

Instytut Chemii Bioorganicznej  
Polska Akademia Nauk  
Zakład Bioinformatyki Strukturalnej

**Analiza strukturalnych uwarunkowań  
wybranych etapów biogenezy małych  
regulatorowych RNA u *Arabidopsis thaliana***

Structural determinants of selected stages of small  
regulatory RNA biogenesis in *Arabidopsis thaliana*

mgr Agnieszka Mickiewicz-Piotrowska

Promotor  
prof. dr hab. Inż. Marta Szachniuk

Promotor pomocniczy  
dr Joanna Sarzyńska

Poznań 2021

*W trakcie realizacji pracy doktorskiej, autorka otrzymywała  
stypendium naukowe:*

*„Stypendium Prezesa PAN dla najlepszych doktorantów”*

*Dziękuję panu  
prof. dr hab. Markowi Figlerowiczowi  
za umożliwienie mi badań w IBCH PAN,  
kierownictwo naukowe,  
cenne wskazówki i życzliwość;*

*Dziękuję mojej promotor  
Pani prof. dr hab. inż. Marcie Szachniuk  
za zaangażowanie w finalizację doktoratu  
oraz pomoc w redakcji pracy;*

*Dziękuję mojej promotor  
Pani dr Joannie Sarzyńskiej  
za opiekę naukową,  
zaangażowanie i wszelką pomoc;*

*Składam serdeczne podziękowania  
za owocną współpracę w miłej atmosferze:  
Panu prof. dr hab. inż. Jackowi Błażewiczowi,  
Panu dr inż. Maciejowi Miłostanowi,  
oraz Pani dr hab. inż. Agnieszce Rybarczyk;*

*Dziękuję także moim Rodzicom oraz Mężowi,  
za nieustające wsparcie, nieocenioną cierpliwość  
oraz wiarę w powodzenie mojego doktoratu;*

## Spis treści

1. Wstęp.....	11
1.1. Wprowadzenie.....	11
1.1.1. Odkrycie zjawiska RNAi.....	12
1.1.2. Mechanizm działania srRNA .....	13
1.2. Małe regulatorowe RNA u roślin .....	14
1.2.1. miRNA .....	14
1.2.2. siRNA.....	17
1.2.2.1. ta-siRNA .....	17
1.2.2.2. ha-siRNA .....	18
1.2.2.3. nat-siRNA.....	19
1.2.3. Egzogenne srRNA.....	19
1.2.4. amiRNA .....	20
1.3. Biogeneza srRNA – schemat .....	21
1.4. Białka uczestniczące w biogenezie roślinnych srRNA .....	22
1.4.1. Rodzina białek RNazy III .....	22
1.4.2. Budowa białek typu Dicer .....	23
1.4.3. Funkcja białek typu Dicer.....	26
1.4.4. Dodatkowe białka uczestniczące w biogenezie srRNA.....	27
1.4.4.1. Białka typu AGO.....	27
1.4.4.2. Polimeraza RNA zależna od RNA .....	28
1.4.4.3. Inne białka pomocnicze .....	29
2. Cel pracy.....	30
3. Materiały i metody.....	31
3.1. Materiały.....	31
3.1.1. Sekwencje RNA .....	31
3.1.2. Sekwencje białek .....	32
3.1.3. Struktury białek.....	33
3.1.4. Programy komputerowe wykorzystane w badaniach .....	34
3.2. Metody.....	36
3.2.1. Drugorzędowa struktura RNA.....	36
3.2.2. Dopasowanie sekwencji aminokwasowych.....	37
3.2.3. Modelowanie struktur białek .....	38
3.2.4. Ocena modeli białek .....	39
3.2.5. Dynamika molekularna .....	39



3.2.6. Przygotowanie szkieletu amiRNA.....	41
4. Wyniki i Dyskusja.....	43
4.1. AmiRNA Designer.....	43
4.1.1. Analiza sekwencji miRNA.....	46
4.1.2. Struktura drugorzędowa duplexu miRNA:target .....	47
4.1.3. Struktura drugorzędowa prekursora miRNA .....	50
4.1.4. Opis działania programu.....	52
4.1.4.1. Etap I – Analiza danych.....	52
4.1.4.2. Etap II – Projektowanie amiRNA.....	57
4.1.4.3. Weryfikacja wyników.....	61
4.1.5. Sposób obliczania profili termodynamicznych. ....	62
4.1.6. Analiza przykładowych wyników .....	65
4.1.7. Szkielet sztucznego miRNA .....	73
4.1.8. Dyskusja .....	77
4.2. Modelowanie AtDCL4 .....	79
4.2.1. Analiza sekwencji białek Dicer i DCL.....	80
4.2.2. Wyszukiwanie szablonów strukturalnych.....	83
4.2.3. Modelowanie homologiczne białka.....	90
4.2.4. Modelowanie kompleksu białko – RNA.....	91
4.2.5. Oddziaływanie białka z RNA.....	98
4.2.6. Dyskusja .....	106
5. Podsumowanie.....	109
6. Streszczenie.....	111
7. Abstract .....	114
8. Bibliografia .....	116
9. Załączniki .....	133

## Wykaz stosowanych skrótów

**Å** – angstrom, jednostka długości równa  $10^{-10}$  m

**aa** – aminokwas (ang. aminoacid)

**AGO** – białko Argonaute (ang. Argonaute)

**BLAST** – narzędzie do poszukiwania lokalnego podobieństwa sekwencji (ang. basic local alignment search tool)

**cDNA** – DNA komplementarny do mRNA (ang. complementary DNA)

**DCL** – białko typu Dicer (ang. Dicer - like protein)

**DRB** – białko wiążące dsRNA (ang. dsRNA binding protein)

**dsRBD** – domena wiążąca dsRNA (ang. dsRNA binding domain)

**dsRNA** – dwuniciowy RNA (ang. double - stranded RNA)

**DUF** – domena o nieznannej funkcji (ang. domain of unknown function)

**hc-siRNA** – heterochromatynowe małe interferujące RNA (ang. heterochromatic siRNA)

**lncRNA** – długi niekodujący RNA (ang. long non - coding RNA)

**MD** – Dynamika molekularna (ang. Molecular Dynamic)

**MSA** – przyrównanie wielu sekwencji (ang. Multiple Sequence Alignment)

**mRNA** – RNA informacyjny (ang. messenger RNA)

**miRNA** – mikroRNA (ang. microRNA)

**nat-siRNA** – małe interferujące RNA pochodzący z naturalnych transkryptów antysensownych (ang. natural antisense transcript siRNA)

**ncRNA** – niekodujące RNA (ang. non - coding RNA)

**NMA** – analiza drgań normalnych (Normal Mode Analysis)

**ns** – nanosekunda

**nt** – nukleotyd (ang. nucleotide)

**PAZ** – domena Piwi/Argonaute/Zwill (ang. Piwi/Argonaute/Zwill)

**PCA** – analiza głównych składowych (PCA, ang. Principal Component Analysis)

**PCR** – łańcuchowa reakcja polimerazy (ang. polymerase chain reaction)

**Pol** – polimeraza RNA (ang. RNA polymerase)

**pre- miRNA** – prekursor miRNA (ang. precursor miRNA)

**pri-miRNA** – pierwotny prekursor miRNA (ang. primary miRNA)

**PTGS** – potranskrypcyjne wyciszenie genów (ang. post - transcriptional gene silencing)

**RdDM** – metylacja DNA kierowana przez RNA (ang. RNA - directed DNA methylation)

**RDR** – polimeraza RNA zależna od RNA (ang. RNA - dependent RNA polymerase)

**RISC** – kompleks wyciszający indukowany przez RNA (ang. RNA induced silencing complex)

**RMSD** – średnie odchylenie kwadratowe (ang. Root Mean Square Deviation)

**RNAi** – interferencja RNA (ang. RNA interference)

**siRNA** – małe interferujące RNA (ang. small interfering RNA)

**sncRNA** – małe niekodujące RNA (ang. small non - coding RNA)

**srRNA** – małe regulatorowe RNA (ang. short - regulatory RNA)

**ssRNA** – jednoniciowe RNA (ang. single - stranded RNA)

**ta-siRNA** – małe interferujące RNA działający w trans (ang. trans - acting siRNA)

**TGS** – transkrypcyjne wyciszenie genów (ang. transcriptional gene silencing)

**vsRNA** – małe interferujące RNA pochodzenia wirusowego (ang. virus - derived siRNA)

## 1. Wstęp

### 1.1. Wprowadzenie

Harmonijny rozwój każdego żywego organizmu jest procesem niezwykle złożonym, wymagającym ekspresji wielu genów. Procesy w komórkach muszą zachodzić równocześnie i bardzo szybko – wymaga to precyzyjnej synchronizacji ekspresji genów, zarówno czasowej jak i ilościowej. Ekspresja genów regulowana jest za pomocą wielu różnych białkowych czynników transkrypcyjnych, przyłączających się do regionu promotorowego genu (Matsui 1980). Dzięki nim geny ulegają transkrypcji, a następnie translacji co prowadzi do powstania określonych białek, niezbędnych do prawidłowego funkcjonowania komórki na danym etapie rozwoju. Niemalże przed 30 laty odkryto inny mechanizm regulacji ekspresji genów określony jako interferencja RNA (ang. RNA interference, RNAi), lub wyciszanie genów (ang. gene silencing) (Napoli 1990, Fire 1998). Polega on głównie na modyfikacji poziomu określonego transkryptu (mRNA) przez hybrydującą do niego cząsteczkę RNA. Niewielkie rozmiary tychże RNA (do 25 nt) oraz funkcja pozwalają określać je wspólnym mianem „małych regulatorowych RNA” (ang. short regulatory RNA, srRNA). Przyłączenie srRNA do mRNA może powodować zablokowanie procesu translacji białka lub też uruchomić szlak degradacji mRNA – w obu przypadkach efektem jest obniżenie poziomu produkcji określonego białka. Regulacja ekspresji genów za pomocą czynników białkowych wymaga czasochłonnej biosyntezy białek regulatorowych. Ponadto białkowe motywy wiążące DNA rozpoznają krótkie sekwencje w regionie promotorowym genu, które nie są unikalne. Regulacja genów na drodze RNAi działa bardzo szybko, gdyż nie wymaga translacji, a także wykazuje ogromną specyficzność wynikającą z komplementarności pojedynczej cząsteczki srRNA do mRNA o długości ponad 20 nt. Niemal wszystkie organizmy wykorzystują do regulacji ekspresji endogenów mechanizm oparty na RNAi. Dodatkowo u roślin RNAi stanowi podstawę obrony przed infekującymi je patogenami. Niniejszy rozdział stanowi krótki przegląd literatury dotyczącej zjawiska potranskrypcyjnego wyciszania genów, w szczególności zaangażowanych w ten proces małych regulatorowych RNA oraz ich biogenezy.

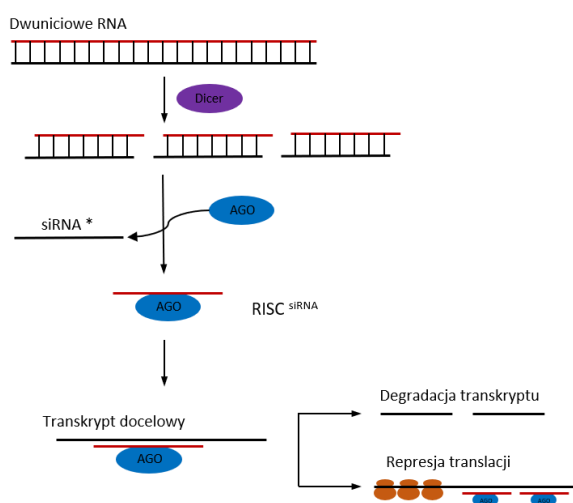
### 1.1.1. Odkrycie zjawiska RNAi

Już w latach dwudziestych ubiegłego stulecia zaobserwowano, że tytoń zainfekowany wirusem pierścieniowej plamistości (ang. tobacco ringspot virus, TobRSV) wykazuje objawy choroby jedynie w liściach, które miały bezpośredni kontakt z wirusem, natomiast górne, młodsze liście pozostają zdrowe, a co więcej – odporne na powtórne zakażenie (Wingard 1928). Zjawisko to – zaobserwowane w przypadku bardzo wielu wirusów i roślin przez nie infekowanych – opisano jako odporność pochodzącą od patogenu (ang. pathogen-derived resistance, PDR) (Lindbo 1992). Początkowo sądzono, że PDR występuje na poziomie białka, jednakże wprowadzenie do genomu rośliny jedynie fragmentu genu wirusowego (co nie prowadziło do powstania pełnego, poprawnie zwinętego białka) również skutecznie przeciwdziało infekcji. Wywnioskowano zatem, że istnieje mechanizm rozpoznawania i usuwania wirusowego RNA przez komórkę roślinną, a proces PDR nazwano RNA-zależną odpornością przeciwwirusową (Covey 1997, Ratcliff 1997). Indukowane infekcją wirusową wyciszenie genu (ang. Virus Induced Gene Silencing, VIGS), występuje dość powszechnie pomiędzy licznymi gatunkami roślin a infekującymi je wirusami (Poogin 2017, Prasad 2019).

