualną analizę przesiewową dla struktury centrum peptydylotransferazy z dużej podjednostki 50S rybosomu bakterii *E. coli* (kod PDB: 3OFQ). Struktura ta została wybrana do testów ponieważ nie zawierała ona żadnego liganda w PTC, została rozwiązana przy pomocy krystalografii rentgenowskiej z rozdzielczością 3,10 Å oraz pochodziła z *E. coli*, które to bakterie mogą przybierać patogenną postać.

Podczas VS dla wszystkich ligandów z bazy danych cząstek lekopodobnych ZINC (15 798 630 związków) program Surflex wygenerował po 10 pól. Następnie, przy pomocy programu LigandRNA, zostały one ocenione, a te z najlepszymi ocenami zostały wytypowane jako potencjalne inhibitory. Ponieważ program LigandRNA na pierwszych kilkuset pozycjach rankingu umieścił grupy związków bardzo podobnych do siebie pod względem struktury chemicznej, wyniki programu zostały sklasteryzowane. Dokonano podziału związków w zależności od liczby atomów ciężkich: grupa <12 atomów ciężkich, grupa 12 at. ciężkich, grupa 13 at. ciężkich, 14, 15 i 16 atomów ciężkich (razem) oraz 17 i więcej atomów ciężkich. Grupa <12 atomów nie była brana pod uwagę, ze względu na niską specyficzność wiązania się tak małych ligandów.

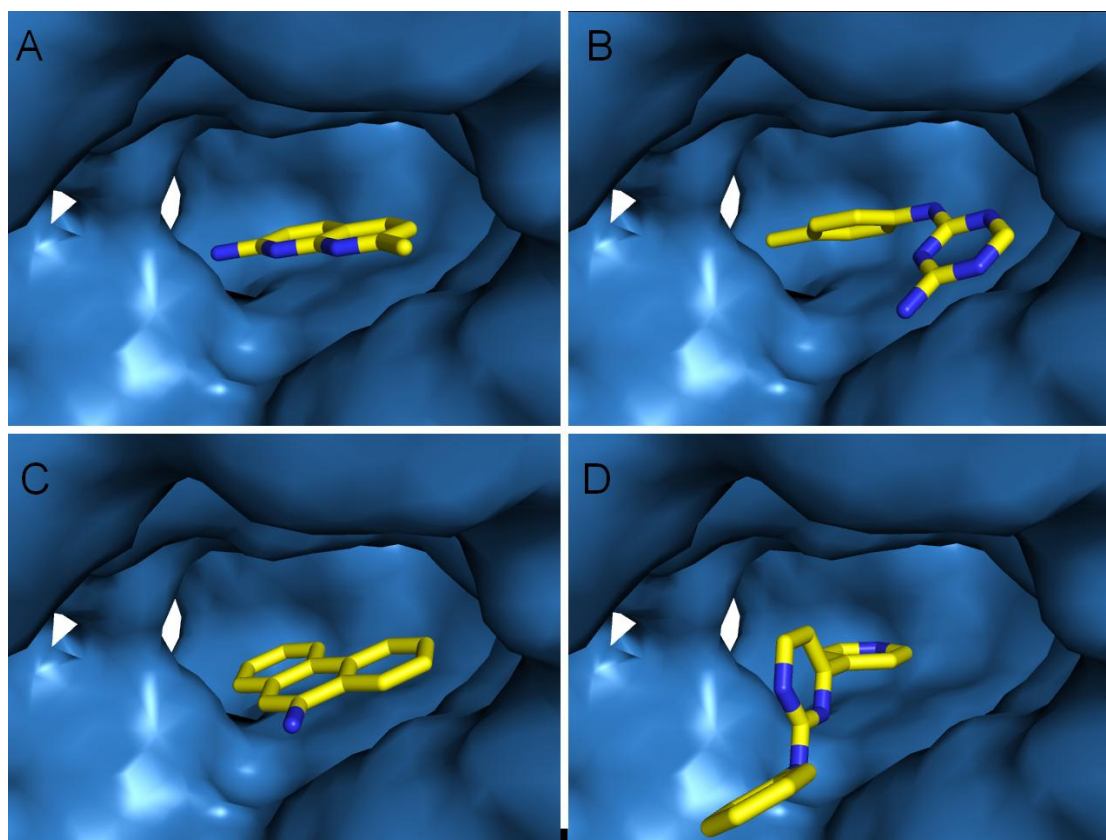
Tabela 17 prezentuje najwyżej ocenione związki spośród pozostałych grup. Ich strukturę chemiczną, kod ZINC oraz liczbę tworzących je atomów ciężkich. Cztery spośród tych związków zostały wybrane do dalszej analizy doświadczalnej, która zostanie przeprowadzona już poza ramami niniejszego projektu w laboratorium prof. Janusza Bujnickiego w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej w Warszawie. Najważniejszym kryterium wyboru związków była wizualna ocena ich dopasowania do receptora (Rycina 16), ale również dostępność danego związku u dostawcy. Preferowane były też związki o większej liczbie atomów. Przymuszczalnie wytypowane związki będą się niespecyficznie wiązać z innymi cząsteczkami w komórce, więc niemożliwe będzie ich zastosowanie jako leków. Jednak celem omawianego testu jest wykazanie, że możliwa jest identyfikacja związków przy pomocy LigandRNA, które wiążą się do danego receptora.

Wyniki

ZINC00370801, 12 at. 	ZINC00506172, 12 at. 	ZINC32198129, 12 at. 	ZINC01597597, 12 at. 
ZINC66054367, 12 at. 	ZINC35269905, 12 at. 	ZINC01631245, 13 at 	ZINC00156168, 13 at. 
ZINC12343992, 13 at. 	ZINC00050187, 13 at 	ZINC26420731, 13 at. 	ZINC01746590, 13 at 
ZINC06117994, 13 at. 	ZINC01871437, 14 at. 	ZINC05973969, 14 at. 	ZINC00189796, 14 at. 
ZINC33893236, 14 at. 	ZINC01769621, 14 at. 	ZINC04390229, 14 at. 	ZINC00039987, 14 at. 
ZINC01509324, 15 at. 	ZINC11535924, 15 at. 	ZINC12959222, 15 at. 	ZINC00089474, 15 at. 

ZINC65506880, 17 at.	ZINC00406567, 18 at.	ZINC50197835, 18 at.	ZINC71835652, 19 at.
ZINC26654647, 19 at.	ZINC00189666, 20 at.	ZINC01668715, 21 at.	ZINC34743425, 22 at.

Tabela 17. Najwyżej ocenione związki z poszczególnych grup (podział według liczby atomów ciężkich) przez LigandRNA, związki wybrane manualnie do dalszych testów doświadczalnych zostały zaznaczone na czerwono.



Rycina 16. Cztery związki wytypowane do testów doświadczalnych. Kody ZINC: Panel A) ZINC32198129. Panel B) ZINC00089474. Panel C) ZINC01509324. Panel D) ZINC26654647. Struktura receptora – kod PDB: 3OFQ.

5. Dyskusja

Wyniki opisane w niniejszej rozprawie doktorskiej koncentrują się wokół opracowania metodologii przewidywania oddziaływań RNA z jonami metali i ligandami. W tym celu stworzono dwa anizotropowe potencjały statystyczne, a następnie programy MetalionRNA i LigandRNA, w których zostały one zaimplementowane. Program MetalionRNA służy do przewidywania miejsc wiązania jonów Mg^{2+} , K^+ i Na^+ w strukturach RNA. Program LigandRNA umożliwia ocenę różnych pól liganda/ów w strukturze RNA. Obie metody zostały przetestowane i porównane z konkurencyjnymi narzędziami.

Oba programy udostępnione zostały na licencji wolnego i otwartego oprogramowania (ang. *General Public License*, GPL) i dołączone do niniejszej rozprawy na płycie CD (patrz: „Załączniki”). Ponadto powstały serwery internetowe, umożliwiające prowadzenie obliczeń „online”. Więcej informacji na ten temat znajduje się w Rozdziale 5.3 „Serwery MetalionRNA i LigandRNA”.

5.1. Omówienie programu MetalionRNA

Zdolność potencjału MetalionRNA do przewidywania miejsc wiązania jonów metali w strukturach RNA.

Działanie programu MetalionRNA opiera się na anizotropowym potencjale statystycznym wyprowadzonym ze znanych kompleksów RNA z jonami metali. W celu zweryfikowania zdolności stworzonego potencjału do identyfikacji miejsc wiązania się kationów do RNA został przeprowadzony pięciokrotny sprawdzian krzyżowy. Test ten dowiódł wysokiej skuteczności działania programu MetalionRNA. Miejsca wiązania się wyżej wymienionych jonów zostały poprawnie przewidziane w przypadku dziesiątek krystalograficznie wyznaczonych pozycji tych jonów z różnych struktur RNA. Co więcej, krzywe ROC bardzo szybko osiągają wartość TPR zbliżoną do wartości maksymalnej (0,9~1,0), przy stosunkowo niskiej wartości FPR (~0,2). Oznacza to, że spośród najwyższej ocenionych przewidywań odsetek tych niepoprawnych jest znikomy w porównaniu z liczbą poprawnie zidentyfikowanych jonów.

Warto zauważyć, że wartości TPR dla niektórych krzywych ROC wynoszą maksymalnie $\sim 0,7$ (RNA z Mg^{2+}), a minimalnie $\sim 0,55$ (RNA z K^+ i RNA z Na^+). Krzywe te odpowiadają sytuacji, gdy przewidywanie było uznawane za poprawne, czyli gdy wartość RMSD dla przewidywanej pozycji jonu do pozycji prawdziwego jonu ze struktury referencyjnej nie przekraczało promienia jonowego tego kationu. Takie wartości TPR oznaczają, że od 45% do 30% jonów w strukturach natywnych w ogóle nie zostało przewidzianych z tak dużą dokładnością. Dla bardziej tolerancyjnych zakresów progu odcięcia jony te zostały przewidziane, a wartości TPR wynosiły $\sim 1,0$.

Podobna skuteczność została osiągnięta w przypadku testu przeprowadzonego na zbiorze struktur DNA z Mg^{2+} . Struktury te nie były zawarte w zbiorze uczącym potencjału, a przewidywania przeprowadzono dla nich przy pomocy potencjału wyprowadzonego ze zbioru struktur RNA z Mg^{2+} . Test ten dowiódł, że stworzony potencjał statystyczny uchwycił ogólny mechanizm wiązania się kationów do kwasów nukleinowych.

Porównanie programu MetalionRNA z innymi metodami

Jak omówiono we wstępie, istnieje szereg metod do przewidywania miejsc wiązania się jonów w strukturach RNA, są to metody kosztowne obliczeniowo i niemożliwe do uruchomienia w trybie automatycznym. Wszystkie z nich symulują zachowanie się systemu w zależności od jego fizyko-chemicznych właściwości i parametrów zadanych przez użytkownika. Służą też innym celom niż w pełni automatyczne narzędzia takie jak MetalionRNA czy FEATURE, które działają na zasadzie „czarnej skrzynki” (ang. *blackbox*), przez co nadają się do analizy dużej liczby struktur. Dlatego też te dwa typy narzędzi nie zostały bezpośrednio porównane.