W podobnym czasie odkryto u roślin zjawisko PTGS (ang. Post Transcriptional Gene Silencing). W celu wzmocnienia intensywności koloru petunii wprowadzono do jej genomu dodatkową kopię genu syntazy chalkonowej. Zaskakującym efektem było uzyskanie roślin o znacznie słabszym kolorze, ponieważ dodatkowa kopia genu obniżyła ilość białka powstającego z tego genu, zamiast ją podwyższyć. Zjawisko to określono jako współblokowanie (ang. co-suppression) (Napoli, 1990). Podobny proces – nazwany tłumieniem (ang. quelling) – zaobserwowano dla genu syntazy karotenoidów u grzyba *Neurospora crassa* (Cogoni, 1999). Eksperymenty różnych grup badawczych doprowadziły do odkrycia podobnego zjawiska u wielu organizmów, między innymi: muszki owocówki (Jensen 1999), pierwotniaka *Paramecium* (Ruiz 1998), nicienia *Caenorhabditis elegans* (Tabara 1999), oraz ssaków w tym człowieka (Bahramian 1999) gdzie nazwano go interferencją RNA – RNAi. Procesy te okazały się na tyle zbliżone w mechanizmie, że obecnie u wszystkich organizmów określa się je wspólną nazwą wyciszania (lub interferencji) RNA (ang. RNA silencing). W 2006 roku Andrew Fire i Craig Mello za badania nad procesem RNAi u *Caenorhabditis elegans* otrzymali nagrodę Nobla w dziedzinie medycyny i fizjologii.

### 1.1.2. Mechanizm działania srRNA

Czynnikiem uruchamiającym wyciszanie genów na drodze RNAi są małe regulatorowe RNA. Pochodzą one z prekursorów posiadających w swojej strukturze fragmenty dwuniciowe, z których jedno z białek typu Dicer wycina dupлексы o długości 20-24 nt. Krótkie fragmenty dsRNA zostają następnie przyłączone do białka typu Argonaute (AGO), jednego ze składników kompleksu białkowego RISC (ang. RNA-Induced Silencing Complex). W tym kompleksie – zwanym także Mikroprocesorem – dupлекс rozplata się. Jedna z jego nici (ang. guide) pozostaje w kompleksie przyłączona do białka AGO, jako wzorzec rozpoznania mRNA, druga natomiast (ang. passenger) – odrzucona. Następnie RISC przyłącza się do mRNA i w ścieżce prowadzącej do degradacji mRNA jedno z białek AGO (należące do klasy RNaz typu V – nazywane również „Slicer”) przecina mRNA w miejscu hybrydyzacji srRNA. Cięcie jest bardzo specyficzne, gdyż zachodzi dokładnie w połowie długości srRNA, pomiędzy 10 a 11 nukleotydem (Elbashir 2001). Następnie RISC oddysocjowuje i może przyłączyć się do kolejnej cząsteczki mRNA. Inną ścieżką, którą może zająć wyciszanie jest represja translacji. Nie dochodzi wówczas do rozcięcia mRNA, a jedynie do zablokowania dalszego procesu biogenezy białka. Związane z transkrypcyjnym srRNA blokuje łączenie się podjednostek rybosomu oraz przyłączenie kolejnych czynników inicjacji translacji. Jeszcze niedawno sądzono, że poziom komplementarności warunkuje mechanizm indukowania sposobu represji genów przez srRNA. Proces ten przedstawiono na Rys. 1.1



**Rys. 1.1.** Zmodyfikowany ogólny schemat działania srRNA w procesie potranskrypcyjnego wyciszania genów (Silva 2015).

Sugeruje się, że srRNA o wysokim stopniu komplementarności działają poprzez degradację ich transkryptów docelowych natomiast te o niższym stopniu komplementarności poprzez represję translacji (Shabalina 2008). Chociaż jest to prawdą w przypadku modeli zwierzęcych, nie ma wyraźnych dowodów na taką samą korelację u roślin (Brodersen 2009), a mechanizm wyboru ścieżki wyciszania nadal pozostaje niejasny. Oprócz powyższych mechanizmów prowadzących do potranskrypcyjnego wyciszania genów, małe regulatorowe RNA mogą działać na wcześniejszym etapie i powodować transkrypcyjne wyciszenie (ang. transcriptional gene silencing, TGS) (Xie 2017, Bai 2018). Powstające w komórce srRNA wraz z białkami AGO kierują procesem metylacji DNA zależnej od RNA. U roślin dotyczy to zwłaszcza transpozonów i sekwencji repetytywnych.

## 1.2. Małe regulatorowe RNA u roślin

Małe regulatorowe RNA, to jednoniciowe cząsteczki RNA o wielkości od 20 do 24 nukleotydów. Endogenne pojawiają się w komórkach w efekcie ekspresji kodujących je endogenów, natomiast egzogenne – w odpowiedzi na infekcje patogenami lub na dwuniciowe RNA wprowadzone metodami inżynierii genetycznej (Pagliarani 2019). Endogenne małe RNA można podzielić na kilka grup w oparciu o różnice w biogenezie i funkcji. Ich ścieżki biogenezy determinowane są głównie lokalizacją komórkową oraz pochodzeniem i strukturą prekursorów z których powstają cząsteczki efektorowe. Powstające w jądrze komórkowym miRNA pochodzą z jednoniciowych prekursorów o strukturze spinki do włosów, natomiast cytoplazmatyczne siRNA – z długich dwuniciowych RNA (dsRNA). Znane obecnie endogenne siRNA należą w większości do jednej z trzech grup: (i) siRNA działające w „trans” - ta-siRNA; (ii) siRNA związane z heterochromatyną – ha-siRNA oraz (iii) naturalne siRNA – nat-siRNA (Borges 2015). Można także wyróżnić egzogenne siRNA oraz „sztuczne” siRNA. Zostaną one pokrótce opisane w kolejnych podrozdziałach.

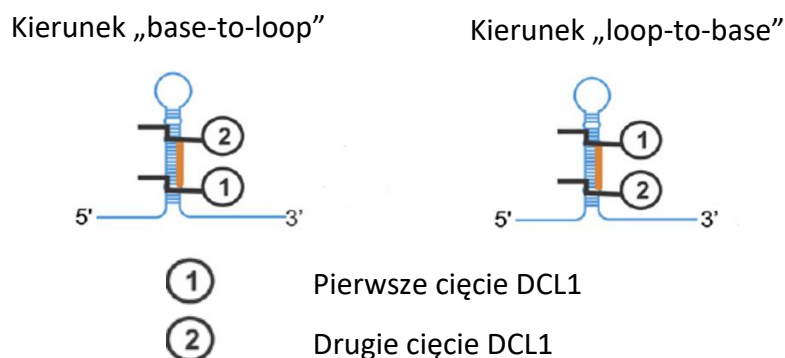
### 1.2.1. miRNA

MikroRNA (miRNA) są powszechne zarówno u roślin jak i zwierząt. Pochodzą z regionów międzygenowych lub intronów. Powstają w postaci prekursorów pierwotnych (pri-miRNA) w procesie transkrypcji prowadzonej przez polimerazę RNA II (Chen 2009) a następnie

ulegają modyfikacjom potranskrypcyjnym obejmującym przyłączenie czapeczki 7-metyloguanozynowej na końcu 5', poliadenylację końca 3'. Mogą także przechodzić proces składania (Xie 2005) lub alternatywnego składania (Szarzyńska 2009). Pri-miRNA przyjmują często skomplikowaną organizację struktury drugorzędowej, zawsze jednak zawierają podstrukturę tzw. spinki do włosów (ang. hairpin structure), złożoną z dwuniciowego trzonu i jednoniciowej pętli. We fragmentach helikalnych (trzon) zawarte są ok. 21 nukleotydy (nt) miRNA. Mogą one występować w 5' lub 3' ramieniu trzonu, zdarzają się także występujące w obu ramionach jednego prekursora, lub też dwa miRNA leżące na tym samym ramieniu. Pomimo uniwersalności procesu RNAi, u roślin i zwierząt istnieją różnice, zarówno w biogenezie miRNA jak i samej strukturze prekursorów (Bologna 2013). Prekursory zwierzęce miRNA są ok. 80 nukleotydowej długości i dwa różne białka – Drosha i Dicer – katalizują kolejne reakcje hydrolizy. Białko Drosha w pierwszym etapie, zachodzącym w jądrze komórkowym, uwalnia pre-miRNA (Han 2006), z którego następnie w cytoplazmie Dicer uwalnia dupleks miRNA/miRNA\*. Roślinne prekursory miRNA są często znacznie dłuższe, do ponad 600 nt, wykazują duże zróżnicowanie strukturalne i są rozcinane w jądrze komórkowym przez białko typu Dicer 1 (ang. Dicer-like protein1, DCL1, Kuhara 2004) według kilku mechanizmów, w zależności od kierunku obróbki i liczby cięć, ostatecznie generując ok. 21-nt dupleks miRNA/miRNA\*. O odmienności systemów roślinnego i zwierzęcego świadczy także fakt, iż siRNA ulegające ekspresji w komórkach roślinnych nie powodują wyciszenia w komórkach zwierzęcych (Chau 2007)

Niedawne badania biochemiczne pokazały, że sposoby wycinania dupleksu miRNA/miRNA\* mogą być zarówno ewolucyjnie zachowane jak i stosunkowo niedawno utworzone. Zaowocowało to powstaniem nowych koncepcji przebiegu procesu wycinania dojrzałego miRNA z jego prekursora. Mechanizm definiowany jest kolejnością wykonywanych cięć: pierwsze od strony trzonu (ang. base-to-loop) lub pierwsze od strony pętli (ang. loop-to-base) (Rys. 1.2). Kolejny element określający mechanizm powstawania miRNA stanowi liczba cięć niezbędna do osiągnięcia jednego z końców miRNA: jedno cięcie to mechanizm „short” (zachowane ewolucyjnie), więcej niż jedno – „long” (ewolucyjnie nowsze) (Bologna 2013). Badania *in silico* drugorzędowych struktur wielu prekursorów miRNA, wskazują możliwość występowania motywów sekwencyjnych i strukturalnych w rejonach otaczających dupleks miRNA:miRNA\*. Przypuszcza się, że motywy te mogą być rozpoznawane przez Mikroprocesor

(Miśkiewicz 2017). Wspiera to hipotezę o dwóch mechanizmach w zależności od położenia pierwszego cięcia oraz o ewolucyjnym zachowaniu mechanizmu „short”.



**Rys. 1.2.** Mechanizm wycinania miRNA z ich prekursorów u roślin.

W 2010 roku ukazały się prace pokazujące zależność pomiędzy strukturami drugorzędowymi niektórych roślinnych pri-miRNA poza regionem definiującym miRNA:miRNA\* a działaniem dojrzałego miRNA. Natywne, zawierające zaburzenia struktury, prekursor powodował znacznie większy efekt wyciszający miRNA niż prekursor takiego samego miRNA o całkowicie sparowanych regionach flankujących miRNA:miRNA\* (Song 2010, Mateos 2010). Potwierdza to przypuszczenia o wpływie budowy prekursora na sposób w jaki DCL1 rozpoznaje miejsca cięcia wpływając na proces powstawania dojrzałych miRNA, a co za tym idzie – ich funkcjonalność.

Mimo, że RNAi jest uniwersalnym procesem, istnieją różnice w sposobie jego funkcjonowania u zwierząt i roślin. MiRNA roślin zazwyczaj są idealnie lub prawie idealnie komplementarne do swoich miejsc docelowych (Axtell 2011). MiRNA zwierzęce wykazują znacznie mniejszą komplementarność. Wyróżnia się w nich tzw. region "seed": najczęściej jest to fragment od drugiego do ósmego nukleotydu z 5' końca miRNA komplementarny do ok. 7 nt miejsca docelowego. Na obu niciach może wystąpić dodatkowy, niesparowany nukleotyd (Bartel, 2009). Inną różnicą jest położenie miejsc docelowych miRNA. U zwierząt znajdują się one głównie w regionach 3' nie ulegających translacji (3' UTR), natomiast u roślin miRNA wiążą się zarówno do miejsc kodujących i regionów UTR. U roślin miRNA regulują głównie wzrost i procesy rozwojowe (Mallory 2006, Liu 2018a), a także są zaangażowane w odpowiedzi na stres biotyczny i abiotyczny (praca przeglądowa Sunkar 2012, Ahmed 2019, Jagannadham 2019). U zwierząt, miRNA są zaangażowane w prawidłowe funkcjonowanie każdej komórki



(Chekulaeva 2009, Gebert 2019), a ich zaburzenia są związane z różnymi chorobami (Meola 2009, Martínez-Jiménez 2019).

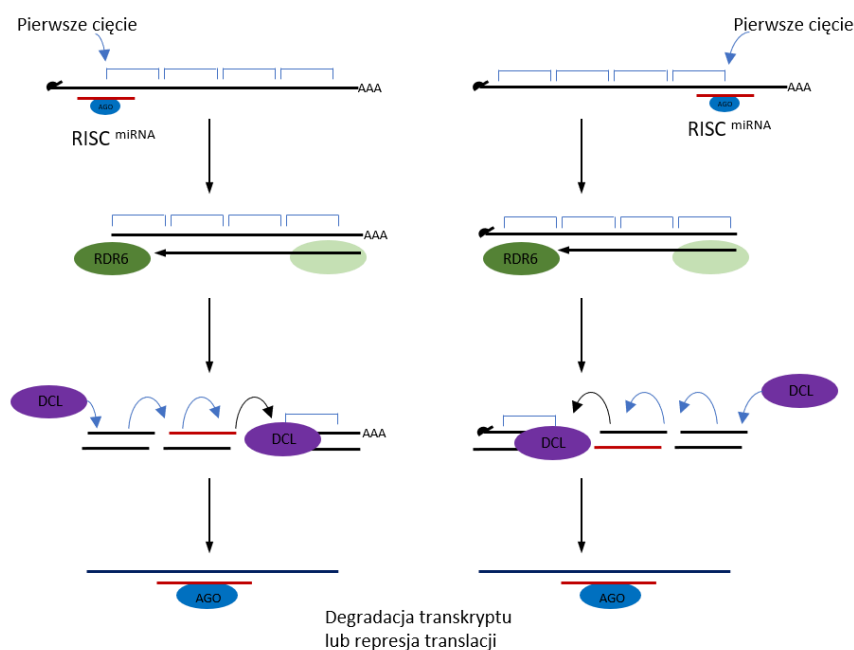
### 1.2.2. siRNA

Drugą ważną klasą cząsteczek małych regulatorowych RNA, które uruchamiają mechanizm RNAi u roślin oraz niektórych zwierząt są siRNA. Pochodzą one z infekujących wirusów, transgenów, odwróconych powtórzeń, centromerów, transpozonów lub sekwencji repetytywnych, a ich funkcją jest zapobieganie ekspresji tych sekwencji. Zasadniczo ich komplementarność do wyciskanego docelowego transkryptu jest pełna. Ich biogeneza nie obejmuje zatem wyszukiwania zaburzeń strukturalnych jak w przypadku miRNA. Długie dwuniciowe RNA (dsRNA) jest cięte – rozpoczynając od końca o łatwiejszej dostępności – kolejno na krótkie (21-24 nt) fragmenty przez Dicer2 lub Dicer1 w przypadku zwierząt oraz DCL2-4 u roślin. siRNA uważa się za oparty na RNA mechanizm immunologiczny do usuwania szkodliwych, czy zbędnych cząsteczek kwasów nukleinowych, lub utrzymywania stabilności genomu (Deng 2018, Axtell 2018).

#### 1.2.2.1. ta-siRNA

siRNA działające w trans wyciszają geny, które są odmienne od *loci* ich pochodzenia, stąd ich nazwa ta-siRNA (ang. trans active siRNA). Są one zakodowane w genomie jako tzw. geny TAS, zaś ich oddziaływanie obejmuje regulację innego genu, niż ten, z którego pochodzą – działanie „*in trans*”. Pierwszorzędowe srRNA (siRNA lub miRNA) inicjują produkcję dwuniciowego RNA, z którego są wycinane drugorzędowe srRNA, stanowiące ta-siRNA. Większość roślinnych ta-siRNA ma długość 21-nt, a ich biogeneza wymaga współdziałania białek RDR6 i DCL4 (Allen 2005, Yoshikawa 2005). Niekodujące transkrypty genów TAS są rozpoznawane przez kompleks AGO-miRNA i rozcinane. Do jednego z fragmentów transkryptu (5' lub 3' – mechanizm nie jest znany) polimeraza RNA zależna od RNA 6 (RDR6) syntetyzuje komplementarną nić. Powstały w ten sposób dsRNA jest cięty przez białko DCL4 na następujące kolejno po sobie fragmenty, stanowiące drugorzędowe siRNA. Początek pierwszego z nich jest zdefiniowany przez pierwsze cięcie białka AGO (Rys. 1.3). Ta-siRNA powstające kolejno z jednego transkryptu mogą regulować ekspresję odrębnych genów.

Udział komórkowej polimerazy RDR6 definiuje substrat dla DCL4 jako idealnie sparowany dwuniciowy RNA.



**Rys. 1.3.** Schemat powstawania ta-siRNA miRNA (Allen 2005) zmodyfikowany. Pierwszym etapem ich powstawania jest kierowane przez miRNA przecięcie transkryptu. Następnie polimeraza RNA RDR6 syntetyzuje komplementarną do przeciętego transkryptu drugą nić RNA i taki dwuniciowy RNA jest sekwencyjnie cięty przez DCL4 począwszy od końca, który powstał po pierwotnym przecięciu transkryptu. Niektóre fragmenty działają jako siRNA, kierując białko typu AGO do rozcięcia transkryptu odmiennego niż ten, z którego pochodzą ta-siRNA (lub represji translacji).

### 1.2.2.2. ha-siRNA

Związane z heterochromatyną siRNA (heterochromatin associated siRNA, ha-siRNA) pochodzą z międzygenowych lub repetytywnych regionów genomu. Mają one zazwyczaj 24 nt długości, a ich biogeneza wymaga białek RDR2, DCL3 i AGO4, AGO6 i AGO9 u *Arabidopsis* (Axtell 2013). Biogeneza ha-siRNA nie jest w pełni wyjaśniona. Aktualnie uważa się, że powtarzalne regiony międzygenowe są transkrybowane przez RNA polimerazę IV (Pol IV) specyficzną dla roślin. Transkrypt Pol IV służy jako matryca dla RDR2 w celu wytworzenia dsRNA. DCL3 rozcina otrzymane dsRNA na 24-nt siRNA i jedna z jego nici jest włączana do AGO4. Polimeraza RNA V (Pol V), także specyficzna dla roślin, wytwarza transkrypty służące jako matryca dla wiązania siRNA i białka AGO4. Po związaniu do transkryptów Pol V, ha-siRNA naprowadzają metylotransferazę – enzym powodujący modyfikację *de novo* chromatyny w *loci*, z których pochodzą transkrypty generowane przez Pol V, w tym metylację

cytozyny w asymetrycznych miejscach CHH (gdzie H = C lub T) i metylację histonów H3K9 (Zhong 2012, Wierzbicki 2012).

### 1.2.2.3. nat-siRNA

Uważa się, że naturalne siRNA (nat-siRNA) pochodzą z dwuniciowego RNA, utworzonego przez hybrydyzację niezależnych transkryptów komplementarnych RNA. Dwie komplementarne nici RNA zostały przepisane z przeciwnych stron w tym samym *locus*, a po rozcięciu na małe fragmenty stanowią nat-siRNA działające w *cis* (Borsani 2005, Zhang 2012, Yuan 2015). Możliwa jest także hybrydyzacja RNA pochodzących z innych *loci* prowadząca do powstania trans-nat-siRNA (Yuan 2015).

### 1.2.3. Egzogenne srRNA

Małe cząsteczki RNA pochodzące z replikacji RNA wirusa obserwowano u roślin zainfekowanych wirusami jeszcze przed odkryciem wyciszania na drodze RNAi związanego z endogennymi srRNA (Hamilton 1999, Ding 2007). Ponieważ srRNA były także związane z potranskrypcyjnym wyciszaniem genów (PTGS) indukowanym obecnością transgeny, zaproponowano hipotezę, że PTGS może stanowić naturalny mechanizm obrony przeciwvirusowej roślin (Hamilton 1999), przy czym molekularne podstawy tego procesu nie były wówczas jasne. Odkrycie endogennych małych RNA w różnych organizmach eukariotycznych, oraz rozległe badania genetyczne i biochemiczne, pomogły ustalić mechanizm wyciszania genów także przez egzogenne srRNA (Zamore 2005; Vaucheret 2006). Strategia RNAi funkcjonuje przede wszystkim jako odpowiedź roślin na infekcje wirusowe, jednakże pokazano, że grzybowe patogeny mogą ją również indukować (Lu 2007). Zarówno egzogenne srRNA pochodzące od wirusów jak i te z transgenów indukujących RNAi nie mają wpływu ani na fenotyp, ani na procesy rozwojowe regulowane przez endogenne srRNA (Shen 2014, Pumplin 2013).

#### 1.2.4. amiRNA

Małe regulatorowe RNA często wykorzystywane są w badaniach funkcjonalnych genów. Pierwotnie stosowano do tego celu długie dwuniciowe RNA. Takie podejście prowadzi do powstania heterogenicznej puli srRNA, co nie gwarantuje wydajnego efektu wyciszenia. Sąsiednie małe regulatorowe RNA pochodzące z tej samej długiej sekwencji mogą wyciszać docelowy gen w różnym stopniu (np.: w 20% i 70%) (Li 2013). Ponadto poszczególne srRNA powstałe w ten sposób mogą oddziaływać niespecyficznie z genami, do których wyciszenia nie były pierwotnie przeznaczone (ang. off-target effect) (Petri 2013), a także wyciszać transkrypt, z którego pochodzą (Molnar 2009).

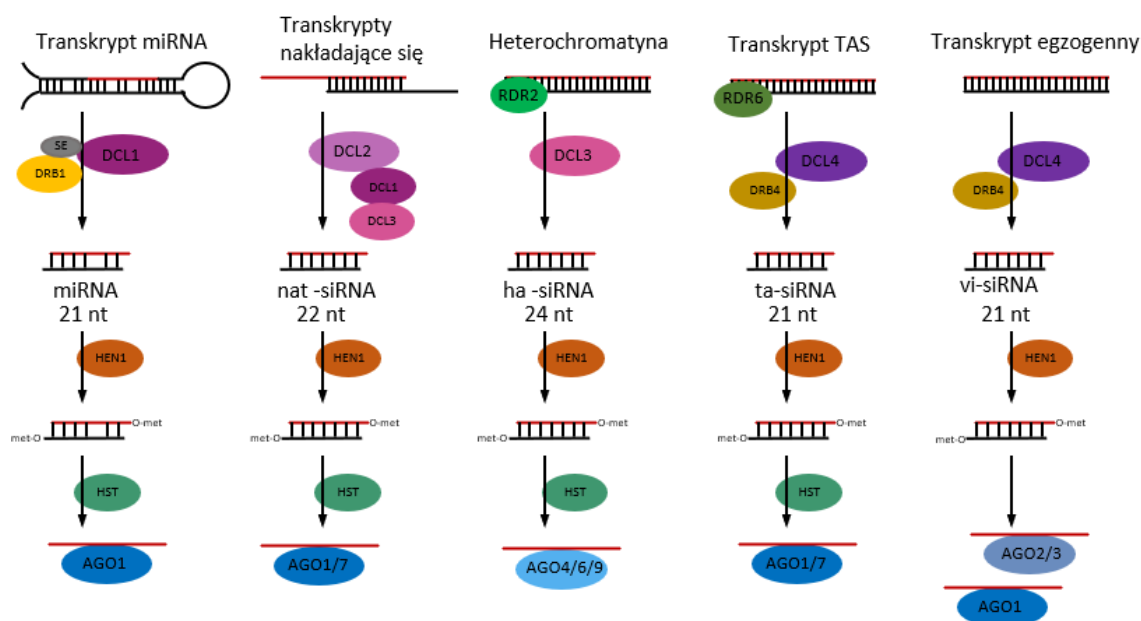
W celu uniknięcia powyższych niedogodności opracowano koncepcję „sztucznych miRNA” – amiRNA. Określenie „sztuczne” miRNA odnosi się do sekwencji samego srRNA nieobecnej pierwotnie u modyfikowanego organizmu a zaprojektowanej w celu wyciszenia konkretnego transkryptu. Stworzenie sztucznego miRNA polega na umieszczeniu sekwencji srRNA w szkielecie funkcjonalnego miRNA w ten sposób, aby natywny proces biogenezy doprowadził do powstania oczekiwanej jednorodnej puli amiRNA. Daje to możliwość analizowania funkcjonalności określonego genu oraz umożliwia regulowanie ekspresji konkretnego genu w celach hodowlanych – np. aby uzyskać odporność na suszę lub zmodyfikować inne cechy rośliny. Dodatkowym atutem tej metody jest możliwość specyficznej ekspresji genu miRNA, będącej odpowiedzią na określony stres biotyczny lub abiotyczny, albo uzyskanie ekspresji tkankowo specyficznej, czy też właściwej dla danego stadium rozwojowego rośliny. Wszystkie te cechy można zapewnić stosując odpowiedni promotor, pod kontrolą którego umieszczony zostanie gen miRNA zawierający sekwencję amiRNA.

Sztuczne miRNA zdaje się być bardzo dobrym narzędziem do badania ekspresji genów zwłaszcza w eksperymentach prowadzonych na organizmach roślinnych. Metoda ta jest o tyle interesująca, że fragmenty kwasów nukleinowych o długości do ok. 30 nt można uzyskać metodami chemicznymi i są one dostępne komercyjnie. Zatem sekwencję amiRNA można dowolnie zaprojektować, a następnie wykorzystując narzędzia biologii molekularnej (jak PCR czy enzymy restrykcyjne) umieścić w szkielecie natywnego miRNA. Jednakże poza techniczną stroną stworzenia amiRNA istotniejszą kwestią wydaje się być wybór właściwej sekwencji,

stanowiącej funkcjonalne małe regulatorowe RNA. Coraz większy postęp informatyzacji zarówno na płaszczyźnie sprzętowej jak i oprogramowania idzie tu w sukurs. Pomimo tego zaprojektowanie sekwencji poprawnie przetwarzanej w procesie biogenezy miRNA i dostarczającej pożądanego efektu wyciszającego jest problemem nietrywialnym.