Porównanie MetalionRNA z FEATURE ponownie dowiodło, że MetalionRNA identyfikuje prawdziwe miejsca wiązania się Mg^{2+} w RNA z relatywnie niskim odsetkiem przewidywań fałszywie dodatnich. Niestety w tym porównaniu było możliwe jedynie odniesienie się do wyników prezentowanych w publikacji (Liang, Banatao i wsp. 2003) (przewidywania dla struktury 1HC8). Serwer WebFEATURE nie posiada możliwości załadowania więcej niż jednej struktury na raz, a przy takiej próbie przestaje działać. Ponadto, wyświetla błędy podczas próby zmiany parametrów i pobrania wyników. Problem został zgłoszony autorom, lecz nie zgodzili się oni na udostępnienie programu w innej formie niż dostępny publicznie serwer. W obliczu takiej sytuacji niemożliwe

Dyskusja

było przetestowania FEATURE na zbiorze testowym MetalionRNA. W zamian za to, dodatkowym testem było użycie zbioru uczącego FEATURE do wyprowadzenia potencjału MetalionRNA, a następnie użycie go do przeprowadzenia przewidywań dla struktury 1HC8. Wyniki dowiodły, że taki potencjał również radzi sobie lepiej niż oryginalny FEATURE, lecz nie tak dobrze jak MetalionRNA wytrenowany na oryginalnym zbiorze. Oznacza to, że stworzony potencjał statystyczny MetalionRNA, opierający się na odległościach i kątach między atomami, jest bardziej specyficzny niż FEATURE, a zbiór uczący, który posłużył do jego wyprowadzenia pewniejszy. Najprawdopodobniej wynika to z faktu, iż oryginalny zbiór uczący MetalionRNA jest niemal dziesięciokrotnie większy niż zbiór FEATURE.

Podsumowanie MetalionRNA

W ramach niniejszej pracy doktorskiej został stworzony program MetalionRNA, który wykorzystując potencjał statystyczny wyprowadzony ze znanych struktur PDB przewiduje miejsca wiązania się jonów w RNA. Przeprowadzone testy i porównania dowiodły, że jego skuteczność jest wyższa niż dotychczas dostępnej metody FEATURE, również opierającej się na empirycznym potencjale.

Szybki wzrost liczby struktur RNA zainspirował wprowadzenie do programu MetalionRNA cotygodniowych aktualizacji statystyk. Serwer MetalionRNA (patrz: Rozdział 5.3 „Serwery MetalionRNA i LigandRNA”) co siedem dni o tej samej porze pobiera nowo zdeponowane struktury PDB o rozdzielczości lepszej niż 2 Å i, jeśli spełniają one dodatkowe kryteria wyznaczone dla kompleksów RNA z Mg^{2+} , RNA z K^+ czy RNA z Na^+ (opisane szerzej w Rozdziale 3.2 „Przygotowanie zbiorów uczących i testowych”), dołącza je do zbiorów uczących potencjału, który przeliczany jest na nowo. Dzięki tej procedurze liczba struktur o gorszej rozdzielczości (2-3 Å) będzie stopniowo przewyższana przez te o lepszej, co w rezultacie doprowadzi do ulepszania potencjału. Serwer MetalionRNA oferuje wybór pomiędzy oryginalnym potencjałem statystycznym prezentowanym w tej pracy a najnowszym.

Program MetalionRNA może zostać wykorzystany jako uzupełnienie doświadczalnego rozwiązywania struktury przestrzennej RNA. Na przykład, niepewne miejsca wiązania się kationów z danych doświadczalnych mogą zostać potwierdzone poprzez porównanie ich z wynikami uzyskanymi z MetalionRNA. Ponadto, MetalionRNA może również zaproponować miejsca wiązania się jonów w strukturach,

w których ich brak, np. w strukturach NMR (Shen, Cai i wsp. 1995) czy modelach teoretycznych. Takie miejsca wiązania można następnie potwierdzić doświadczalnie przy pomocy metody Fentona. Wykorzystuje ona fakt, że jony Mg^{2+} mogą zostać zastąpione przez Fe^{2+} , które w roztworze wodnym powodują powstanie rodników hydroksylowych, tnących łańcuch główny RNA. Dzięki identyfikacji miejsc cięcia, już przy pomocy standardowych metod biochemicznych, możliwe jest określenie przybliżonego miejsca wiązania się kationu Fe^{2+} , a co za tym idzie Mg^{2+} (Berens, Streicher i wsp. 1998).

Niewątpliwą zaletą programu MetalionRNA jest fakt, że jest on szybki, łatwo dostępny przez przyjazny użytkownikowi serwer internetowy oraz, że nie wymaga on specjalistycznej wiedzy i umiejętności do jego uruchomienia i interpretacji wyników. Miejsca wiązania się jonów przewidziane przez MetalionRNA są zwracane w postaci rankingu od miejsca ocenionego jako najbardziej prawdopodobne aż do najmniej prawdopodobnego. Dzięki temu użytkownik łatwo może wnioskować o sile wiązania jonów, czy decydować ile z nich jest przewidywaniami „silnymi”, np. w warunkach zwiększającego się stężenia jonów w roztworze.

5.2. Omówienie programu LigandRNA

Zdolność potencjału LigandRNA do przewidywania oddziaływań RNA z ligandami

W celu zweryfikowania zdolności stworzonego potencjału do identyfikacji póź liganda bliskich natywnym przeprowadzono test na zbiorze 42 kompleksów. By test był jak najbardziej obiektywny, struktury te zostały wybrane w taki sposób, żeby zagwarantować możliwie jak największą różnorodność ligandów. Dodatkowo zapewniono, że struktury testowe, ani im podobne, tzn. należące do tego samego klastra, nie znajdowały się w zbiorze uczącym potencjału, który posłużył do oceny wygenerowanych dla nich póź.

Przeprowadzony test dowiódł wysokiej skuteczności działania LigandRNA. Stworzony program poprawnie wybrał pozycje ligandów bliskie tym rozwiązaniem doświadczalnie w wielu różnych strukturach RNA. Aż w przypadku 20 spośród

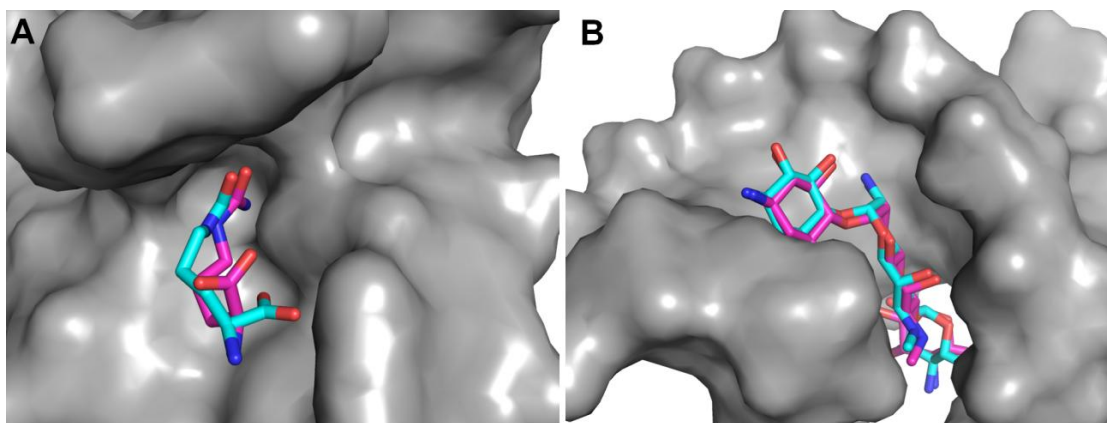
Dyskusja

23 struktur, dla których istniała poza o RMSD $<1 \text{ \AA}$, program wybrał pozy liganda bliskie natywnym (tzn. z RMSD $<3 \text{ \AA}$).

Dodatkowym testem dla programu LigandRNA było użycie go do oceny poz ligandów otrzymanych z VS przeprowadzonego dla PTC dużej podjednostki rybosomu bakteryjnego. Najwyżej ocenione pozy zostały poddane klasteryzacji i manualnej analizie, a cztery z nich wytypowano do dalszych badań doświadczalnych. Są one obecnie prowadzone w laboratorium prof. Bujnickiego w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej w Warszawie. Jeśli zaproponowane ligandy okazałyby się inhibitorami badanego regionu, test ten ostatecznie dowiódłby skuteczności działania LigandRNA.

Porównanie programu LigandRNA z innymi metodami

Porównanie programu LigandRNA z alternatywnymi metodami wykazało, że odsetek najlepszych wyników wygenerowanych przez LigandRNA jest porównywalny z metodą MORDOR (około 31% przypadków) i wyższy od wyników metod RiboDock, DrugScoreRNA czy Dock6 (patrz: Tabela 16). Ponadto, program LigandRNA jako jedyny spośród testowanych metod znalazł pozę bliską natywną w przypadku dwóch kompleksów. Pierwszy z nich to aptamer RNA w kompleksie z cytruliną (kod PDB: 1KOD). LigandRNA najwyżej ocenił pozę o RMSD $1,9 \text{ \AA}$, natomiast Dock6 i DrugScoreRNA odpowiednio o RMSD równym $5,3$ i $11,2 \text{ \AA}$. Drugi kompleks to struktura rRNA centrum dekodującego ze związanym z nią antybiotykiem, apramycyną (kod PDB: 2OE5), dla którego LigandRNA zidentyfikował pozę o odległości RMSD $0,7 \text{ \AA}$, natomiast Dock6 pozę o RMSD $5,6 \text{ \AA}$, a DrugScoreRNA o $3,3 \text{ \AA}$. Obie struktury i związane z nimi ligandy, a także pozy najlepiej ocenione przez LigandRNA, prezentuje Rycina 17.

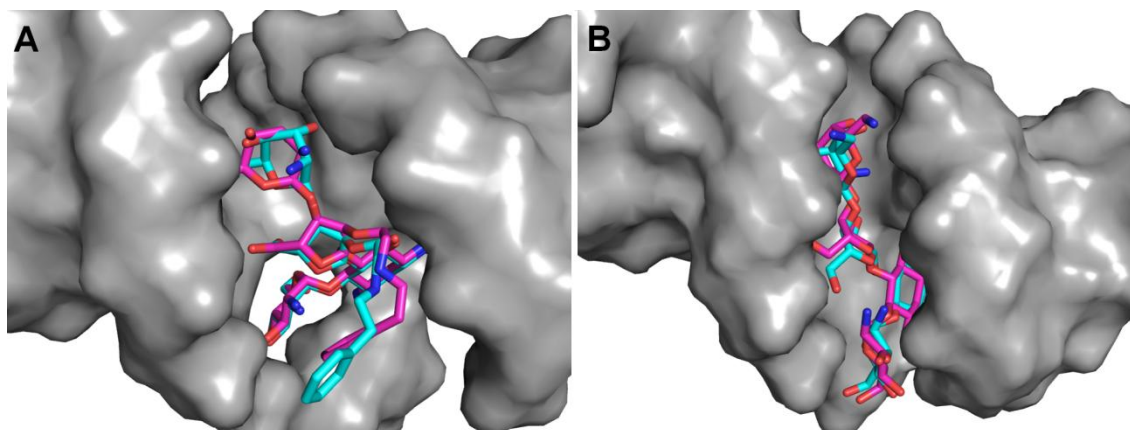


Rycina 17. Panel A) Aptamer RNA w kompleksie z cytruliną (kod PDB: 1KOD). Panel B) Centrum dekodujące rybosomu w kompleksie z apramycyną (pdb id: 2OE5). Doświadczalnie rozwiązane pozycje ligandów przedstawione są na jasnoniebiesko, natomiast pozycje wygenerowane przy pomocy Dock6 i ocenione najwyżej przez LigandRNA ukazane są na różowo.