### 1.3. Biogeneza srRNA – schemat

Małe regulatorowe RNA u roślin biorą udział w procesach rozwojowych, zapewniają stabilność genomu oraz stanowią ochronę przed różnego rodzaju patogenami (Iki 2017). Ich różne rodzaje charakteryzują się odmiennymi szlakami biogenezy. Poniższa ilustracja pokazuje ich zestawienie, a także stanowi wstęp do kolejnego rozdziału, gdzie omówione zostaną białka uczestniczące w powstawaniu srRNA.



**Rys. 1.4.** Szlaki biogenezy srRNA u roślin. Dwuniciowe RNA powstaje w wyniku działania komórkowych polimeraz RNA zależnych od DNA w procesie transkrypcji lub pochodzi od patogenu w przypadku egzogennych transkryptów. Transkrypty pochodzące z heterochromatyny oraz genów TAS wymagają aktywności polimeraz RNA zależnych od RNA w celu utworzenia komplementarnej nici RNA. Białka DCL odpowiadają za wycinanie właściwego fragmentu z dsRNA. Kolejnym etapem jest metylacja 3' końców fragmentów (HEN1) i eksport do cytoplazmy (HST) – jeśli poprzednie etapy zachodziły w jądrze komórkowym.

## 1.4. Białka uczestniczące w biogenezie roślinnych srRNA

Proces biogenezy roślinnych srRNA nie został jeszcze poznany we wszystkich szczegółach. Zasadniczym jego etapem jest wycinanie dojrzałych cząsteczek srRNA z ich prekursorów gdzie kluczową rolę pełnią białka typu Dicer – DCL (ang. Dicer-like proteins) zawierające centrum katalityczne przeprowadzające reakcję hydrolizy wiązania fosfodiesterowego. Stosunkowo najwięcej wiadomo o szlaku związanym z miRNA. W pozostałych natomiast białka DCL2-4 mogą występować zamiennie, a rola tej redundancji nie została jeszcze poznana. Ich sposoby powstawania przedstawiono na Rys. 1.4 zaś funkcje poszczególnych białek zostaną opisane w kolejnych podrozdziałach.

### 1.4.1. Rodzina białek RNazy III

Białka typu Dicer należą do rodziny rybonukleaz typu III (RNazy III), przeprowadzających reakcję hydrolizy dwuniciowego RNA. Rodzina ta jest wszechobecna wśród organizmów żywych. Można podzielić ją na trzy grupy różniące się wielkością i budową, jednakże zachowujące zasadnicze funkcje (Rys. 1.5) (Zhang 2004, Song 2017).



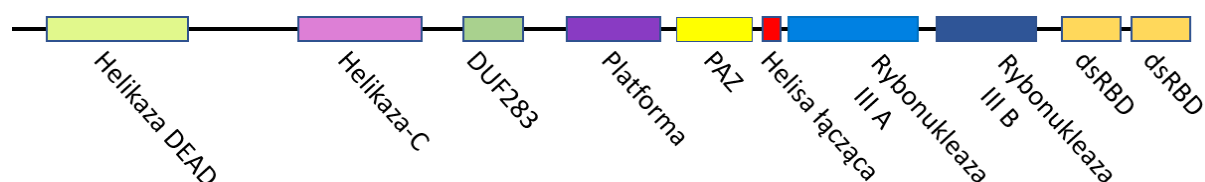
**Rys. 1.5.** Schematyczna reprezentacja domen w RNazach III, wg Zhang 2004. Zaznaczono domenę helikazy, której aktywność nie została dowiedziona dla wszystkich białek DCL oraz domenę PAZ, odpowiadającą za wiązanie jednoniciowego wystającego końca 3' z dwuniciowego RNA. Domen tych nie ma w pozostałych klasach RNaz typu III, zaś domeny bogate w prolinę (ang. Pro-rich), argininę i serynę (ang. RS-rich) są prawdopodobnie odpowiedzialne za oddziaływanie z innymi białkami (Johanson 2013, Song 2017).

W królestwie Prokariota występuje tylko klasa I, która jest odpowiedzialna za dojrzewanie rRNA, biogenezę małego jądrowego RNA (snRNA) oraz małego jądrowego RNA (snoRNA) a także przecina niektóre mRNA. Białko to występuje w postaci monomeru zbudowanego z domeny RNazyIII i domeny wiążącej dwuniciowe RNA – dsRBD, zatem do przecięcia obu nici dsRNA konieczna jest dimeryzacja. U eukariontów również występują

białka I klasy, które określane są jako RTL (Ribonuclease Three-Like), i odpowiadają funkcji białka bakteryjnego (Conrad 2002). Dwie pozostałe klasy różnią się między sobą zarówno budową jak i funkcją oraz stanowią istotną różnicę pomiędzy roślinami a zwierzętami. Cechą wspólną białek klasy II i III jest ich budowa obejmująca dwie domeny rybonukleazowe, przez co białko funkcjonuje jako monomer. U zwierząt białka klasy II (Drosha) i III (Dicer) uczestniczą w procesie RNAi, każde z nich funkcjonuje pojedynczo i katalizuje tylko jeden rodzaj reakcji na specyficznym substracie (Carmell 2004). Do prawidłowego działania wymagają szeregu białek pomocniczych, w oddziaływaniu z którymi prawdopodobnie biorą udział rejony bogate w prolinę (Pro-rich) oraz argininę i serynę (RS-rich). RNaza III – Dicer (Bernstein 2001) jest jednym z pierwszych poznanych białek przetwarzających dsRNA na specyficzną odpowiedź (w odróżnieniu od niespecyficznego degradacji). Rozcina ona dwuniciowe cząsteczki dsRNA precyzyjnie na fragmenty od 21 nt do 25nt zależnie od wielkości odpowiednich domen w endonukleazie (Hannon 2002). Fragmenty te stanowią funkcjonalne srRNA w komórce. Cechą charakterystyczną produktów wszystkich RNaz III jest 2 nt wystający 3' koniec zakończony grupą hydroksylową podczas gdy na 5' końcu pozostaje grupa fosforanowa. U roślin nieobecne są białka klasy II, podczas gdy klasa III reprezentowana jest zazwyczaj przez 4-8 rodzajów białka typu Dicer – DCL (ang. Dicer-like) (Agrawal 2003). Zostaną one opisane w dalszej części pracy.

#### 1.4.2. Budowa białek typu Dicer

Białka typu Dicer są to duże, około 190 kDa, wielodomenowe enzymy, które zazwyczaj składają się z helikazy DEAD, helikazy-C, domeny DUF283, domeny PAZ, dwóch domen rybonukleazowych III (RIIIA i RIIB) i jednej lub dwóch domen wiążących dsRNA (dsRBD) (Rys. 1.6).



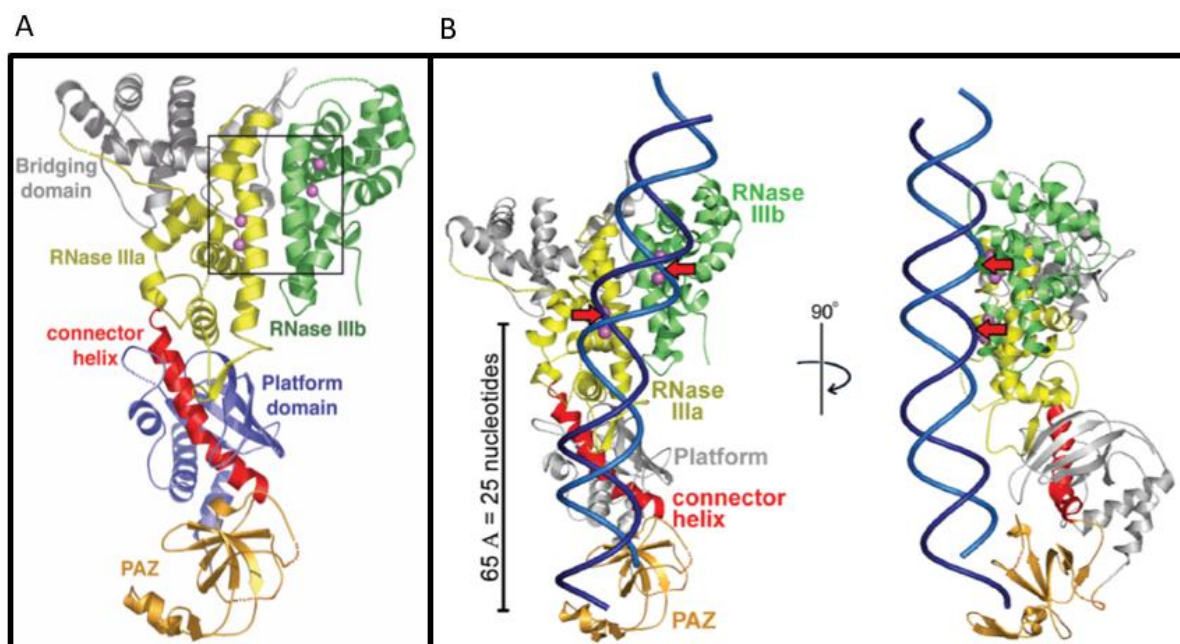
**Rys. 1.6.** Schematyczna reprezentacja budowy domenowej roślinnego białka DCL.

Domena helikazy obejmująca motyw DEAD i domena końca C helikazy tworzą razem region helikazowy, którego funkcją jest rozplatanie dwuniciowego RNA. Funkcją domeny DUF283 (ang. domain of unknown function) zdaje się być wiązanie białek DRB4 (AtDCL4) lub HYL1 (AtDCL1) (Qin 2010). PAZ – wiąże jeden z końców dsRNA zwłaszcza 2 nt wystające po stronie 3' terminalnej, może także posiadać kieszeń wiążącą 5' koniec. Platforma stanowi prawdopodobnie domenę stabilizującą, zaś w połączeniu z helisą łączącą reguluje odległość pomiędzy domenami PAZ a domenami katalitycznymi stanowiąc wyznacznik długości generowanych przez DCL produktów. Dwie domeny RNazy III (A i B) tworzą wewnątrzcząsteczkowy dimer, który (w obecności jonów  $Mg^{2+}$ ) rozrywa pojedyncze wiązanie fosfodiesterowe na każdej z nici dsRNA stanowiąc centrum katalityczne białka. Domeny dsRBD obecne u roślinnych DCL w liczbie jednej lub dwóch wiążą dwuniciowe RNA (dsRBD, ang. dsRNA binding domain).

Jedyną poznaną strukturę całego białka typu Dicer stanowi pochodzący z pierwotniaka *Giardia intestinalis* GiDCR (MacRae 2006, 2007). Jest on mniejszy niż pozostałe białka typu dicer. Składa się z platformy stabilizującej, domeny wiążącej dsRNA – PAZ, helisy łączącej i dwóch domen rybonukleazowych – w jednej z nich posiada sekwencję nieobecną u większości białek tego typu, tzw. mostek (Rys. 1.7A). W strukturze białka GiDCR brak domen: helikazy, DUF i dsRBD, lecz mimo to biochemiczne testy potwierdziły jego aktywność (MacRae 2006). Można przyjąć, że GiDCR jest to „molekularna linijka”, czyli część funkcjonalna i wystarczająca do precyzyjnego wycięcia fragmentu dsRNA o określonej długości, również w innych białkach, także tych z rodziny DCL. Przyłożenie do poznanej struktury dwuniciowego RNA – pomiędzy domenę wiążącą wolny koniec dsRNA a domenami RNaz przeprowadzającymi cięcie – potwierdza odległość zbliżoną do długości srRNA (Rys. 1.7B).

Zaledwie kilka struktur poszczególnych domen z białek Dicer lub DCL zostało doświadczalnie określonych na poziomie atomowym: (i) domena DUF283 z AtDCL4 (Qin 2010), (ii) druga domena wiążąca dsRNA (dsRBD) z DCL1 (Burdisso 2012), (iii) domena RIIB z ludzkiej Dicer – HsDCR, (iv) domena RIIB wraz z dsRBD pochodząca z mysiej Dicer – MmDCR (Du 2008, Takeshita 2007) oraz (v) fragment obejmujący region od platformy przez domenę PAZ do helisy łączącej (ang. connector helix) pochodzący z ludzkiej Dicer w kompleksie z siRNA (Tian 2014).





**Rys. 1.7.** Struktura białka GIDCR poznana eksperymentalnie metodą krystalografii rentgenowskiej (MacRae 2006). **A.** kolorami zaznaczono poszczególne domeny; **B.** Ilustracja oddziaływania białka z jego substratem dsRNA.