Ocena łączona LigandRNA-Dock6

Z przeprowadzonego testu, którego wyniki podsumowują Tabela 14 i Tabela 15, wynika, że zarówno program LigandRNA jak i Dock6 wybierają pozycje bliskie natywnym ze zbliżoną skutecznością. Dodatkowo biorąc pod uwagę fakt, że metody te opierają się na skrajnie różnych podejściach (empirycznym i pola sił), zdecydowano się na połączenie ocen tych dwóch metod i sprawdzenie skuteczności takiego mieszanego podejścia. Wagi przypisane ocenom obu programów były równe i wynosiły 0,5, gdyż, jak wspomniano, programy te wykazały w teście podobną skuteczność.

Zgodnie z oczekiwaniami, zastosowanie łączonej oceny LigandRNA-Dock6 przyniosło zwiększenie odsetka poprawnych przewidywań. Dowodzi tego zarówno fakt, że polepszony został współczynnik korelacji pomiędzy oceną a wartościami RMSD poszczególnych póż (patrz: Rycina 15), jak i wynik porównania z innymi metodami. Liczba najlepszych rozwiązań mieszanej metody jest wyższa niż każdej z nich zastosowanej z osobna. Co więcej, poprawnie wytypowała ona pozycje bliskie natywnym w przypadku dwóch kompleksów, dla których zawiodły pozostałe podejścia. Wspólna ocena LigandRNA-Dock6 zaowocowała identyfikacją póż o odległościach RMSD równym 2 i 1,2 Å odpowiednio dla kompleksów: aminoglikozydu z grupą L-HABA związaną z rRNA centrum dekodującego oraz paromycyny również związaną z centrum dekodującym, w okolicy miejsca A (kody PDB: 2PWT, 1FYP). Przedstawione wyniki prezentuje Rycina 18.



Rycina 18. Panel A) rRNA centrum dekodującego w kompleksie z aminoglikozydem z grupą L-HABA. Panel B) rRNA miejsca A w kompleksie z paromycyną. Doświadczalnie rozwiązane pozy ligandów przedstawione są na jasnoniebiesko, natomiast pozy wygenerowane przy pomocy Dock6 i ocenione najwyżej przez kombinację ocen LigandRNA-Dock6 ukazane są na różowo.

Wzrost poprawności otrzymywanych wyników w przypadku zastosowania oceny łączonej LigandRNA-Dock6 najprawdopodobniej wynika z faktu, że generowały je funkcje o bardzo różnym, przez co komplementarnym, charakterze. LigandRNA jest empirycznym potencjałem statystyczny podczas, gdy Dock6 ocenia pozy liganda na podstawie kalkulacji energii oddziaływań.

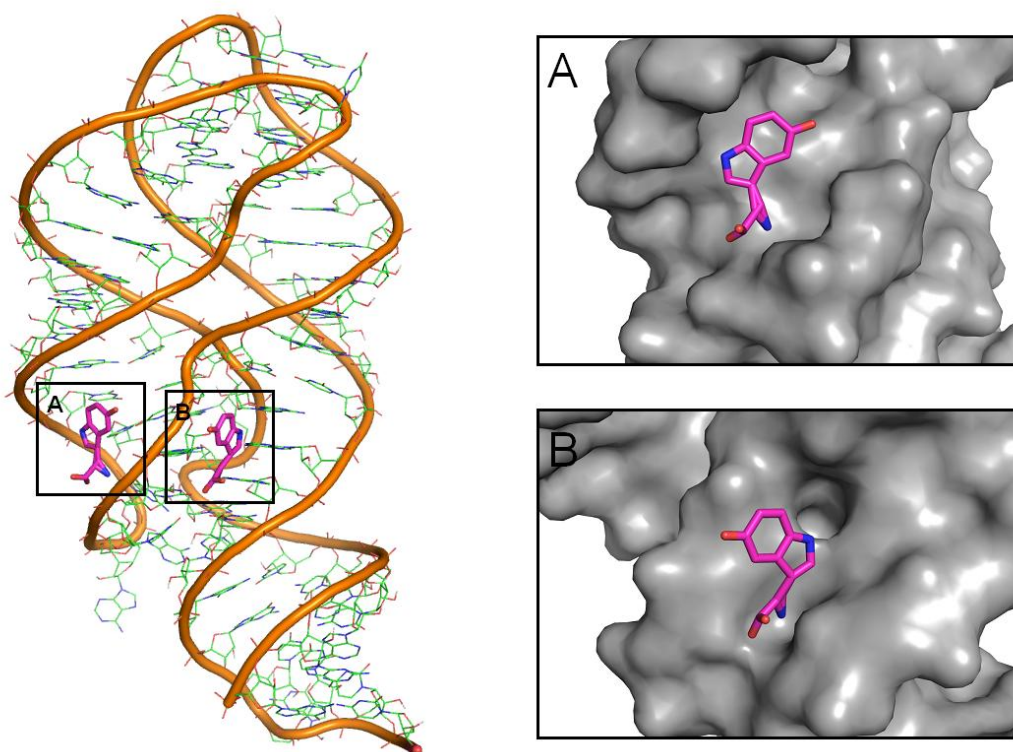
Podsumowanie LigandRNA

W ramach prezentowanej pracy doktorskiej został stworzony program LigandRNA. Wykorzystuje on potencjał statystyczny wyprowadzony ze znanych struktur PDB do oceny poz ligandów w strukturach RNA, które zostały wygenerowane przez niezależne narzędzia. Dowiedziono, że stworzony potencjał działa z większą skutecznością niż istniejące metody. Połączenie ocen LigandRNA i Dock6 daje dodatkowe zwiększenie dokładności przewidywań, dlatego też zalecane jest użycie LigandRNA z tą metodą.

Program LigandRNA jako jedyna metoda, spośród wymienionych w testach, umożliwia ocenienie poz ligandów wygenerowanych przez różne narzędzia do dokowania. Ponadto, jako jedyny jest szybki i łatwy w użyciu, dostępny poprzez serwer internetowy. Użytkownik nie musi posiadać specjalistycznej wiedzy, czy doświadczenia, by uzyskać wiarygodne wyniki. LigandRNA jako rezultat obliczeń

zwraca ranking zadanych pól liganda/ów w zależności od otrzymanych przez nie ocen, z czego użytkownik może łatwo wnioskować o sile wiązania.

Ciekawostką może być fakt, że program LigandRNA posłużył do oceny wyników dokowania przeprowadzonego przez autorkę niniejszej rozprawy na konkurs RNA Puzzles (Cruz, Blanchet i wsp. 2012). Konkurs ten jest organizowany od trzech lat, a jego idea została zaczerpnięta z konkursu modelowania białek, CASP. Grupy naukowców z całego świata stają przed zadaniem wymodelowania struktury przestrzennej RNA. Dostępne są dla nich jedynie informacje o sekwencji nukleotydowej, czy, jak w przypadku tegorocznego konkursu, o strukturze chemicznej liganda. Zadanie polegało na wymodelowaniu struktury ryboprzełącznika i jego kompleksu z zadaniem ligandem, 5-hydroksytryptofanem. Do dokowania został użyty program Dock6, następnie otrzymane pozy zostały ocenione przy pomocy LigandRNA oraz oceny łączonej LigandRNA-Dock6, które w omawianym przypadku dały zgodny wynik i przedstawiały ten sam ranking. Najwyżej ocenione dwie pozy z dokowania, które zostały wysłane na konkurs, prezentuje Rycina 19. Wyniki zmagania nie są jeszcze znane.





Rycina 19. Struktura modelu ryboprzełącznika oraz dwie najwyżej ocenione pozy 5-hydroksytryptofanu z nim związane.

5.3. Serwery MetalionRNA i LigandRNA

Zarówno program MetalionRNA, jak i LigandRNA są bezpłatnie dostępne na serwerach internetowych pod adresami: <http://metalionrna.genesilico.pl/> oraz <http://ligandrna.genesilico.pl/>. Serwery zostały zaimplementowane przez mgr Kaję Milanowską.

Serwer MetalionRNA umożliwia przewidzenie zdefiniowanej, bądź domyślnej liczby jonów magnezu sodu lub potasu w dowolnej strukturze przestrzennej RNA (lub DNA) przesłanej przez użytkownika. Formularz umożliwia załadowanie pliku (w formacie PDB) lub podanie kodu PDB struktury, którą następnie serwer sam pobiera z bazy danych PDB. Dodatkowo serwer umożliwia zdefiniowanie wartości innych niż domyślne: minimalnej odległości między przewidywanymi jonami oraz ich promienia. Ponadto użytkownik może wybrać rozdzielczość sieci, na której prowadzone są obliczenia (0,25 lub 0,5 Å) oraz typ potencjału (potencjał zaprezentowany w publikacji (Philips, Milanowska i wsp. 2012) lub zaktualizowany). W wyniku działania serwer zwraca pliki PDB z przewidzianymi jonami oraz ze strukturą RNA i policzonym dla niej powierzchniowym potencjałem statystycznym (w wynikowym pliku PDB kolumna z wartością czynnika temperaturowego (ang. *b-factor*) jest zastąpiona wartościami powierzchniowego potencjału statystycznego).

Serwer LigandRNA natomiast umożliwia ocenę zadanych pól liganda (akceptowany format pliku: MOL2) w strukturze RNA (akceptowany format pliku: PDB lub kod PDB). Podobnie jak w przypadku programu MetalionRNA użytkownik może wybrać typ potencjału (taki jak prezentowany w publikacji LigandRNA lub zaktualizowany). Ponadto, dla pól wygenerowanych przy pomocy programu Dock6 istnieje możliwość wyznaczenia łączonej oceny kombinacją metod: LigandRNA-Dock6. Wynikiem działania serwera jest plik tekstowy z rankingiem i oceną pól (od najlepiej ocenionej aż po najgorszą) oraz cztery pliki PDB ze strukturą RNA i policzonym dla niej powierzchniowym potencjałem statystycznym: dla wszystkich typów atomów liganda oraz dla atomów: C, N i O. Wygląd strony powitalnej serwera LigandRNA prezentuje Rycina 20.

**Laboratory of Bioinformatics and Protein Engineering**

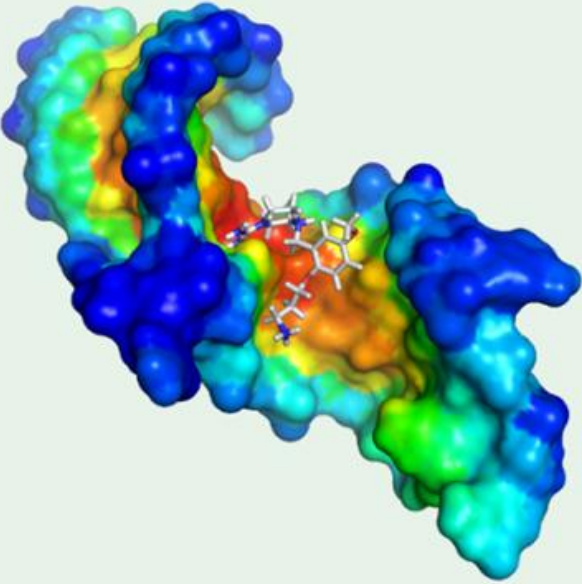
[Home](#)
[Submit](#)
[Instructions](#)
[Contact](#)
[Genesilico Homepage](#)

WELCOME TO

LigandRNA

A ligand binding site predictor

LigandRNA is a novel bioinformatic method for predicting ligand-binding sites in RNA structures.



Rycina 20. Zrzut ekranu ze strony powitalnej serwera LigandRNA.