Na podstawie obserwacji w mikroskopie elektronowym (EM) zaproponowano ogólny pokrój całej HsDCR, w tym lokalizację poszczególnych domen (Lau 2012, Liu 2018b). Dotychczas niewiele jest danych dotyczących interakcji Dicer z jego substratem dsRNA. Ostatnie badania struktury fragmentu HsDCR w kompleksie z RNA wykazały obecność kieszeni wiążącej 2 nukleotydy 3' koniec w domenie PAZ oraz kieszeni fosforanowej wiążącej 5' koniec w domenie platformy (Tian 2014). Na podstawie danych z analizy Cryo-EM otrzymanych z rozdzielczością 29-31Å zostały także skonstruowane modele kompleksów całej HsDCR z RNA (Taylor 2013), a później także z rozdzielczością 4,4Å (Liu 2018b). Dodatkowe informacje dotyczące interakcji poszczególnych domen białka DCL z RNA mogą być uzyskane ze struktur krystalicznych homologicznych domen pochodzących z innych białek w kompleksie z RNA. Zostały one określone dla bakteryjnej RNazy III (Gan 2006, 2008), eukariotycznej domeny RNazy III (Liang 2014) oraz domen PAZ z białek AGO i PIWI (Ma 2012, Nakanishi 2013) oraz RNazy III z ludzkiej DROSHA (Kwon 2016).

### 1.4.3. Funkcja białek typu Dicer

Białka Dicer oraz DCL przeprowadzają reakcję hydrolizy dwuniciowego RNA. Efektem ich działania są małe regulatorowe RNA, służące do modulacji poziomu ekspresji genów. Pomimo uniwersalności procesu RNAi, różnice w biogenezie miRNA u roślin i zwierząt mają swoje odzwierciedlenie zarówno w budowie jak i sposobie funkcjonowania białek w nią zaangażowanych.

Istnieje tylko jeden gen kodujący białko Dicer w komórkach ssaczy, dwa u owadów i grzybów i 4-8 genów białek typu Dicer (DCL) w roślinach (Kapoor 2008). Roślina modelowa, *Arabidopsis thaliana* koduje cztery DCL (AtDCL1-4). **AtDCL1** głównie uczestniczy w biogenezie miRNA (Reinhart 2002, Bartel 2004, Voinnet 2009), ale może także przetwarzać dwuniciowe RNA (Bouche 2006). **AtDCL2** jest częścią systemu obrony przeciwvirusowej i generuje 22-nt siRNA pochodzące z wirusowego dsRNA (Xie 2004, Mlotshwa 2008). Ponadto jest zaangażowane w generowanie nat-siRNA. **AtDCL3** wymagane jest do endogennego tworzenia 24-nt siRNA (ha-siRNA, ra-siRNA) z długich dwuniciowych RNA wytwarzanych przez roślinną polimerazę RNA zależną od RNA-2 (RDR2). SiRNA generowane przez AtDCL3 biorą udział w modyfikacji chromatyny, a zatem wpływają na jej strukturę przyczyniając się do zachowania stabilności genomu (Chan 2004, Zhong 2012). **AtDCL4** jest głównym czynnikiem odpowiedzialnym za odporność roślin na wirusy przez cięcie dwuniciowych RNA powstających w procesie ich namnażania. AtDCL4 generuje także pierwotne i drugorzędowe siRNA o wielkości 21-nt (ta-siRNA) z długich dwuniciowych RNA wytwarzanych przez roślinną polimerazę RNA zależną od RNA-6 (RDR6) z transkryptów roślinnych genów TAS (Dunoyer 2005, Gascioli 2005). Po rozcięciu transkryptu genu TAS, za pomocą miRNA oba fragmenty (jeden zawierający wolny 5' zaś drugi 3' koniec) stanowią matrycę dla RDR6, która syntetyzuje do nich nici komplementarne (Rajeswaran 2012). W rezultacie otrzymuje się dwie długie dwuniciowe cząsteczki RNA nieposiadające zaburzeń strukturalnych. AtDCL4 rozcina takie dsRNA w sposób sekwencyjny, odcinając kolejne 21-nt fragmenty począwszy od wolnego końca i generuje zbiór krótkich drugorzędowych srRNA. Niedawne doniesienia wykazały, że AtDCL4 mogą także rozcinać jednoniciowy RNA, zawierający strukturę spinki do włosów powstałą w wyniku przedłużonej transkrypcji (ang. read-through transcripts) (Liu 2012b, Fukudome 2017).

#### 1.4.4. Dodatkowe białka uczestniczące w biogenezie srRNA

##### 1.4.4.1. Białka typu AGO

Białka typu AGO należą do rodziny białek Argonaute o dużej zachowawczości ewolucyjnej zarówno u eukariontów jak i prokariotów. Stanowią one jeden z czynników koordynujących regulację ekspresji genów w komórkach. Rodzina białek Argonaute dzieli się na trzy grupy: (i) białka typu AGO (ang. AGO-like proteins) szeroko rozpowszechnione wśród wszystkich żywych organizmów, (ii) białka typu PIWI obecne w komórkach rozrodczych zwierząt, oraz (iii) występujące u *Cenorhabditis elegans*. W dalszej części opisane zostaną białka należące tylko do pierwszej z wymienionych klas, jako zaangażowane w procesy RNAi u roślin.

Białka AGO stanowią część kompleksu RISC i wraz ze związaną cząsteczką srRNA stanowią bezpośredni czynnik odpowiedzialny za rozpoznanie i przecięcie transkryptu skierowanego do degradacji (Moazed 2009). Od niedawna znana jest pełna struktura eukariotycznych białek AGO (Schirle 2014, Elkayam 2017). Zbudowane są z czterech domen: N, PAZ, MID i PIWI (Kapoor 2008). Domena **N** umożliwia aktywność katalityczną (Derrien 2018), odgrywa również rolę w rozplataniu dupletu RNA i selekcji nici o mniej stabilnym końcu 5' (Hauptmann 2013, Kwak 2013). Domena **PAZ** wiąże charakterystyczny 2-nt wystający koniec dwuniciowego RNA, który może stanowić duplet miRNA/miRNA\* lub dwuniciowy siRNA. Domena **MID** wiąże 5' monofosforan srRNA do białka. Uczestniczy także w wyborze nici pozostającej w RISC oraz odrzucanej. Domena **PIWI** wykazuje duże podobieństwo do RNAzy H (Höck 2008). Jej funkcją jest przecinanie transkryptu w miejscu wyznaczonym przez 10 i 11 nt małego regulatorowego RNA (ang. RNA guide strand). Jako, że tylko jedna nić zostaje przecięta, cała cząsteczka srRNA pozostaje związana z AGO, a zatem może funkcjonować przez wiele obrotów enzymu.

*Arabidopsis thaliana* koduje 10 białek AGO (Morel 2002, Zheng 2007), funkcje niektórych z nich zdają się pokrywać lub nie są do końca poznane. Głównym białkiem zaangażowanym w potranskrypcyjne wyciszanie większości transgenów jest **AGO1** (Carbonell 2012, Bajczyk 2019). Występuje zarówno w cytoplazmie jak i w jądrze komórkowym (Dolata 2018). **AGO10** odpowiada za regulację niektórych procesów rozwojowych (Liu 2009, Ji 2011)

oraz stanowi negatywny regulator AGO1 na poziomie białkowym (Mallory 2009). **AGO2** (i w mniejszym stopniu **AGO3**) regulują odpowiedź rośliny na zakażenie bakteryjne (Zhang 2011). Najbardziej do nich podobne **AGO7** (i w mniejszym stopniu **AGO1**) odpowiadają za powstawanie ta-siRNA (Adenot 2006, Qu 2008). **AGO4**, **AGO6** i **AGO9** duplikują swoje role w akumulacji ha-siRNA oraz metylacji DNA i transkrypcyjnego wyciszania genów transpozonów i sekwencji repetytywnych (Qi 2006, Zheng 2007, Havecker 2010). **AGO5** wiąże się z siRNA o różnych długościach, zawierających 5' terminalną cytozyną (Takeda 2008). Prawdopodobnie bierze udział w wyciszaniu RNA wirusowego.

#### 1.4.4.2. Polimeraza RNA zależna od RNA

Już w 1993 roku odkryto, że RDR wyizolowana z pomidora przeprowadza *in vitro* reakcję syntezy komplementarnego RNA (cRNA) na matrycy RNA jak i DNA przy użyciu starterów bądź bez nich (Schiebel 1993b). Później wykazano, że takie RNA odgrywa istotną rolę w wyciszaniu genów na drodze RNAi, zaś produkty działania roślinnych RDR mogą być substratem dla białek DCL.

*Arabidopsis thaliana* posiada sześć genów kodujących RDR, z tego aktywność wykazano eksperymentalnie dla trzech z nich RDR1, RDR2 i RDR6 (Xie 2004, Polydore 2018). Są one na tyle specyficzne, że każda z nich uczestniczy w innym szlaku powielania sygnału. **RDR1** odpowiada u *Arabidopsis* za obronę przed patogenami. Jej ekspresja jest indukowana czynnikami biotycznymi i abiotycznymi. **RDR2** stanowi obronę przed transpozonami i sekwencjami repetytywnymi zawartymi w genomie. RDR2 razem z DCL3 (ang. Dicer-like protein 3) tworzą 24-nukleotydowe siRNA z retroelementów, transpozonów, endogennych powtórzeń i odwróconych powtórzeń a także transgenów zawierających sekwencje powtarzające się (Chan 2004, Xie 2004). **RDR3**, **RDR4** i **RDR5** zostały odkryte na podstawie analiz bioinformatycznych. Ich aktywność nie została potwierdzona, zaś ich podobieństwo może świadczyć, że powstały wskutek dwukrotnej duplikacji pierwotnego genu. **RDR6** jest zaangażowana w produkcję ta-siRNA, oraz związana ze szlakiem regulacji genów odpowiedzialnych za rozwój organów wegetatywnych (Peragine 2004).

Dowiedziano, że RDR, ani żadne białko o podobnej roli, nie występuje u muszki owocówki i ssaków (Stein 2003).

#### 1.4.4.3. Inne białka pomocnicze

Oprócz enzymów przeprowadzających reakcje prowadzące do powstania małych regulatorowych RNA występuje jeszcze szereg białek pomocniczych. Ich rolą jest ułatwienie przyjęcia właściwej pozycji enzymu i substratu lub transport, czy też przygotowanie substratu do reakcji enzymatycznej.

Właściwa akumulacja miRNA zależy między innymi od aktywności dwóch białek jądrowych: **HEN1** i **HYL1** (Park 2002, Han 2004, Vazquez 2004, Yang 2014). Jądrowa metylotransferaza, HUA ENHENCER1 (**HEN1**), metyluje grupę 2' hydroksylową 3' końca dojrzałego miRNAs (Yu 2005, Huang 2009). Modyfikacje takie wpływają na właściwe rozpoznanie srRNA przez odpowiednie białko AGO. **HYL1** (znane też jako DRB1) jest jądrowym białkiem wiążącym dwuniciowe RNA, które razem z DCL1 tworzy kompleks regulujący cięcie pierwotnego transkryptu miRNA (Yang 2006, Eamens 2009). Analogiczną do niego funkcję pełni **DRB4**, które współpracuje z pozostałymi roślinnymi białkami DCL (Nakazawa 2007, Qu 2008). **SERRATE** (SE) prawdopodobnie działa jako rusztowanie, jako białko zdolne do wiązania zarówno białka jak i RNA decyduje o właściwym położeniu prekursora miRNA w katalitycznym wnętrzu roślinnego DCL1 (Machida 2011). Oba białka HYL1 i SE wspierają aktywność DCL1 w biogenezie miRNA zarówno na poziomie pri- jak i pre-miRNA, przez co znacząco zwiększają wydajność reakcji cięcia (Dong 2008). **HASTY** (HST) koduje białko, które uczestniczy w transporcie miRNA lub pre-miRNA z jądra do cytoplazmy (Park 2005).

## 2. Cel pracy

Celem niniejszej pracy było poznanie strukturalnych uwarunkowań biogenezy roślinnych srRNA na przykładzie wybranych białek oraz cząstek RNA pochodzących z rośliny modelowej *Arabidopsis thaliana*. Małe regulatorowe RNA powstające jako produkty aktywności białek DCL1 mają długość 21 nt (miRNA). Powstające w efekcie działania białka DCL4 – siRNA – są tej samej długości i uruchamiają ten sam mechanizm wyciszania genów. Jednakże sam dupleks miRNA złożony z nici wiodącej (miRNA, ang. guide) oraz towarzyszącej (miRNA\*, ang. passenger) ma odmienną strukturę drugorzędową, gdyż w odróżnieniu od dupleksu siRNA zawiera nieregularności. Głównym przedmiotem badań była struktura dupleksu miRNA:miRNA\* oraz oddziaływanie pomiędzy nicią wiodącą miRNA a miejscem docelowym w mRNA. Białko DCL4 jako generujące małe regulatorowe RNA z długich dwuniciowych RNA nie zawierających nieregularności wydaje się mieć strukturę sztywniejszą, dostosowaną do jednorodnej budowy swoich substratów. Dlatego jako drugi obiekt badań strukturalnych uwarunkowania biogenezy srRNA wybrano białko AtDCL4. Jego model zaproponowany w ramach przedłożonej rozprawy może być bardziej wiarygodny biologicznie i lepiej uzasadniony za pomocą metod *in silico* niż w przypadku pozostałych białek z rodziny DCL, których substraty nie są jednoznacznie scharakteryzowane strukturalnie.