5.4. Ograniczenia i perspektywy rozwoju

Ograniczenia i perspektywy rozwoju MetalionRNA

Wszystkie omawiane w niniejszej pracy programy do przewidywania miejsc wiązania kationów umożliwiają przewidywanie jedynie Mg^{2+} . Obecna implementacja MetalionRNA jest zdolna do przewidywania miejsc wiązania Mg^{2+} , K^+ i Na^+ . Są to trzy rodzaje kationów, które najczęściej wiążą się z RNA. W czasie powstawania niniejszego projektu, były to jedyne jony, dla których w bazie PDB znajdowała się wystarczająco duża liczba struktur do stworzenia wiarygodnego potencjału statystycznego. Jednak w ostatnich latach obserwuje się znaczący wzrost liczby rozwiązanych doświadczalnie kompleksów RNA i jonów metali, co w przyszłości powinno umożliwić stworzenie potencjału statystycznego również dla innych typów jonów (np. Ca^{2+}).

Szybki wzrost liczby rozwiązanych struktur nie oznacza jednak mniejszej liczby błędów. Szczególnie w strukturach o gorszej rozdzielczości $>3 \text{ \AA}$ zdarza się, że kationy błędnie zidentyfikowane są jako cząsteczki wody, ponieważ zarówno H_2O jak i Mg^{2+} i Na^+ posiadają po dziesięć elektronów i niemożliwe jest ich łatwe rozróżnienie. Identyfikacja jonów metali jest również trudna w przypadku struktur rozwiązywanych za pomocą NMR. W związku z tym często ich w nich brakuje. Stąd też warto byłoby ocenić jak program MetalionRNA radzi sobie z przewidywaniem miejsc wiązania jonów w strukturach o gorszej jakości, a następnie wprowadzić konieczne modyfikacje potencjału i programu, by przystosować go do pracy z tego rodzaju danymi. Obecnie nie ma ograniczeń, co do rozdzielczości struktur, dla których robione są przewidywania, jednak zaleca się użycie struktur o rozdzielczości $<3 \text{ \AA}$, gdy jest to możliwe.

Prezentowana wersja programu MetalionRNA posiada funkcję wyznaczenia domyślnej liczby jonów dla RNA. Liczba ta jest stała i zależna jedynie od jego długości (liczby reszt nukleotydowych). W kolejnych implementacjach wprowadzenie stężenia jonów jako parametru pozwoliłoby na mniej restrykcyjne wyznaczanie liczby przewidywań. Wymagałoby to dogłębnej analizy znanych struktur i stworzenia funkcji zależności liczby obecnych jonów od stężenia jonów, przy jakim RNA zostało skryształizowane, oraz jego wielkości.

Porównanie z FEATURE zasugerowało, że kationy współzawodniczą ze sobą o najbardziej atrakcyjne energetycznie miejsca w trakcie wiązania się z RNA. Może świadczyć o tym fakt, że miejsce wiązania K^+ zostało zidentyfikowane przez FEATURE jako Mg^{2+} , a w rankingu MetalionRNA dla jonów K^+ najlepsze przewidywania znajdowały się w miejscach wiązania Mg^{2+} . Warto więc rozważyć umożliwienie przeprowadzenia przewidywań dla różnych typów jonów jednocześnie, co zbliżyłoby proces przewidywania do warunków panujących w komórce. Będzie to możliwe dopiero, gdy liczba struktur rozwiązanych przy różnym stężeniu różnych jonów osiągnie poziom wystarczający do stworzenia statystycznie istotnego zbioru uczącego i testowego.

Ograniczenia i perspektywy rozwoju LigandRNA

Niewątpliwą zaletą programu LigandRNA jest fakt, że może on posłużyć do oceny póż ligandów wygenerowanych dowolną metodą, przy użyciu różnych programów. Daje to możliwość porównania wyników otrzymanych z różnych narzędzi do dokowania przy pomocy niezależnej funkcji oceny. Jednak, fakt, że program LigandRNA nie posiada własnego modułu do generowania póż ligandów jest też ograniczeniem jego funkcjonalności. Użytkownik musi sięgać po zewnętrzne narzędzia, które nie zawsze są łatwe w użyciu. Problem staje się jeszcze poważniejszy w przypadku, gdy program do dokowania w ogóle nie wygeneruje pozy o RMSD bliskim strukturze natywnej. Wtedy niemożliwe jest znalezienie przez LigandRNA poprawnego rozwiązania, gdyż takie nie istnieje. Dlatego też następnym krokiem w rozwoju programu LigandRNA powinna być próba stworzenia wewnętrznej funkcji do generowania póż, bądź też integracja LigandRNA z istniejącym narzędziem do dokowania (np. AutoDock).

Pierwszy krok w tym kierunku został już wykonany. Program LigandRNA może zostać użyty do oceny póż ligandów wygenerowanych przy pomocy Dock6, przy uwzględnieniu oceny przypisanej przez ten program. Dock6 szuka optymalnej pozy w przestrzeni rozwiązań poprzez identyfikację tej wiążącej się najkorzystniej pod względem energetycznym. Istnieje, więc możliwość, że Dock6 osiągnie minimum lokalne i nie znajdzie najlepszego rozwiązania. Dzięki asyście LigandRNA składowa oceny pochodzi również z potencjału statystycznego, co zwiększa szanse na odrzucenie póż z lokalnego minimum i znalezienie tej bliskiej natywnej.

Dyskusja

Na dzień dzisiejszy zarówno do wyznaczenia potencjału statystycznego, jak i do przewidywań LigandRNA, służą tylko cztery podstawowe typy reszt nukleotydów (adeniny, guaniny, cytozyny lub uracylu). Modyfikowane reszty są traktowane w taki sposób, jak gdyby były swoimi niemodyfikowanymi pierwowzorami. Oznacza to, że dodatkowe atomy z modyfikacji nie są brane pod uwagę w obliczeniach. Oraz, że w takich resztach część zdefiniowanych par atomów może zostać nieodnaleziona. Dowiedziono, że wpływ modyfikowanych reszt na statystyki i kształt potencjału jest znikomy. Zarówno brane pod uwagę (w formie opisanej powyżej), jak i zupełnie ignorowane w obliczeniach, nie zmieniają wyniku przewidywań (dane nie są prezentowane w niniejszej pracy). Jednakże, z biologicznego punktu widzenia ich obecność jest kluczowa dla wiązania się wielu ligandów. Na przykład metylacje znacząco zmniejszają siłę wiązania rRNA z antybiotykami. Dlatego też niezwykle istotne dla dalszego rozwoju LigandRNA byłoby uwzględnianie modyfikowanych reszt, a w szczególności dodatkowych atomów i nadanie im specjalnych wag w obliczeniach.

Wspólne perspektywy rozwoju MetalionRNA i LigandRNA

Kolejnym krokiem w rozwoju programów MetalionRNA i LigandRNA mogłaby być próba ich integracji zarówno ze sobą nawzajem jak i z innymi narzędziami do komputerowego modelowania struktur i ich kompleksów. Podejście takie wpisuje się w najnowsze trendy badań, które kładą nacisk na „globalne” podejścia do zagadnienia oddziaływań między cząsteczkami w komórce. Istnieje silna potrzeba badania struktur i ich dynamiki nie tylko w kontekście pojedynczych cząsteczek, czy kompleksów (np. białko-białko, białko-ligand, białko-kwas nukleinowy, kwas nukleinowy-ligand), ale coraz większy nacisk kładzie się na analizę dużych systemów takich jak kompleksy makromolekularne złożone z wielu komponentów. Szybki rozwój i coraz większa skuteczność technik do przewidywania struktur, analizy dynamiki molekularnej czy dokowania molekularnego pozwala na badanie całych systemów powiązanych ze sobą komponentów w taki sposób aby jak najlepiej odwzorować oddziaływania występujące w komórce. Takie kompleksowe podejście pozwala nie tylko badać coraz większe systemy, ale także przewidywać mechanizmy działania maszyn molekularnych regulujących podstawowe procesy życiowe w komórce, ale przede wszystkim daje szansę na odkrycie sposobów ich kontroli np. w celach terapeutycznych.

Połączenie obu metod dałoby możliwość dokowania ligandów do struktur RNA w warunkach przypominających ich prawdziwe środowisko w komórce, w którym występują jony oddziałujące z molekułami. Jest to o tyle kluczowe, że obecność jonów znajdujących się zarówno w rozpuszczalniku jak i związanych z RNA znacząco moduluje oddziaływanie RNA z innymi związkami. W związku z tym niezwykle istotne jest uwzględnienie wpływu kationów w procesie dokowania. Niestety większość programów do tego służących, zarówno tych dla białek, jak i RNA (np. Dock6) zupełnie ignoruje te cząsteczki, upuszczając w ten sposób obliczenia, ale również znacząco zmniejszając szansę na znalezienie poprawnego rozwiązania. Działanie takiej kompleksowej metody mogłoby odbywać się na następującej zasadzie. W pierwszym kroku, jeszcze przed przeprowadzeniem właściwego dokowania, moduł MetalionRNA dodałby jony do struktury RNA a następnie w kolejnym kroku moduł LigandRNA generowałby pozycje zadanych ligandów i ocenił je już z uwzględnieniem powstałego wpływu jonów. Ponieważ wiadomo jest, że niektóre ligandy tworzą wiązanie z RNA tylko i wyłącznie poprzez jon metalu, możliwa byłaby też funkcja dokowania liganda ze wcześniej zdefiniowanym jonom w strukturze receptora, który nie znajdowałby się tam bez obecności liganda.

Innym zastosowaniem programu MetalionRNA mogłoby być połączenie go z programami do komputerowego modelowania struktur przestrzennych RNA, z użyciem szablonu (Rother, Rother i wsp. 2011) czy też *de novo* (Das i Baker 2007). Takie połączenie dałoby nadzieję na otrzymanie modeli z jonami, przez co bliższych strukturom natywnym. Programy *ab initio* (Parisien i Major 2008, Das, Karanicolas i wsp. 2010) które symulują zwijanie się łańcucha RNA zyskałyby jednak najbardziej, gdyż jak wspomniano w Rozdziale 1.1 „Rola kationów w procesie zwijania się i stabilizacji struktur RNA” kationy mają kluczowy wpływ na ten proces i tylko w ich obecności struktura może przybrać poprawną formę trzeciorzędową.

Programy do modelowania wielopodjednostkowych makrokompleksów (Wriggers 2010, Tjioe, Lasker i wsp. 2011) również mogłyby zostać wzbogacone zarówno o moduł MetalionRNA jak i LigandRNA. Dodanie jonów metali czy ocena poprawności umiejscowienia zadanych ligandów pozwoliłoby na urealnienie generowanych modeli oraz poprawę ich jakości. Dodatkowym zastosowaniem modułu LigandRNA mogłaby być możliwość poszukiwania inhibitorów zbudowanych

Dyskusja

już kompleksów, które oddziaływałyby z kilkoma jednostki rozpatrywanego systemu (np. na ich pograniczu).

W dalszej perspektywie należałoby zaprojektować platformę wielozadaniową do modelowania struktur od pojedynczych cząsteczek (białek, kwasów nukleinowych), po kompleksy złożone z wielu komponentów, aż po całe systemy oddziałujących ze sobą kompleksów. Włączenie w tego typu procedury metod do przewidywania lokalizacji jonów i ligandów z całą pewnością zwiększyłoby skuteczność przewidywań, ale przede wszystkim upodobniłoby opis środowiska cząsteczek do tego występującego w komórce (bogatego w jony, ligandy i wodę).

Perspektywy zastosowania anizotropowego potencjału statystycznego

Opracowany przez autorkę niniejszej pracy anizotropowy potencjał statystyczny (patrz: Rozdział 3.1 „Potencjał statystyczny”) posiada nieocenioną cechę łatwego dostosowania go do opisu innych typów oddziaływań, takich jak RNA z białkami czy białek z białkami, jonami metali, ligandami. Do tego celu wymagana jest jedynie zmiana definicji par atomów receptora oraz typów atomów liganda. Przykładowo w przypadku przewidywania oddziaływań białek z jonami metali konieczne byłoby stworzenie listy par atomów reszt aminokwasowych oraz zdefiniowanie typów jonów z nimi oddziałujących. Wykonania tego zadania podjął się już lic. Krzysztof Formanowicz, który podczas swojego licencjatu zmodyfikował potencjał MetalionRNA oraz zaimplementował go w metodzie do identyfikacji miejsc wiązania kationów w strukturach białek. Metoda ta została nazwana MetalionProt i służy do przewidywania pozycji pięciu rodzajów kationów (sodu, potasu, magnezu, cynku i żelaza). Przeprowadzone testy oraz porównanie z analogicznymi metodami dały zadowalające wyniki, co świadczy o uniwersalności proponowanego potencjału i możliwości przeniesienia go na kolejne typy oddziaływań (np. białek z ligandami). Program MetalionProt dalej będzie rozwijany w grupie prof. Bujnickiego, a sam potencjał zostanie włączony do nowego programu do modelowania homologicznego struktur przestrzennych białek uwzględniającego obecność jonów.

5.5. Modelowanie struktury kompleksów RNA z ligandami a komputerowe projektowanie leków

Ostatnimi laty znaleziono wiele katalitycznych cząsteczek RNA, które poprzez oddziaływanie ze specyficznymi ligandami pełnią niebagatelne funkcje w komórce. Odkrycia te spowodowały spekulacje na temat możliwości zastosowania tych cząsteczek i ich właściwości w nauce czy medycynie. Wiele grup podjęło tę tematykę i obecnie prowadzi badania nad oddziaływaniami RNA z ligandami. Rosnące zainteresowania badaczy tym polem nauki spowodowało też duży postęp w dziedzinie komputerowego przewidywania oddziaływań. Metody teoretyczne są szybsze i tańsze od tradycyjnych doświadczeń laboratoryjnych. Dlatego też wzrasta potrzeba ich tworzenia, a stosowanie jest coraz powszechniejsze. O tej tendencji świadczy chociażby fakt, że na tegorocznym konkursie RNA Puzzles jednym ze zadań było wymodelowanie struktury ryboprzełącznika i jego kompleksu z zadaniem ligandem, 5-hydrokсыtryptofanem.

Komputerowe poszukiwanie inhibitorów ryboprzełączników

Stosunkowo duża liczba ryboprzełączników występująca w bakteriach skłania do prób ich wykorzystania jako celów dla leków przeciwbakteryjnych. Na przykład, prowadzone są badania nad wykorzystaniem ryboprzełącznika guaninowego w terapii przeciw infekcjom powodowanym przez *S. aureus* (Ster, Allard i wsp. 2013). Jest on interesującym celem, gdyż znany jest analog liganda wiążącego się z tym ryboprzełącznikiem (2,5,6-triaminopirymidyn-4-ol, PC1), który blokuje syntezę monofosforanu guanozyny, co skutkuje śmiercią bakterii. Trwają również prace nad znalezieniem inhibitorów innych ryboprzełączników. Poszukiwane są homologi ligandów, które silnie wiązałyby się z wybranym przełącznikiem RNA, ale nie oddziaływały równocześnie z innymi RNA lub białkami w komórce (Blount i Breaker 2006, Long, Ji i wsp. 2010). Poza klasycznymi badaniami przesiewowymi *in vitro* istnieją dwa podejścia teoretyczne do tego problemu. Pierwsze z nich, to zastosowanie VS jako pierwszego etapu badań. Pozwala to na wstępne przesianie tysięcy związków w stosunkowo krótkim czasie i przy minimalnych kosztach, by wytypować zbiór związków do dalszych analiz. Technika ta jest już z powodzeniem stosowana podczas poszukiwania ligandów wiążących się z białkami (Badrinarayan i Sastry 2011, Ou-Yang, Lu i wsp. 2012). Obecnie są podejmowane pierwsze próby jej wykorzystania dla celów RNA (Penchovsky i Stoilova 2013). Przykładowo, odkryto cztery nowe

Dyskusja

ligandy oddziałujące z ryboprzełącznikami purynowymi, co zostało potwierdzone doświadczalnie (Daldrop, Reyes i wsp. 2011). Program LigandRNA doskonale wpisuje się w trend tych badań, gdyż oferuje możliwość oceny pól ligandów z VS. Inną techniką jest próba zaprojektowania związku w zależności od konformacji i właściwości fizyko-chemicznych receptora (Billingsley 2008, Jentzsch i Hines 2012).

Ograniczenia metod komputerowych

Struktury rybozymów charakteryzują się dużo mniejszą stabilnością niż enzymy białkowe. Sprawia to, że dokowanie do kwasów nukleinowych jest tematyką znacznie trudniejszą, ze względu na duże zmiany konformacyjne, jakie mogą zachodzić podczas wiązania się liganda. By zwiększyć stabilności struktur wprowadzane są modyfikacje zasad reszt nukleotydowych, lecz te z kolei mogą negatywnie wpływać na zdolności katalityczne rybozymu (Sioud 2006). Kolejnym problemem jest wybór miejsc docelowych w mRNA, który jest kluczowy dla przeprowadzenia VS, dokowania czy chemioinformatycznego zaprojektowania leku. Należy również pamiętać, że wyniki otrzymane metodami *in silico* czy *de novo* są przewidywaniami komputerowymi i zawsze wymagają potwierdzenia przy pomocy badań doświadczalnych.

6. Wnioski

W wyniku prac, które opisane zostały w niniejszej rozprawie doktorskiej, opracowano:

- anizotropowy potencjał statystyczny opisujący oddziaływania RNA z jonami magnezu, sodu i potasu;
- narzędzie MetalionRNA do przewidywania miejsc wiązania się jonów magnezu, sodu i potasu w strukturach przestrzennych RNA;
- anizotropowy potencjał statystyczny opisujący oddziaływania RNA z niskocząsteczkowymi ligandami;
- narzędzie LigandRNA do oceny i rankingu pól liganda/ów w strukturach przestrzennych RNA.

Załączniki

Potencjał statystyczny (CD)

Na płycie CD załączony został potencjał statystyczny.

Kod programów MetalionRNA oraz LigandRNA (CD)

Na płycie CD załączony został kod programów MetalionRNA, LigandRNA oraz skrypty służące do wyliczenia potencjałów statystycznych.

Zbiór plików PDB wykorzystanych do wytrenowania potencjałów oraz przeprowadzenia testów programów (CD)

Wszystkie pliki wykorzystane do wyprowadzenia potencjałów, a także do przeprowadzenia testów programów MetalionRNA oraz LigandRNA, nagrano na płytę CD dołączoną do rozprawy doktorskiej. Dla każdego testowanego kompleksu RNA z ligandem dołączono także zestaw wygenerowanych pól.

Spis rycin

Rycina 1. Trzy typy wiązania się jonów Mg^{2+} z RNA. Na czerwono zaznaczono tlen, na szaro wodór, natomiast magnez zaznaczony został na zielono. Panel A) Mg^{2+} bezpośrednio związany z atomami RNA. Panel B) Mg^{2+} oddziałujący z RNA poprzez jedną warstwę wody. Panel C) Mg^{2+} oddziałujący z RNA poprzez więcej niż jedną warstwę wody. (Draper, Grilley i wsp. 2005).	15
Rycina 2. Erytromycyna związana w dużej podjednostce rybosomu <i>Deinococcus</i> (kod PDB: 1JZY).	24
Rycina 3. Hygromycyna B związana w małej podjednostce rybosomu <i>Thermus</i> (kod PDB: 1HNZ).	26
Rycina 4. Panel A) idea potencjału izotropowego, zależnego tylko od odległości, w tym przypadku między atomem RNA O4' a Mg^{2+} (zielona sfera). Panel B) idea potencjału anizotropowego zależnego zarówno od odległości między parą atomów RNA O4' a Mg^{2+} jak i od kąta tworzonego przez atomy pary [C4', O4'] i Mg^{2+} .	35
Rycina 5. Panel A) Potencjał statystyczny otrzymany dla pary RNA [P, OP2] (na rycinie [a, b]) oraz jonów Mg^{2+} (na rycinie przykładowe położenie tego jonu c). Panel B) Znormalizowany potencjał statystyczny, przedstawiony na Panelu A. Potencjał ukazany jest w skali szarości (im ciemniejszy kolor tym mniejsza wartość potencjału).	37
Rycina 6. Potencjał statystyczny dla pary atomów RNA [P, OP2] i kationu Mg^{2+} . Trzy zaznaczone obszary (po dwa razy każdy, ponieważ wykres jest symetryczny) odpowiadają trzem możliwym stanom które przyjmuje Mg^{2+} związany z RNA (patrz: Rozdział 1.1 „Charakterystyka oddziaływań RNA z Mg^{2+} ”)(Draper 2004). Źródło: (Philips, Milanowska i wsp. 2012).	50
Rycina 7. Przykładowy fragment pliku z potencjałem statystycznym wyprowadzonym dla pary atomów RNA [C1', O4'] z łańcucha głównego oraz jonu Na^+ . W pierwszej kolumnie znajduje się wartość odległości wyrażona w Angstromach, w drugiej wartość kąta wyrażona w stopniach, natomiast trzecia kolumna zawiera wartość potencjału dla tego przedziału przestrzeni.	51

Rycina 8. Rozkład potencjału dla par: Panel A) [N1, C6] i C.ar. Panel B) [C5', O5'] i O.3. Panel C) [P, OP1] i N.3. Panel D) [P, OP1] i C.3.....	53
Rycina 9. Schematyczna ilustracja sposobu obliczania potencjału statystycznego dla węzłów sieci w odległości $<9 \text{ \AA}$ od dwóch par atomów RNA [P, OP2] i identyfikacja najbardziej prawdopodobnych miejsc wiązania jonu metalu. Źródło: (Philips, Milanowska i wsp. 2012).....	55
Rycina 10. Krzywe ROC dla zestawu struktur RNA z Mg^{2+} . Mały wykres przedstawia zakres od 0 do $0,2 \text{ \AA}$ całego wykresu w skali logarytmicznej.....	58
Rycina 11. Krzywe ROC dla zestawu struktur RNA z Na^+ . Mały wykres przedstawia zakres od 0 do $0,2 \text{ \AA}$ całego wykresu w skali logarytmicznej.....	58
Rycina 12. Krzywe ROC dla zestawu struktur RNA z K^+ . Mały wykres przedstawia zakres od 0 do $0,2 \text{ \AA}$ całego wykresu w skali logarytmicznej.....	59
Rycina 13. Krzywe ROC dla zestawu struktur DNA z Mg^{2+} . Mały wykres przedstawia zakres od 0 do $0,2 \text{ \AA}$ całego wykresu w skali logarytmicznej.....	59
Rycina 14. Panel A) Struktura rRNA (kod PDB: 1HC8) z doświadczalnie wyznaczonymi pozycjami jonów Mg^{2+} (zielone sfery z etykietami). Najlepsze przewidywania dla Mg^{2+} programu MetalionRNA ukazane zostały jako czerwone sfery. Panel B) przedstawia powierzchniowy anizotropowy potencjał statystyczny obliczony dla tej struktury. Im kolor bardziej czerwony tym mniejszy potencjał i większe prawdopodobieństwo wiązania się jonu w tym miejscu.....	63
Rycina 15. Przykłady zależności oceny pozy i jej wartości RMSD do struktury referencyjnej. Strzałką zostały zaznaczone pozy zidentyfikowane przez metody jako najlepsze.....	69
Rycina 16. Cztery związki wytypowane do testów doświadczalnych. Kody ZINC: Panel A) ZINC32198129. Panel B) ZINC00089474. Panel C) ZINC01509324. Panel D) ZINC26654647. Struktura receptora – kod PDB: 3OFQ.	72