Cel zdefiniowany w niniejszej pracy cel osiągnięto realizując następujące zadania badawcze:

- Charakterystyka termodynamiczna profili dupleksów mikroRNA z sekwencjami docelowymi oraz dupleksów miRNA:miRNA\* zawartych w prekursorach miRNA,
- Opracowanie metody oraz stworzenie narzędzia obliczeniowego AmiRNA Designer umożliwiającego projektowanie sekwencji wyciszających zadany gen,
- Charakterystyka porównawcza sekwencji białek roślinnych DCL1-4.
- Opracowanie modelu struktury przestrzennej białka AtDCL4 w kompleksie z jego substratem dsRNA, oraz zbadanie związku pomiędzy strukturą enzymu a długością wycinanych przez niego fragmentów dsRNA.

### 3. Materiały i metody

#### 3.1. Materiały

##### 3.1.1. Sekwencje RNA

Sekwencje prekursorów miRNA zostały pobrane z bazy danych miRBase rel18 (Griffiths-Jones 2008) (w tym przypadku ograniczono do organizmu – *A. thaliana*). Sekwencje miRNA oraz ich miejsca oddziaływania w mRNA (targety) zostały pobrane z bazy danych Arabidopsis Small RNA Project (ASRP, <https://asrp.danforthcenter.org> (Gustafson 2005)). Uwzględniono zarówno targety potwierdzone eksperymentalnie jak i wyznaczone komputerowo. Wszystkie sekwencje miRNA wykorzystane do wyznaczenia zależności sekwencyjnych oraz profili termodynamicznych zostały wymienione w załączniku nr 1.

Analiza potencjalnych niespecyficzných miejsc oddziaływania amiRNA oraz amiRNA\* została wykonana przy wykorzystaniu danych z PlantGDB (Dong 2004, Duvick 2008) służących jako źródło wszystkich genów ulegających ekspresji w *A. thaliana*. Fragment tego pliku pokazano w załączniku nr 2.

Sekwencje badanych genów docelowych Imm1 i Agb1 zostały pobrane z bazy danych TAIR (<http://www.arabidopsis.org/>). Ich sekwencje kodujące zamieszczono w załączniku nr 3.

Sekwencje wygenerowane przez program „AmiRNA Designer” zostały przeanalizowane pod kątem efektu niespecyficznego oddziaływania (ang. off-target) przy użyciu skryptów w Perlu napisanych przez dr hab. inż. Agnieszkę Rybarczyk (Politechnika Poznańska, Wydział Informatyki i Telekomunikacji, Instytut Informatyki). Metoda określania bezpośredniej interakcji między transkryptem a miRNA opiera się na wyrażeniach regularnych zastosowanych do wyszukiwania motywów w ich sekwencji. W przypadku każdego miRNA, poszukiwanie miejsca wiązania przeprowadzono wzdłuż całych transkryptów z przesunięciem jednego nukleotydu.

### 3.1.2. Sekwencje białek

W niniejszej pracy wykorzystano trzy zestawy sekwencji białek.

Pierwszy z nich stanowił sekwencje roślinnych białek DCL zarówno potwierdzonych eksperymentalnie jak i przetłumaczonych na kod białkowy z pełnych poznanych i złożonych sekwencji genomowych (cDNA). Sekwencje białkowe wszystkich czterech białek AtDCL posłużyły jako wzorce do wyszukania homologicznych sekwencji w genomach roślinnych zdeponowanych w bazie danych Plant Genome Database (<http://www.plantgdb.org>, (Dong 2004) opcja wyszukiwania sekwencją białkową). Przeszukano każdy z opisanych tam gatunków roślin oddzielnie. Otrzymany w ten sposób zestaw został dodatkowo zweryfikowany przez przeszukanie nim bazy danych GenBank za pomocą programu BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>).

Drugi zestaw składał się ze zweryfikowanych metodami eksperymentalnymi sekwencji białek zawierających domenę RNAzy III zdeponowanych w bazie danych UniProt (<http://www.uniprot.org>). Sekwencje wybrano w taki sposób, by uniknąć nadreprezentacji białek ssaków lub kręgowców, dla których znaleziono znacznie więcej trafień niż dla białek roślinnych czy owadzych. Do tego zestawu dodano także bakteryjną RNAzę typu III.

Dla każdej z sekwencji wyszukanych w bazach danych oznaczono domeny białkowe wykorzystując narzędzia ScanProsite (<http://prosite.expasy.org/scanprosite>, (De Castro 2006)) oraz SMART (<http://smart.embl-heidelberg.de>, (Schultz 1998)). Wyniki tych analiz wykorzystano do wykonania porównań wielosekwencyjnych (ang. MSA - multiple sequence alignment): 1) porównanie globalne całych dostępnych sekwencji aminokwasowych, 2) porównanie fragmentów zawierających odpowiednik białka GiDicer – stanowiących funkcjonalny rdzeń enzymów DCL, 3) porównanie bloków domen oraz, 4) porównanie każdej domeny oddzielnie. Niektóre z porównań zamieszczono w załączniku nr 5.

Konserwatywne regiony ze wszystkich porównań posłużyły jako wzorzec do wyszukiwania sekwencji białek o poznanej strukturze z zastosowaniem metody wykorzystującej profile sekwencyjne PSI-BLAST (Altschul 1997) oraz DELTA-BLAST (Boratyn 2012). Ponadto uzyskane w ten sposób struktury poszczególnych domen oraz bloków domen zostały wykorzystane do wyszukania dostępnych struktur za pomocą metaserwera GeneSilico



(Kurowski 2003) oraz serwera I-TASSER, który wykorzystuje metodę przewlekania do budowania modelu białko 3D de novo (Roy 2010, Zhang 2008). Wyniki działania obu metod zostały ze sobą porównane. Otrzymane w efekcie sekwencje białek o znanej strukturze 3D zostały zawarte w trzecim zestawie danych. Te sekwencje poddano przyrównaniu strukturalnemu w oparciu o nakładanie się domen z wykorzystaniem algorytmu STAMP zaimplementowanego w pakiecie VMD (Humphrey 1996). Bazując na tak otrzymanym przyrównaniu strukturalnym dokonano ręcznego udokładnienia nakładania sekwencji. Wszystkie przyrównania częściowe zostały złożone razem i ręcznie poprawione w oparciu o wyniki analiz strukturalnych. Tak przygotowano ostateczne przyrównania – sekwencji celu do szablonów – które stanowiły dane wejściowe do modelowania białka w programie Modeller.

### 3.1.3. Struktury białek

Struktury białek pobrano z bazy danych PDB na podstawie wyszukiwania PSI BLAST poszczególnymi domenami białek roślinnych DCL. Dodatkowo rozpatrywano białka o mniejszym podobieństwie sekwencyjnym ale o znanej funkcjonalności podobnej do poszczególnych domen analizowanych białek DCL.

**Tab. 3.1.** Wykaz szablonów używanych jako wzorce do modelowania homologicznego AtDCL4.

Modelowana domena	Nazwa	PDB ID	Zakres sekwencji
<b>PAZ</b>	GiDCR	2qvw	150-251
	DmAGO1	1r4k	32-423
	HsAGO1	1si3	249-346
	HsPIWIL1	3o7v	402-501
	HsDCR	4ngd	914-1031
<b>Helisa łącząca</b>	GiDCR	2qvw	242-340
	HsDrosha	5b16	850-973
	HsDCR	2ngd	1023-1053
	AaRIII	2nuf	14-44
	MmDCR1	3c4b	1661-1193
<b>RNAzy</b>	AaRIII	2nuf	<b>A:</b> 4-148 <b>B:</b> 36-148
	MmDCR1	3c4b	<b>A:</b> 1650-1834 <b>B:</b> 1685-1834

Dodatkowo jako wzorce położenia RNA względem domeny PAZ posłużyły struktury:

- i) TtAgo (PDB ID: 3dlb) , TtAgo (PDB ID: 3dlh), HsAgo1 (PDB ID: 1si3), TtAgo (PDB ID: 3f73), HsPIWIL1 (PDB ID: 307v), HsDCR (PDB ID: 4nha)

W tym przypadku porównywane było zarówno zakotwiczenie końca 3' RNA w kieszeni PAZ, jak i rozkład potencjału elektrostatycznego na powierzchni białka w miejscu, gdzie znajdował się 5' koniec RNA.

- ii) Struktury fragmentu obejmującego domeny PAZ, platformę i helisę łączącą białka HsDCR w kompleksach z dsRNA (PDB ID: 4ngb, 4ngc, 4ngd, 4ngf, 4ngg, 4nh3, 4nh5 i 4nh6).

#### 3.1.4. Programy komputerowe wykorzystane w badaniach

**UNAFold** (Markham 2008) został włączony do programu AmiRNA Designer, ze względu na opcję przewidywania struktury dupleksu z dwóch nici RNA.

**psRNATarget** (<https://plantgrn.noble.org/psRNATarget/home> Dal 2011) była używana do weryfikacji zaprojektowanych cząsteczek amiRNA jako potencjalne srRNA.

Wszystkie przyrównania sekwencji aminokwasowych białek (lub ich fragmentów) były prowadzone z wykorzystaniem dostępnych narzędzi internetowych **CLUSTALW2** (Higgins 1994, Larkin 2007) oraz **T-Coffee** (Notredame 2000).

Programy **BioEdit** (Hall 1999) i **GenDoc** (Nicholas 1997) zostały wykorzystane do wizualizacji, edycji oraz oceny przyrównań sekwencji białkowych.

**Jpred3** (<http://www.compbio.dundee.ac.uk/www-jpred/>, (Cole 2008)) i **JUFO 3D** (<http://www.meilerlab.org>, (Meiler 2003)) użyto do określania drugorzędowej struktury białka.

**PCONS5** (Wallner 2005) na meta-serwerze **Genesilico** (Kurowski 2003) analizuje zestaw modeli białek i szuka powtarzających się trójwymiarowych wzorów strukturalnych. Wykorzystywany do wyszukania wzorców dla zadanych sekwencji modelowanych białek.

**Modeller** (Šali 1993) jest to program umożliwiający modelowanie homologiczne białka. Pozwoliła na wygenerowanie struktury trzeciorzędowej białka w oparciu o zadany szablon (lub kilka) oraz zadane więzy odległościowe pomiędzy określonymi atomami.

**MetaMQAPII metaserver** (Pawlowski 2008) wykorzystano do oceny struktur białek.

Potencjał elektrostatyczny na powierzchni białka został wygenerowany za pomocą narzędzi pakietu **PyMOL – APBS Tools** (Baker 2001).

Oprogramowanie **PDB2PQR** (Dolinsky 2004, 2007) było wykorzystywane do przypisania cząstkowych ładunków dla każdego atomu na podstawie pola siłowego CHARMM (Brooks 1983).

Stan protonacji reszt aminokwasowych został ustalony programem **PROPKA** przy pH 7 (Li 2005).

**KoBaMin** (<http://chopra-modules.science.purdue.edu/modules/kobamin/html> Rodriguez 2012) wykonuje minimalizację struktury białka w oparciu o potencjały statystyczne zbudowane w oparciu znane struktury białek.

**ProQ** (<https://proq.bioinfo.se/ProQ/ProQ.html> Wallner 2003) w oparciu o cechy strukturalne przewiduje jakość modelu białka. Oceniano za jego pomocą jakość modeli białek (wyjętych z ich kompleksów z RNA) po każdym etapie modelowania oraz dynamiki molekularnej.

Podczas testowania i interpretacji wyników poszczególnych etapów modelowania białka niezbędne było wykorzystanie programów pozwalających na wizualizację plików w formacie PDB: **PyMOL** (DeLano 2002), **Swiss PDB Viewer** (Guex 1997) oraz **VMD** (Humphrey 1996).

Programy do wizualizacji zostały wykorzystane także do wykonania ilustracji zamieszczonych w rozprawie.

## 3.2. Metody

### 3.2.1. Drugorzędowa struktura RNA

Do uzyskania drugorzędowej struktury RNA stosowano program UNAFold (Markham 2008). Jest to kompleksowy pakiet oprogramowania do przewidywania zwiwania się jednoniciowego RNA lub hybrydyzacji między dwoma pojedynczymi nićmi kwasów nukleinowych. Nazwa pochodzi od „Unified Nucleic Acid Folding”. Wykorzystuje się tu założenie, że struktura cząsteczki RNA o całkowitej najniższej energii swobodnej, jako najtrwalsza termodynamicznie, z największym prawdopodobieństwem jest odwzorowaniem struktury rzeczywiście występującej w organizmach żywych. Program ten jest pochodną programu mfold, który do obliczenia minimalnej energii swobodnej ( $\Delta G$  – wyrażaną w kcal/mol) wykorzystuje model najbliższego sąsiedztwa. Jego parametry zaś oparte są na podstawie uzyskanych doświadczalnie danych termodynamicznych.