Rycina 17. Panel A) Aptamer RNA w kompleksie z cytruliną (kod PDB: 1KOD). Panel B) Centrum dekodujące rybosomu w kompleksie z apramycyną (pdb id: 2OE5). Doświadczalnie rozwiązane pozycje ligandów przedstawione są na jasnoniebiesko, natomiast pozycje wygenerowane przy pomocy Dock6 i ocenione najwyżej przez LigandRNA ukazane są na różowo.	78
Rycina 18. Panel A) rRNA centrum dekodującego w kompleksie z aminoglikozydem z grupą L-HABA. Panel B) rRNA miejsca A w kompleksie z paromycyną. Doświadczalnie rozwiązane pozycje ligandów przedstawione są na jasnoniebiesko, natomiast pozycje wygenerowane przy pomocy Dock6 i ocenione najwyżej przez kombinację ocen LigandRNA-Dock6 ukazane są na różowo.	79
Rycina 19. Struktura modelu ryboprzełącznika oraz dwie najwyżej ocenione pozycje 5-hydroksytryptofanu z nim związane.	80
Rycina 20. Zrzut ekranu ze strony powitalnej serwera LigandRNA.	82

Spis tabel

Tabela 1. Znane ligandy oddziałujące z ryboprzełącznikami oraz ich charakterystyka.	19
Tabela 2. Wybrane antybiotyki wiążące się z rRNA w dużej podjednostce rybosomu.	23
Tabela 3. Wybrane antybiotyki wiążące się z rRNA w małej podjednostce rybosomu.	25
Tabela 4. Lista kodów z bazy danych PDB struktur RNA z jonami magnezu, sodu i potasu, użytych do wyprowadzenia potencjału statystycznego oraz przeprowadzenia walidacji krzyżowej.	38
Tabela 5. Lista kodów z bazy danych PDB struktur DNA z jonami magnezu, użytych do dodatkowych testów.	39
Tabela 6. Lista kodów z bazy danych PDB struktur ze zbioru uczącego programu FEATURE.	39
Tabela 7. Lista kodów z bazy danych PDB struktur kompleksów RNA z ligandami, użytych do wyprowadzenia potencjału statystycznego oraz przeprowadzenia walidacji krzyżowej.	40
Tabela 8. Klastry ligandów i ich identyfikatory PDB.	41
Tabela 9. Zbiór testowy programu LigandRNA.	43
Tabela 10. Sprzęt komputerowy.	48
Tabela 11. Pary atomów RNA zdefiniowane do wyprowadzenia potencjału RNA z jonami metali.	49
Tabela 12. Pary atomów RNA zdefiniowane do wyprowadzenia potencjału RNA z ligandami.	52
Tabela 13. Porównanie wyników działania programu MetalionRNA (z oryginalnym potencjałem i potencjałem wyprowadzonym ze zbioru uczącego programu FEATURE) z wynikami działania programu FEATURE. Przewidywania były prowadzone dla struktury rRNA z podjednostki 23S (kod PDB: 1HC8). Jony „silnie” związane zostały oznaczone asteriskiem.	62

Tabela 14. Procent najlepiej ocenionych póz o RMSD do struktury natywnej odpowiednio mniejszym niż 1, 1,5; 2 Å, uzyskany dla zbioru 42 kompleksów, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD <2 Å.....66

Tabela 15. Procent najlepiej ocenionych póz o RMSD do struktury natywnej odpowiednio mniejszym niż 1, 1,5; 2 Å, uzyskany dla podzbioru kompleksów, dla których Dock6 wygenerował przynajmniej jedną pozę o RMSD <1 Å.....66

Tabela 16. Wyniki testu programu przeprowadzonego na wybranych 42 kompleksach RNA z ligandami. Pierwsza kolumna zawiera kody PDB kompleksów, następane wartości RMSD (w Angstromach) najlepiej ocenionych póz przez analizowane programy.67

Tabela 17. Najwyżej ocenione związki z poszczególnych grup (podział według liczby atomów ciężkich) przez LigandRNA, związki wybrane manualnie do dalszych testów doświadczalnych zostały zaznaczone na czerwono.72

Bibliografia

- Ames, T. D. i R. R. Breaker (2011). "Bacterial aptamers that selectively bind glutamine." RNA Biol **8**(1): 82-89.
- Auslander, S., P. Ketzer i J. S. Hartig (2010). "A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression." Mol Biosyst **6**(5): 807-814.
- Badrinarayan, P. i G. N. Sastry (2011). "Virtual high throughput screening in new lead identification." Comb Chem High Throughput Screen **14**(10): 840-860.
- Baker, J. L., N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge i R. R. Breaker (2012). "Widespread genetic switches and toxicity resistance proteins for fluoride." Science **335**(6065): 233-235.
- Ban, N., P. Nissen, J. Hansen, M. Capel, P. B. Moore i T. A. Steitz (1999). "Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit." Nature **400**(6747): 841-847.
- Batey, R. T. (2011). "Recognition of S-adenosylmethionine by riboswitches." Wiley Interdiscip Rev RNA **2**(2): 299-311.
- Beebe, J. A., J. C. Kurz i C. A. Fierke (1996). "Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNA^{Asp}." Biochemistry **35**(32): 10493-10505.
- Berens, C., B. Streicher, R. Schroeder i W. Hillen (1998). "Visualizing metal-ion-binding sites in group I introns by iron(II)-mediated Fenton reactions." Chem Biol **5**(3): 163-175.
- Berisio, R., J. Harms, F. Schluenzen, R. Zarivach, H. A. Hansen, P. Fucini i A. Yonath (2003). "Structural insight into the antibiotic action of telithromycin against resistant mutants." J Bacteriol **185**(14): 4276-4279.
- Berisio, R., F. Schluenzen, J. Harms, A. Bashan, T. Auerbach, D. Baram i A. Yonath (2003). "Structural insight into the role of the ribosomal tunnel in cellular regulation." Nat Struct Biol **10**(5): 366-370.
- Billingsley, M. L. (2008). "Druggable targets and targeted drugs: enhancing the development of new therapeutics." Pharmacology **82**(4): 239-244.
- Blount, K. F. i R. R. Breaker (2006). "Riboswitches as antibacterial drug targets." Nat Biotechnol **24**(12): 1558-1564.
- Borda, E. J. i S. T. Sigurdsson (2004). "Interactions of the antibiotics neomycin B and chlortetracycline with the hammerhead ribozyme as studied by Zn²⁺-dependent RNA cleavage." Bioorg Med Chem **12**(5): 1023-1028.

Borovok, I., B. Gorovitz, R. Schreiber, Y. Aharonowitz i G. Cohen (2006). "Coenzyme B12 controls transcription of the *Streptomyces* class Ia ribonucleotide reductase nrdABS operon via a riboswitch mechanism." J Bacteriol **188**(7): 2512-2520.

Bottegoni, G. (2011). "Protein-ligand docking." Front Biosci **16**: 2289-2306.

Bottger, E. C. (2006). "The ribosome as a drug target." Trends Biotechnol **24**(4): 145-147.

Brodersen, D. E., W. M. Clemons, Jr., A. P. Carter, R. J. Morgan-Warren, B. T. Wimberly i V. Ramakrishnan (2000). "The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit." Cell **103**(7): 1143-1154.

Brooks, B. R., C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York i M. Karplus (2009). "CHARMM: the biomolecular simulation program." J Comput Chem **30**(10): 1545-1614.

Carter, A. P., W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly i V. Ramakrishnan (2000). "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics." Nature **407**(6802): 340-348.

Chen, P. J., H. L. Wu, C. J. Wang, J. H. Chia i D. S. Chen (1997). "Molecular biology of hepatitis D virus: research and potential for application." J Gastroenterol Hepatol **12**(9-10): S188-192.

Christian, E. L., N. M. Kaye i M. E. Harris (2000). "Helix P4 is a divalent metal ion binding site in the conserved core of the ribonuclease P ribozyme." Rna **6**(4): 511-519.

Cochrane, J. C., S. V. Lipchok i S. A. Strobel (2007). "Structural investigation of the GlnS ribozyme bound to Its catalytic cofactor." Chem Biol **14**(1): 97-105.

Cock, P. J., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski i M. J. de Hoon (2009). "Biopython: freely available Python tools for computational molecular biology and bioinformatics." Bioinformatics **25**(11): 1422-1423.

Corbino, K. A., J. E. Barrick, J. Lim, R. Welz, B. J. Tucker, I. Puskarz, M. Mandal, N. D. Rudnick i R. R. Breaker (2005). "Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria." Genome Biol **6**(8): R70.

Cruz, J. A., M. F. Blanchet, M. Boniecki, J. M. Bujnicki, S. J. Chen, S. Cao, R. Das, F. Ding, N. V. Dokholyan, S. C. Flores, L. Huang, C. A. Lavender, V. Lisi, F. Major, K. Mikolajczak, D. J. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K. M. Weeks, C. Waldsich, M. Wildauer, N. B. Leontis i E. Westhof

- (2012). "RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction." RNA **14**(4): 610-625.
- Daldrop, P., F. E. Reyes, D. A. Robinson, C. M. Hammond, D. M. Lilley, R. T. Batey i R. Brenk (2011). "Novel ligands for a purine riboswitch discovered by RNA-ligand docking." Chem Biol **18**(3): 324-335.
- Das, R. i D. Baker (2007). "Automated de novo prediction of native-like RNA tertiary structures." Proc Natl Acad Sci U S A **104**(37): 14664-14669.
- Das, R., J. Karanicolas i D. Baker (2010). "Atomic accuracy in predicting and designing noncanonical RNA structure." Nat Methods **7**(4): 291-294.
- Deigan, K. E. i A. R. Ferre-D'Amare (2011). "Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs." Acc Chem Res **44**(12): 1329-1338.
- DeLano, W. L. (2002). "The PyMOL Molecular Graphics System."
- DeRose, V. J. (2002). "Two decades of RNA catalysis." Chem Biol **9**(9): 961-969.
- DeRose, V. J. (2003). "Metal ion binding to catalytic RNA molecules." Curr Opin Struct Biol **13**(3): 317-324.
- Detering, C. i G. Varani (2004). "Validation of automated docking programs for docking and database screening against RNA drug targets." J Med Chem **47**(17): 4188-4201.
- Dieterich, C. i P. F. Stadler (2013). "Computational biology of RNA interactions." Wiley Interdiscip Rev RNA **4**(1): 107-120.
- Douthwaite, S. (1992). "Functional interactions within 23S rRNA involving the peptidyltransferase center." J Bacteriol **174**(4): 1333-1338.
- Douthwaite, S., D. Fourmy i S. Yoshizawa (2005). "Nucleotide methylations in rRNA that confer resistance to ribosome-targeting antibiotics." Fine-tuning of RNA functions by modification and editing **12**: 287-309.
- Draper, D. E. (2004). "A guide to ions and RNA structure." RNA **10**(3): 335-343.
- Draper, D. E. (2008). "RNA folding: thermodynamic and molecular descriptions of the roles of ions." Biophys J **95**(12): 5489-5495.
- Draper, D. E., D. Grilley i A. M. Soto (2005). "Ions and RNA folding." Annu Rev Biophys Biomol Struct **34**: 221-243.
- Dunkle, J. A., L. Xiong, A. S. Mankin i J. H. Cate (2010). "Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action." Proc Natl Acad Sci U S A **107**(40): 17152-17157.
- Eddy, S. R. (2001). "Non-coding RNA genes and the modern RNA world." Nat Rev Genet **2**(12): 919-929.