**Model najbliższego sąsiedztwa** (NN model, ang. the Nearest-Neighbour model) został pierwotnie opracowany w celu przewidywania energii swobodnej tworzenia się podwójnej helisy DNA z dwóch rozdzielonych nici. Później został rozszerzony o zwiwanie pojedynczej nici DNA a także RNA. Podstawą, na której opiera się model najbliższego sąsiedztwa jest fakt, że energia tworzenia motywu strukturalnego (np.: pary zasad, niesparowania, wybrzuszenia czy pętli) zależy nie tylko od jego sekwencji, ale także od sekwencji w najbliższym sąsiedztwie. Eksperymenty termochemiczne (topnień UV) pozwoliły na skonstruowanie macierzy motywów 2x2 nt, dla których oznaczono energię tworzenia. Na energię motywu składa się (i) tworzenie wiązań wodorowych pomiędzy resztami zasad azotowych, (ii) oddziaływania warstwowe pomiędzy sąsiednimi zasadami z tej samej nici oraz (iii) spadek entropii układu. W przypadku RNA, macierz motywów nie jest ograniczona tylko do parowania typu Watson-Crick, ale zawiera także inne możliwe pary, które tworzą drugorzędową strukturę RNA. Model NN jest wspierany eksperymentami kalorymetrycznymi oraz badania absorpcji UV. Pomiar ciepła absorbowanego i emitowanego przez substancję chemiczną podczas jej ogrzewania pozwalają wnioskować o entalpii i entropii jej powstawania. Szereg eksperymentów wykonanych dla różnych sekwencji RNA i DNA umożliwił wyznaczenie wartości entalpii i entropii dla każdego z motywów.

W pracy wykorzystano następujące parametry i predefiniowane skrypty:

- reguły energii: RNA(2.3) (Walter 1994)
- temperaturę 22°C.
- hybrid-2s.pl (dla dwóch sekwencji RNA)
- hybrid-ss-2s.pl (dla pojedynczej sekwencji RNA)

### 3.2.2. Dopasowanie sekwencji aminokwasowych

Przyrównanie sekwencji białek polega na znalezieniu podobieństw w sekwencjach aminokwasowych dwóch lub więcej białek na całej długości ich cząsteczki (dopasowanie globalne) lub fragmentów (dopasowanie lokalne). W latach dziewięćdziesiątych na podstawie przyrównań wysoce spokrewnionych sekwencji i obserwacji występujących substytucji aminokwasów opracowano macierze podstawień, w których każdej substytucji przypisano określoną wartość prawdopodobieństwa, z jaką dany aminokwas może być wymieniony na inny. Dodatkowo wartości odpowiadają substytucjom, których prawdopodobieństwo jest większe niż prawdopodobieństwo zdarzenia losowego, natomiast wartości ujemne – substytucjom o niższym prawdopodobieństwie. Przerwom w przyrównaniu sekwencji także jest przypisana wartość liczbowa. Przerwy odpowiadają insercjom i delecjom, których prawdopodobieństwo ewolucyjne jest stosunkowo niskie, więc wartości te są ujemne i noszą nazwę kary za wprowadzenie przerwy oraz kary za wydłużanie przerwy. W oparciu o wartości z macierzy oraz kary za przerwy najlepsze dopasowanie wyszukiwane jest algorytmami, które maksymalizują wartość miary przyrównania. Można do tego zastosować algorytmy dokładne oparte na metodzie programowania dynamicznego (np. algorytm Needlemana i Wunscha (Needleman 1970)) lub metody heurystyczne, dające rozwiązania suboptymalne. Ze względu na złożoność problemu w praktyce zazwyczaj stosuje się metody heurystyczne.

Program CLUSTALW2 (Higgins 1994, Larkin 2007) służy do generowania dopasowań dla wielu sekwencji białek jednocześnie. Algorytm opiera się na dopasowaniu najbardziej podobnych sekwencji (ang. neighbour joining), następnie dopasowuje się kolejne sekwencje o coraz niższym podobieństwie. Najpierw dokonywane są obliczenia macierzy przybliżonej odległości między każdą parą sekwencji w oparciu o wyniki dopasowania sekwencji parami.

Następnie algorytm wykorzystuje metodę łączenia sąsiedniego z ukorzeniem w punkcie środkowym (ang. midpoint rooting), do utworzenia drzewa filogenetycznego, które jest używane do generowania globalnego przyrównania. Drzewo służy jako zgrubny szablon dla określenia spokrewnionych grup, w obrębie których insercje i delecje są zbliżone.

Innym wykorzystywanym w niniejszej pracy algorytmem jest T-Coffee (Notredame 2000). Także jest to algorytm heurystyczny. Podobnie tworzona jest biblioteka dopasowań zadanych sekwencji każdej z każdą (zarówno globalne jak i lokalne). Każdej pozycji w dopasowanych parach sekwencji przypisywana jest wartość liczbowa, a następnie zliczane są analogiczne wartości dla każdej pary dopasowanych sekwencji. Wszystkie wartości są dodatnie i nie ma potrzeby wprowadzania kar za przerwy – przerwa jest wartością zerową.

Drzewo filogenetyczne analizowanych sekwencji roślinnych białek DCL jest tworzone wraz z ustaleniem optymalnego przyrównania i zostało zaprezentowane w załączniku 5. Oba programy używane były w pracy jako narzędzia webowe z domyślnymi parametrami.

### 3.2.3. Modelowanie struktur białek

Generowanie trzeciorzędowych struktur białek zostało przeprowadzone metodą modelowania homologicznego z zastosowaniem oprogramowania Modeller (Sali 1994). Jako że sekwencje domen białka AtDCL4 nie miały dużego podobieństwa do sekwencji najlepszych szablonów, posłużono się metodami ustalającymi struktury drugorzędowe, a następnie według nich poprawiano dopasowane. Modeller nie kopiuje z szablonu współrzędnych atomów tylko ustala odległości i kąty pomiędzy atomami, a następnie przenosi je jako więzy przestrzenne (ang. restrains) na odpowiednie atomy aminokwasów modelowanego białka (według zadanego dopasowania sekwencyjnego). Zaimplementowane w programie mechanizmy porównują je z idealnymi wartościami długości wiązań i kątów. Dodatkowo w pracy zastosowano własne więzy pochodzące z analizy bioinformatycznej: (i) przewidywane struktury drugorzędowe dla regionów, które nie mają homologicznego odpowiednika w szablonie, (ii) zależności odległościowe pomiędzy aminokwasami tworzącymi centrum katalityczne AtDCL4 pochodzące ze struktury AA\_RNAselll, (iii) fragmenty ocenione jako wymodelowane poprawnie w programie I-Tasser. Model budowany jest tak, aby zminimalizować naruszenie wszystkich wiązań. W końcowym etapie budowy modelu

przeprowadzana jest minimalizacja energii w polu siłowym CHARMM22 (MacKerell 1998, <http://yuri.harvard.edu/>), która zapewnia poprawną stereochemię i korzystne oddziaływania pomiędzy grupami funkcyjnymi.

I-Tasser (Iterative Threading ASSEmbly Refinement) to narzędzie internetowe, umożliwiające modelowanie białka metodą przewlekania (ang. threading), na podstawie podobieństwa krótkich fragmentów modelowanej sekwencji do zdeponowanych w bazie danych PDB struktur poznanych metodami eksperymentalnymi. Często wykorzystywany jest w konkursach CASP. Jednakże wyniki modelowania struktury całego białka ATDCL4 tą metodą (a nawet poszczególnych domen) nie były zadowalające, zatem program ten nie mógł służyć jako główne narzędzie do wygenerowania modelu. Mimo to dostarczył kilku istotnych wskazówek, które zostały wykorzystane jako dodatkowe więzy dla Modellera.

#### 3.2.4. Ocena modeli białek

Ocena modelu dokonywana była przez programy ProQ oraz MetaMQAPII (Pawlowski 2008). Programy te oceniają jak bardzo fragmenty modelu białka są podobne do znanych struktur występujących w innych białkach. Sprawdzają między innymi prawidłowość położenia poszczególnych reszt aminokwasowych w kontekście struktury (np. ułożenie reszt hydrofobowych wewnątrz struktury białka). Metody te dopisują do pliku ze strukturą modelu dodatkową kolumnę „ $\beta$ -factor” zawierającą liczbową wartość oceny poprawności każdej reszty aminokwasowej. W zależności od wartości tej liczby programy do wizualizacji (np. PyMOL) przypisują nieprawidłowo wymodelowanym fragmentom kolor czerwony, a prawidłowo wymodelowanym – niebieski, oraz odcienie kolorów żółtego i zielonego dla wartości pośrednich.

#### 3.2.5. Dynamika molekularna

W celu określenia stabilności otrzymanych modeli AtDCL4 związanych z cząsteczką RNA zbudowane *in silico* kompleksy poddawano symulacjom dynamiki molekularnej wykorzystując pola siłowe mechaniki molekularnej. Symulacja dynamiki molekularnej (MD) jest metodą pozwalającą na badanie w czasie ewolucji układu atomów, których zmiany położenia

i prędkości są efektem przyłożonego pola siłowego. Ruchy te można opisać klasycznym równaniem Newtona dla układu wielu cząstek. Równania ruchu Newtona całkuje się w bardzo krótkich odcinkach czasu stosując np. algorytm Verleta (Verlet 1967). Pole siłowe to funkcja energii potencjalnej oraz jej parametry (pochodzące z badań eksperymentalnych oraz teoretycznych obliczeń). Człony pola siłowego obejmują energię wiązań, kątów, kątów torsyjnych, oddziaływania Van der Waalsa i oddziaływania elektrostatyczne. Parametry te są charakterystyczne dla rodzaju badanych cząsteczek. Pola siłowe stosowane w badaniach obiektów biologicznych to np. Charmm lub Amber.

Wynikiem symulacji dynamiki molekularnej jest trajektoria ruchu atomów, przedstawiona jako zestaw struktur zapisanych w określonych odstępach czasu. Przykładowe parametry, służące do analizy trajektorii i danych z MD to:

- wykres energii całkowitej, kinetycznej i potencjalnej w czasie, po zrównoważeniu układu powinny one fluktuować wokół stałej wartości.
- RMSD (ang. root mean square deviation) - średnie odchylenie kwadratowe wszystkich, bądź wybranych, atomów białka w czasie. Zazwyczaj wyznaczane jest względem konformacji początkowej lub struktury poznanej eksperymentalnie.
- wiązania wodorowe – ich ogólna liczba i stabilność podczas symulacji lub wartości odległości donor-akceptor, czy kątów donor-wodór-akceptor. Mogą one także odzwierciedlać ogólne właściwości termodynamiczne układu.

Analiza drgań normalnych (ang. normal mode analysis, NMA) dostarcza informacji o ruchach wokół struktury wejściowej, która jest w położeniu równowagi. Globalne ruchy białek (fluktuacje) generowane za pomocą drgań normalnych dobrze odwzorowują ich znaczenie funkcjonalne. Do analizy drgań normalnych wykorzystany został model sieci elastycznej, w którym oddziaływania pomiędzy elementami struktury modelowane są za pomocą potencjału harmonicznego. W przypadku analiz przeprowadzonych w tej pracy białko reprezentowane było za pomocą atomów C $\alpha$ . Na podstawie zbioru fluktuacji można wyznaczyć drgania normalne, wykazujące swego rodzaju wzorzec, gdzie pewne fragmenty struktury poruszają się w tym samym kierunku w zgodnym czasie (Hinsen 2000).



Analiza drgań normalnych (NMA) na potrzeby tej pracy została przeprowadzona w aplikacji webowej Elnémo, wykorzystującej model sieci elastycznej (ENM), z domyślnymi parametrami (Li 2005, Suhre 2004).

Po każdej dokonanej manualnie zmianie w strukturze wykonywano minimalizację energii i krótką symulację dynamiki molekularnej (MD) modyfikowanego fragmentu białka stosując pozycyjne więzy harmoniczne ze stałą siłową  $50.0 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$  na atomach C $\alpha$  dla pozostałej części modelu. Ostateczne wersje modeli kompleksów poddano relaksacji obejmującej symulację MD z zastosowaniem więzów harmonicznych ze stałą siłową  $0.1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$  na atomach C $\alpha$  RNazy III (reszty 1117-1464 pomijając fragmenty: 1140-1145, 1170-1232, 1272-1325 i pętlę 1403-1417) oraz stałą siłową  $1.0 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$  dla atomów szkieletu RNA, w celu dopasowania domen PAZ i platformy do struktury RNA. Symulacje prowadzono w programie CHARMM (Brooks 1983, 2009) oraz w polu siłowym Charmm22/CMAP dedykowanym dla białek (Mackereil 1998, 2004) i polu siłowym Charmm27 dedykowanym dla kwasów nukleinowych (Foloppe 2000, Denning 2011) w modelu ciągłym rozpuszczalnika w uogólnionym modelu Borna, przy użyciu metody GBSW (ang. Generalized Born with a simple SWitching) z domyślnymi parametrami programu CHARMM (Im 2003, Chen 2005). W symulacjach MD w modelu rozpuszczalnika GBSW zastosowano dynamikę Langevina ze współczynnikiem tarcia  $10 \text{ ps}^{-1}$

### 3.2.6. Przygotowanie szkieletu amiRNA.

Użyto komercyjnych plazmidów pENTR4 i pMDC43 firmy Gateway System Invitrogen.