- Edwards, A. L., F. E. Reyes, A. Heroux i R. T. Batey (2010). "Structural basis for recognition of S-adenosylhomocysteine by riboswitches." Rna **16**(11): 2144-2155.
- Ennifar, E., J. C. Paillart, A. Bodlener, P. Walter, J. M. Weibel, A. M. Aubertin, P. Pale, P. Dumas i R. Marquet (2006). "Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell." Nucleic Acids Res **34**(8): 2328-2339.
- Epshtein, V., A. S. Mironov i E. Nudler (2003). "The riboswitch-mediated control of sulfur metabolism in bacteria." Proc Natl Acad Sci U S A **100**(9): 5052-5056.
- Erdmann, V. V. A., J. Barciszewski i J. Brosius (2006). RNA towards medicine, Springer-Verlag Berlin Heidelberg.
- Ferre-D'Amare, A. R. i W. G. Scott (2010). "Small self-cleaving ribozymes." Cold Spring Harb Perspect Biol **2**(10): a003574.
- Foloppe, N., N. Matassova i F. Aboul-Ela (2006). "Towards the discovery of drug-like RNA ligands?" Drug Discov Today **11**(21-22): 1019-1027.
- Fuchs, R. T., F. J. Grundy i T. M. Henkin (2006). "The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase." Nat Struct Mol Biol **13**(3): 226-233.
- Fulle, S. i H. Gohlke (2010). "Molecular recognition of RNA: challenges for modelling interactions and plasticity." J Mol Recognit **23**(2): 220-231.
- Garst, A. D., A. L. Edwards i R. T. Batey (2011). "Riboswitches: structures and mechanisms." Cold Spring Harb Perspect Biol **3**(6).
- Geissmann, T., M. Possedko, E. Huntzinger, P. Fechter, C. Ehresmann i P. Romby (2006). "Regulatory RNAs as mediators of virulence gene expression in bacteria." Handb Exp Pharmacol(173): 9-43.
- Gohlke, H., M. Hendlich i G. Klebe (2000). "Knowledge-based scoring function to predict protein-ligand interactions." J Mol Biol **295**(2): 337-356.
- Gripenland, J., S. Netterling, E. Loh, T. Tiensuu, A. Toledo-Arana i J. Johansson (2010). "RNAs: regulators of bacterial virulence." Nature reviews. Microbiology **8**(12): 857-866.
- Grosshans, H. i W. Filipowicz (2008). "Molecular biology: the expanding world of small RNAs." Nature **451**(7177): 414-416.
- Grundy, F. J., S. C. Lehman i T. M. Henkin (2003). "The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes." Proc Natl Acad Sci U S A **100**(21): 12057-12062.
- Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace i S. Altman (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell **35**(3 Pt 2): 849-857.

- Guilbert, C. i T. L. James (2008). "Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets." J Chem Inf Model **48**(6): 1257-1268.
- Hampel, K. J. i M. M. Tinsley (2006). "Evidence for preorganization of the glmS ribozyme ligand binding pocket." Biochemistry **45**(25): 7861-7871.
- Han, Q., Q. Zhao, S. Fish, K. B. Simonsen, D. Vourloumis, J. M. Froelich, D. Wall i T. Hermann (2005). "Molecular recognition by glycoside pseudo base pairs and triples in an apramycin-RNA complex." Angew Chem Int Ed Engl **44**(18): 2694-2700.
- Hansen, J. L., J. A. Ippolito, N. Ban, P. Nissen, P. B. Moore i T. A. Steitz (2002). "The structures of four macrolide antibiotics bound to the large ribosomal subunit." Mol Cell **10**(1): 117-128.
- Hansen, J. L., P. B. Moore i T. A. Steitz (2003). "Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit." J Mol Biol **330**(5): 1061-1075.
- Harms, J. M., F. Schlunzen, P. Fucini, H. Bartels i A. Yonath (2004). "Alterations at the peptidyl transferase centre of the ribosome induced by the synergistic action of the streptogramins dalbavancin and quinupristin." BMC Biol **2**: 4.
- Heilman-Miller, S. L., D. Thirumalai i S. A. Woodson (2001). "Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations." J Mol Biol **306**(5): 1157-1166.
- Hermann, T. (2005). "Drugs targeting the ribosome." Curr Opin Struct Biol **15**(3): 355-366.
- Hermann, T. i E. Westhof (1998). "Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations." Structure **6**(10): 1303-1314.
- Holbrook, S. R. (2008). "Structural principles from large RNAs." Annu Rev Biophys **37**: 445-464.
- Holmstrom, E. D., J. L. Fiore i D. J. Nesbitt (2012). "Thermodynamic origins of monovalent facilitated RNA folding." Biochemistry **51**(18): 3732-3743.
- Hoogstraten, C. G. i M. Sumita (2007). "Structure-function relationships in RNA and RNP enzymes: recent advances." Biopolymers **87**(5-6): 317-328.
- Irwin, J. J., T. Sterling, M. M. Mysinger, E. S. Bolstad i R. G. Coleman (2012). "ZINC: A Free Tool to Discover Chemistry for Biology." J Chem Inf Model.
- Jain, A. N. (2003). "Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine." J Med Chem **46**(4): 499-511.
- Jentsch, F. i J. V. Hines (2012). "Interfacing medicinal chemistry with structural bioinformatics: implications for T box riboswitch RNA drug discovery." BMC Bioinformatics **13 Suppl 2**: S5.

- Johansen, L. E., P. Nygaard, C. Lassen, Y. Agerso i H. H. Saxild (2003). "Definition of a second *Bacillus subtilis* pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL)." J Bacteriol **185**(17): 5200-5209.
- Kellenberger, E., J. Rodrigo, P. Muller i D. Rognan (2004). "Comparative evaluation of eight docking tools for docking and virtual screening accuracy." Proteins **57**(2): 225-242.
- Kieft, J. S., E. Chase, D. A. Costantino i B. L. Golden (2010). "Identification and characterization of anion binding sites in RNA." RNA **16**(6): 1118-1123.
- Klein, D. J. i A. R. Ferre-D'Amare (2006). "Structural basis of glmS ribozyme activation by glucosamine-6-phosphate." Science **313**(5794): 1752-1756.
- Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling i T. R. Cech (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*." Cell **31**(1): 147-157.
- Kuntz, I. D., J. M. Blaney, S. J. Oatley, R. Langridge i T. E. Ferrin (1982). "A geometric approach to macromolecule-ligand interactions." J Mol Biol **161**(2): 269-288.
- Lang, P. T., S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James i I. D. Kuntz (2009). "DOCK 6: combining techniques to model RNA-small molecule complexes." Rna **15**(6): 1219-1230.
- Li, C. F., M. Costa, G. Bassi, Y. K. Lai i F. Michel (2011). "Recurrent insertion of 5'-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs." Rna **17**(7): 1321-1335.
- Li, Y., J. Shen, X. Sun, W. Li, G. Liu i Y. Tang (2010). "Accuracy assessment of protein-based docking programs against RNA targets." J Chem Inf Model **50**(6): 1134-1146.
- Liang, M. P., D. R. Banatao, T. E. Klein, D. L. Brutlag i R. B. Altman (2003). "WebFEATURE: An interactive web tool for identifying and visualizing functional sites on macromolecular structures." Nucleic Acids Res **31**(13): 3324-3327.
- Long, Q., L. Ji, H. Wang i J. Xie (2010). "Riboflavin biosynthetic and regulatory factors as potential novel anti-infective drug targets." Chem Biol Drug Des **75**(4): 339-347.
- MacKerell, A. D., Jr., N. Banavali i N. Foloppe (2000). "Development and current status of the CHARMM force field for nucleic acids." Biopolymers **56**(4): 257-265.
- Mandal, M., B. Boese, J. E. Barrick, W. C. Winkler i R. R. Breaker (2003). "Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria." Cell **113**(5): 577-586.
- Mandal, M. i R. R. Breaker (2004). "Adenine riboswitches and gene activation by disruption of a transcription terminator." Nat Struct Mol Biol **11**(1): 29-35.

- Mandal, M., M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo i R. R. Breaker (2004). "A glycine-dependent riboswitch that uses cooperative binding to control gene expression." Science **306**(5694): 275-279.
- Markham, G. D., J. P. Glusker i C. W. Bock (2002). "The arrangement of first- and second-sphere water molecules in divalent magnesium complexes: results from molecular orbital and density functional theory and from structural crystallography." J Phys Chem B **106**(19): 5118-5134.
- Mastroiannopoulos, N. P., J. B. Uney i L. A. Phylactou (2010). "The application of ribozymes and DNAzymes in muscle and brain." Molecules **15**(8): 5460-5472.
- Mironov, A. S., I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov i E. Nudler (2002). "Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria." Cell **111**(5): 747-756.
- Misra, V. K. i D. E. Draper (2000). "Mg(2+) binding to tRNA revisited: the nonlinear Poisson-Boltzmann model." J Mol Biol **299**(3): 813-825.
- Moitessier, N., E. Westhof i S. Hanessian (2006). "Docking of aminoglycosides to hydrated and flexible RNA." J Med Chem **49**(3): 1023-1033.
- Montange, R. K. i R. T. Batey (2008). "Riboswitches: emerging themes in RNA structure and function." Annu Rev Biophys **37**: 117-133.
- Morley, S. D. i M. Afshar (2004). "Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock." J Comput Aided Mol Des **18**(3): 189-208.
- Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell i A. J. Olson (2009). "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility." J Comput Chem **30**(16): 2785-2791.
- Mulhbacher, J., E. Brouillette, M. Allard, L. C. Fortier, F. Malouin i D. A. Lafontaine (2010). "Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways." PLoS Pathog **6**(4): e1000865.
- Mulhbacher, J., P. St-Pierre i D. A. Lafontaine (2010). "Therapeutic applications of ribozymes and riboswitches." Curr Opin Pharmacol **10**(5): 551-556.
- Murray, J. B., A. A. Seyhan, N. G. Walter, J. M. Burke i W. G. Scott (1998). "The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone." Chem Biol **5**(10): 587-595.
- Nahvi, A., N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown i R. R. Breaker (2002). "Genetic control by a metabolite binding mRNA." Chem Biol **9**(9): 1043.
- Nakano, S., D. M. Chadalavada i P. C. Bevilacqua (2000). "General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme." Science **287**(5457): 1493-1497.

- O'Rear, J. L., S. Wang, A. L. Feig, L. Beigelman, O. C. Uhlenbeck i D. Herschlag (2001). "Comparison of the hammerhead cleavage reactions stimulated by monovalent and divalent cations." Rna **7**(4): 537-545.
- Ou-Yang, S. S., J. Y. Lu, X. Q. Kong, Z. J. Liang, C. Luo i H. Jiang (2012). "Computational drug discovery." Acta Pharmacol Sin **33**(9): 1131-1140.
- Parisien, M. i F. Major (2008). "The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data." Nature **452**(7183): 51-55.
- Park, I. K., J. Y. Kim, E. H. Lim i S. Shin (2000). "Spectinomycin inhibits the self-splicing of the group 1 intron RNA." Biochem Biophys Res Commun **269**(2): 574-579.
- Patel, U., Y. P. Yan, F. W. Hobbs, Jr., J. Kaczmarczyk, A. M. Slee, D. L. Pompliano, M. G. Kurilla i E. V. Bobkova (2001). "Oxazolidinones mechanism of action: inhibition of the first peptide bond formation." J Biol Chem **276**(40): 37199-37205.
- Penchovsky, R. i C. C. Stoilova (2013). "Riboswitch-based antibacterial drug discovery using high-throughput screening methods." Expert Opin Drug Discov **8**(1): 65-82.
- Petrov, A. S., J. C. Bowman, S. C. Harvey i L. D. Williams (2011). "Bidentate RNA-magnesium clamps: on the origin of the special role of magnesium in RNA folding." Rna **17**(2): 291-297.
- Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng i T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.
- Pfeffer, P. i H. Gohlke (2007). "DrugScoreRNA--knowledge-based scoring function to predict RNA-ligand interactions." J Chem Inf Model **47**(5): 1868-1876.
- Philips, A., K. Milanowska, G. Lach, M. Boniecki, K. Rother i J. M. Bujnicki (2012). "MetalionRNA: computational predictor of metal-binding sites in RNA structures." Bioinformatics **28**(2): 198-205.
- Piccirilli, J. A., J. S. Vyle, M. H. Caruthers i T. R. Cech (1993). "Metal ion catalysis in the Tetrahymena ribozyme reaction." Nature **361**(6407): 85-88.
- Pioletti, M., F. Schlunzen, J. Harms, R. Zarivach, M. Gluhmann, H. Avila, A. Bashan, H. Bartels, T. Auerbach, C. Jacobi, T. Hartsch, A. Yonath i F. Franceschi (2001). "Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3." Embo J **20**(8): 1829-1839.
- Poehlsgaard, J. i S. Douthwaite (2005). "The bacterial ribosome as a target for antibiotics." Nat Rev Microbiol **3**(11): 870-881.
- Porse, B. T. i R. A. Garrett (1999). "Sites of interaction of streptogramin A and B antibiotics in the peptidyl transferase loop of 23 S rRNA and the synergism of their inhibitory mechanisms." J Mol Biol **286**(2): 375-387.

- Poulsen, S. M., C. Kofoed i B. Vester (2000). "Inhibition of the ribosomal peptidyl transferase reaction by the mycarose moiety of the antibiotics carbomycin, spiramycin and tylosin." J Mol Biol **304**(3): 471-481.
- Pyle, A. M. (2002). "Metal ions in the structure and function of RNA." J Biol Inorg Chem **7**(7-8): 679-690.
- Reiter, N. J., A. K. Osterman i A. Mondragon (2012). "The bacterial ribonuclease P holoenzyme requires specific, conserved residues for efficient catalysis and substrate positioning." Nucleic Acids Res.
- Rivas, E. i S. R. Eddy (2001). "Noncoding RNA gene detection using comparative sequence analysis." BMC Bioinformatics **2**: 8.
- Rodionov, D. A., A. G. Vitreschak, A. A. Mironov i M. S. Gelfand (2002). "Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms." J Biol Chem **277**(50): 48949-48959.
- Rodionov, D. A., A. G. Vitreschak, A. A. Mironov i M. S. Gelfand (2003). "Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch?" Nucleic Acids Res **31**(23): 6748-6757.
- Romer, R. i R. Hach (1975). "tRNA conformation and magnesium binding. A study of a yeast phenylalanine-specific tRNA by a fluorescent indicator and differential melting curves." Eur J Biochem **55**(1): 271-284.
- Rose, P. W., B. Beran, C. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlic, M. Quesada, G. B. Quinn, J. D. Westbrook, J. Young, B. Yukich, C. Zardecki, H. M. Berman i P. E. Bourne (2011). "The RCSB Protein Data Bank: redesigned web site and web services." Nucleic Acids Res **39**(Database issue): D392-401.
- Roth, A., W. C. Winkler, E. E. Regulski, B. W. Lee, J. Lim, I. Jona, J. E. Barrick, A. Ritwik, J. N. Kim, R. Welz, D. Iwata-Reuyl i R. R. Breaker (2007). "A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain." Nat Struct Mol Biol **14**(4): 308-317.
- Rother, M., K. Rother, T. Puton i J. M. Bujnicki (2011). "ModeRNA: a tool for comparative modeling of RNA 3D structure." Nucleic Acids Res **39**(10): 4007-4022.
- Schlunzen, F., J. M. Harms, F. Franceschi, H. A. Hansen, H. Bartels, R. Zarivach i A. Yonath (2003). "Structural basis for the antibiotic activity of ketolides and azalides." Structure **11**(3): 329-338.
- Schlunzen, F., E. Pyetan, P. Fucini, A. Yonath i J. M. Harms (2004). "Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin." Mol Microbiol **54**(5): 1287-1294.
- Schlunzen, F., R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath i F. Franceschi (2001). "Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria." Nature **413**(6858): 814-821.

- Schnabl, J. i R. K. Sigel (2010). "Controlling ribozyme activity by metal ions." Curr Opin Chem Biol **14**(2): 269-275.
- Scott, L. G. i M. Hennig (2008). "RNA structure determination by NMR." Methods Mol Biol **452**: 29-61.
- Scott, W. G., M. Martick i Y. I. Chi (2009). "Structure and function of regulatory RNA elements: ribozymes that regulate gene expression." Biochim Biophys Acta **1789**(9-10): 634-641.
- Serra, M. J., J. D. Baird, T. Dale, B. L. Fey, K. Retatagos i E. Westhof (2002). "Effects of magnesium ions on the stabilization of RNA oligomers of defined structures." Rna **8**(3): 307-323.
- Shan, S., A. V. Kravchuk, J. A. Piccirilli i D. Herschlag (2001). "Defining the catalytic metal ion interactions in the Tetrahymena ribozyme reaction." Biochemistry **40**(17): 5161-5171.
- Shen, L. X., Z. Cai i I. Tinoco, Jr. (1995). "RNA structure at high resolution." Faseb J **9**(11): 1023-1033.
- Sioud, M. (2006). "Ribozymes and siRnas: from structure to preclinical applications." Handb Exp Pharmacol(173): 223-242.
- Sippl, M. (1993). "Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures." J Comput Aided Mol Des **7**: 473-501.
- Sippl, M. J. (1995). "Knowledge-based potentials for proteins." Curr Opin Struct Biol **5**(2): 229-235.
- Soukup, G. A. i R. R. Breaker (1999). "Design of allosteric hammerhead ribozymes activated by ligand-induced structure stabilization." Structure **7**(7): 783-791.
- Stein, A. i D. M. Crothers (1976). "Equilibrium binding of magnesium(II) by Escherichia coli tRNA^{fMet}." Biochemistry **15**(1): 157-160.
- Ster, C., M. Allard, S. Boulanger, M. Lamontagne Boulet, J. Mulhbach, D. A. Lafontaine, E. Marsault, P. Lacasse i F. Malouin (2013). "Experimental treatment of Staphylococcus aureus bovine intramammary infection using a guanine riboswitch ligand analog." J Dairy Sci **96**(2): 1000-1008.
- Sudarsan, N., J. K. Wickiser, S. Nakamura, M. S. Ebert i R. R. Breaker (2003). "An mRNA structure in bacteria that controls gene expression by binding lysine." Genes Dev **17**(21): 2688-2697.
- Tan, Z. J. i S. J. Chen (2005). "Electrostatic correlations and fluctuations for ion binding to a finite length polyelectrolyte." J. Chem. Phys. **122**.
- Tan, Z. J. i S. J. Chen (2010). "Predicting ion binding properties for RNA tertiary structures." Biophys J **99**(5): 1565-1576.

- Thomas, J. R. i P. J. Hergenrother (2008). "Targeting RNA with small molecules." Chem Rev **108**(4): 1171-1224.
- Tinoco, I., Jr. i C. Bustamante (1999). "How RNA folds." J Mol Biol **293**(2): 271-281.
- Tjioe, E., K. Lasker, B. Webb, H. J. Wolfson i A. Sali (2011). "MultiFit: a web server for fitting multiple protein structures into their electron microscopy density map." Nucleic Acids Res **39**(Web Server issue): W167-170.
- Trausch, J. J., P. Ceres, F. E. Reyes i R. T. Batey (2011). "The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer." Structure **19**(10): 1413-1423.
- Travers, K. J., N. Boyd i D. Herschlag (2007). "Low specificity of metal ion binding in the metal ion core of a folded RNA." Rna **13**(8): 1205-1213.
- Tu, D., G. Blaha, P. B. Moore i T. A. Steitz (2005). "Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance." Cell **121**(2): 257-270.
- Urbanke, C., R. Romer i G. Maass (1975). "Tertiary structure of tRNAPhe (yeast): kinetics and electrostatic repulsion." Eur J Biochem **55**(2): 439-444.
- Vicens, Q. i E. Westhof (2001). "Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site." Structure (Camb) **9**(8): 647-658.
- Vicens, Q. i E. Westhof (2002). "Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding a site." Chem Biol **9**(6): 747-755.
- Vicens, Q. i E. Westhof (2003). "Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide." J Mol Biol **326**(4): 1175-1188.
- Walter, N. G. i J. M. Burke (1998). "The hairpin ribozyme: structure, assembly and catalysis." Curr Opin Chem Biol **2**(1): 24-30.
- Warner, D. F., S. Savvi, V. Mizrahi i S. S. Dawes (2007). "A riboswitch regulates expression of the coenzyme B12-independent methionine synthase in Mycobacterium tuberculosis: implications for differential methionine synthase function in strains H37Rv and CDC1551." J Bacteriol **189**(9): 3655-3659.
- Westhof, E. i P. Auffinger (2006). RNA Tertiary Structure. Encyclopedia of Analytical Chemistry, John Wiley & Sons, Ltd.
- Wheeler, T. M., K. Sobczak, J. D. Lueck, R. J. Osborne, X. Lin, R. T. Dirksen i C. A. Thornton (2009). "Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA." Science **325**(5938): 336-339.
- Wilson, D. N., J. M. Harms, K. H. Nierhaus, F. Schlunzen i P. Fucini (2005). "Species-specific antibiotic-ribosome interactions: implications for drug development." Biol Chem **386**(12): 1239-1252.

Winkler, W. C., S. Cohen-Chalamish i R. R. Breaker (2002). "An mRNA structure that controls gene expression by binding FMN." Proc Natl Acad Sci U S A **99**(25): 15908-15913.

Winkler, W. C., A. Nahvi, N. Sudarsan, J. E. Barrick i R. R. Breaker (2003). "An mRNA structure that controls gene expression by binding S-adenosylmethionine." Nat Struct Biol **10**(9): 701-707.

Woodson, S. A. (2005). "Metal ions and RNA folding: a highly charged topic with a dynamic future." Curr Opin Chem Biol **9**(2): 104-109.

Wriggers, W. (2010). "Using Situs for the integration of multi-resolution structures." Biophys Rev **2**(1): 21-27.

Zhou, J., D. A. Sayre, J. Wang, N. Pahadi i H. O. Sintim (2012). "Endo-S-c-di-GMP analogues-polymorphism and binding studies with class I riboswitch." Molecules **17**(11): 13376-13389.