Genomowe DNA (gDNA) z *Arabidopsis thaliana* izolowano buforem MJ: (200 mM TrisHCl (pH 7,5), 250 mM NaCl, 25 mM EDTA, 0,5% SDS). Całkowite RNA z *Nicotiana benthamiana* izolowano metodą TRIzolową (Chomczyński 2006) i rozdzielano elektroforetycznie na 1.5% żelu agarozowym. Barwiono bromkiem etydyny (0.5  $\mu\text{g}/\text{ml}$ ) i analizowano z użyciem systemu dokumentacji żeli Typhoon 8600 (dołączone do niego oprogramowanie ImageQuant służyło do oceny jakościowej RNA.) Syntezę pierwszej nici cDNA przeprowadzano na 1  $\mu\text{g}$  roztworu całkowitego RNA jako matrycy, z użyciem startera oligo dT i enzymu RevertAid M-MuLV Reverse Transcriptase (Fermentas) zgodnie z procedurą

producenta. 1 µl matrycy (gDNA lub cDNA) wykorzystywano do reakcji PCR z 1 U *Taq* DNA polimerazy (Fermentas) w całkowitej objętości 20 µl zgodnie z procedurą producenta.

Do amplifikacji metodą PCR fragmentów intronu użyto następujących starterów wprowadzonych w miejscach rozpoznanych przez enzymy restrykcyjne:

Numer	Nazwa	Sekwencja 5' -> 3'
132	1NcoFor	TTACCATGGGTAAACCAGCTGTAGCTAAA
133	1SalMluR	TTAGTCGACTATACGCGTGAAAATATGGGTTAACAAACAAAGG
134	2SalFor	TAGTCGACACATCTGCTTTTTGAAAAGTTGTC
135	2BamRev	TTAGGATCCTTAAAGGCTATTTTCGAAGTCAGTC
136	3KpnFor	TTAGGTACCATCAAACGAGTCTGCTACTATG
137	3EcoVRev	TTAGATATCCTTCAAGAGCAATATCACGATG
146	GFP6s	GGACGGAACATCCTCGGC

**Tab. 3.2.** Wykaz starterów używanych do reakcji PCR.

Pary starterów, jakich używano do amplifikacji metodą PCR konstruktów ulegającego ekspresji w *N. benthamiana*: 132.1NcoFor z 137.3EcoVRev oraz 134.2SalFor z 137.3EcoVRev

Reakcję PCR prowadzono w następujących warunkach : pierwszy cykl: 94 °C, 1 min; 24 kolejne cykle: 94 °C, 40 sec; 58 °C, 30 sec; 72 °C, 30 sec , wydłużanie końcowe: 72 °C, 5 min. Następnie, 10 µl roztworu reakcyjnego analizowano elektroforetycznie w 0.9% żelu agarozowym barwionym 0.5 µg/ml bromku etydy. Żel skanowano aparatem Typhoon 8600 a oprogramowania ImageQuant użyto do analizy jakościowej.

Ekspresję przejściową *in vivo* prowadzono po inokulacji liści *Nicotiana benthamiana* roztworem *Agrobacterium tumefaciens* transformowanych wektorem ekspresyjnym. Po 3, 5 i 7 dniach izolowano RNA z zainfekowanej tkanki.

## 4. Wyniki i dyskusja

### 4.1. AmiRNA Designer

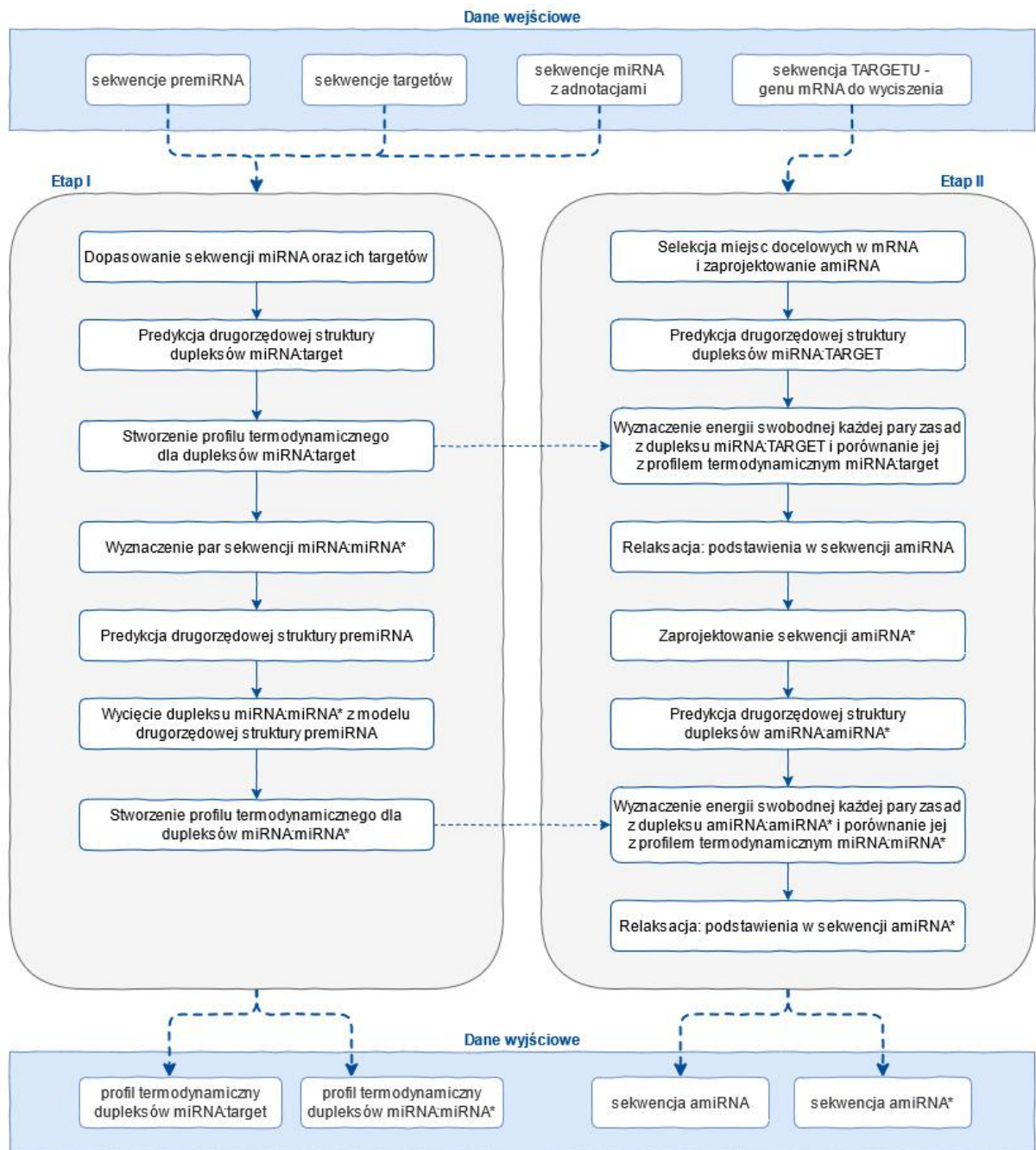
Jednym z pierwszych opisanych sposobów generowania amiRNA dostępnym do dziś jest WMD3 z laboratorium Weigel'a (Schwab 2006). Jest to system dedykowany dla roślin. Opracowano w nim zarówno sposób wyboru sekwencji potencjalnych amiRNA jak i biochemiczne procedury dostarczenia ich do komórek roślinnych. Szkielet, w którym umieszczone zostaje amiRNA przypomina prekursor charakterystyczny dla zwierząt. Sama sekwencja amiRNA ma spełnić następujące kryteria dopasowania do genu docelowego: (i) całkowite dopasowanie na pozycjach 2-12 amiRNA, (ii) 1 lub 2 niedopasowania na 3' końcu amiRNA oraz (iii) całkowita energia hybrydyzacji amiRNA do genu docelowego ma się zawierać w przedziale od -35 do -38 kcal/mol. Kilka innych zespołów zajmujących się bioinformatyką zaproponowało alternatywne sposoby projektowania sekwencji zarówno samego amiRNA jak i regionów je otaczających (Baby 2012, Laganà 2014, Fahlgren 2016). W większości skupiały się one wokół sekwencji amiRNA oraz energii swobodnej jaką wykazuje cały dupleks miRNA:miRNA\* (szerzej opisane w rozdziale Wyniki i Dyskusja). Inną nową aplikacją jest P-SAMS: the Plant Small RNA Maker Suite (Fahlgren 2016). Jej algorytm opiera się na sekwencyjnym podobieństwie do miejsc natywnego oddziaływania miRNA.

Opisano próby wykorzystania sztucznych miRNA do uzyskania odporności na infekcje wirusowe m.in. u pomidora i *Arabidopsis thaliana*. Jednakże wyciszanie – przy użyciu różnych metod projektowania sekwencji wyciszających i odmiennych szkieletów istniejących miRNA – zachodziło z różną, trudną do przewidzenia wydajnością (Ai 2011, Vu 2013, Carbonell 2018). Sugeruje to istnienie jeszcze wielu niejasności w mechanizmie wyciszania genów u roślin, a co więcej potrzebę projektowania sekwencji sztucznych miRNA w oparciu o termodynamiczne parametry, nie z poziomu całych cząsteczek RNA, ale z poziomu niższego poszczególnych par nukleotydowych, zarówno w dupleksie amiRNA z sekwencją docelową (wyciszaną) jak i w prekursorze amiRNA. Badania natywnych sekwencji i struktur prekursorów miRNA oraz analiza regionów flankujących dupleks miRNA:miRNA\* wskazały potencjalne fragmenty, na które warto zwrócić baczniejszą uwagę projektując sztuczne małe regulatorowe

RNA. Szczególnie istotne wydają się same miejsca cięcia w prekursorze (prowadzące do powstania dojrzałego miRNA) oraz odległości do 20 nt przed i za miRNA (Miśkiewicz 2017). Dlatego w ramach niniejszej pracy powstała koncepcja opracowania nowego algorytmu do projektowania amiRNA, uwzględniająca w/w przesłanki.

Efektem niniejszej pracy doktorskiej jest program AmiRNA Designer, którego podstawową funkcją jest projektowanie sekwencji sztucznych miRNA i miRNA\* na podstawie cech natywnych komórkowych miRNA, ich prekursorów i docelowych miejsc oddziaływania (w dalszej części pracy miejsca docelowe będą określane mianem „target” w celu uproszczenia zapisu, zwłaszcza w kontekście dupleksu miRNA:target.) Metoda zaimplementowana w programie działa dwuetapowo. W pierwszym etapie analizuje sekwencje i drugorzędowe struktury istniejących (i) prekursorów miRNA oraz (ii) kompleksów miRNA:target, w celu uzyskania dwóch profili termodynamicznych. Drugi etap obejmuje proces projektowania sekwencji potencjalnych amiRNA dla zadanego genu i sprawdzanie ich zgodności z profilami z pierwszego etapu.

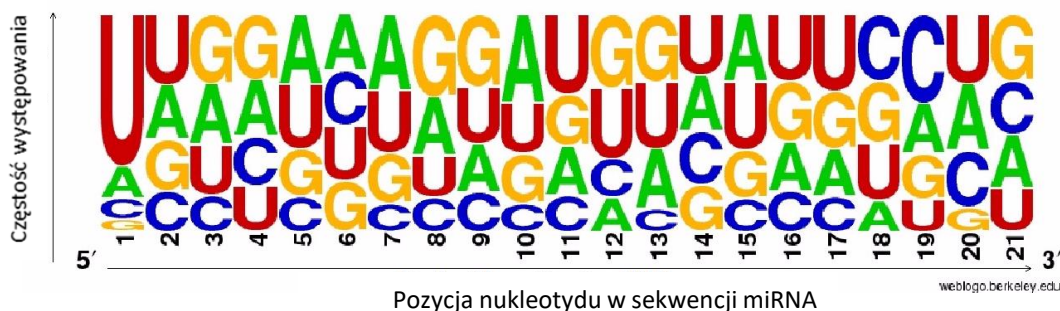
Na koniec każda z wygenerowanych przez program sekwencji jest przyrównywana do transkryptomu *Arabidopsis thaliana* w celu znalezienia potencjalnych niespecyficznych oddziaływań (ang. off-target). Wykorzystywany jest w tym celu skrypt napisany w języku Perl przez dr. Agnieszkę Rybarczyk. (rozdział 4.1.4.3). Schemat działania programu przedstawia Rys. 4.1.



**Rys. 4.1.** Schemat działania programu AmiRNA Designer: Etap I analiza danych. Etap II projektowanie nowych sekwencji.

#### 4.1.1. Analiza sekwencji miRNA

W celu określenia rozkładu nukleotydów w istniejących miRNA posłużono się pulą 157 znanych sekwencji dojrzałych miRNA<sup>1</sup> (Załącznik nr 1), pochodzących z *Arabidopsis thaliana*, wszystkich o długości 21 nt. Uzyskane wyniki wykazały, że u prawie 70% wszystkich miRNA pierwszym nukleotydem w sekwencji jest nukleotyd zawierający jako zasadę azotową uracyl (U1), a u ponad 40% dziesiąty nukleotyd ma adeninę (A10) (Rys. 4.2.). Ponadto, cytydyna na trzynastej pozycji (C13) oraz guanozyna na dwudziestej (G20) występują rzadziej niż wynika to ze statystycznego prawdopodobieństwa obliczonego wg. rozkładu normalnego.



**Rys. 4.2.** Rozkład nukleotydów na poszczególnych pozycjach w 21-nt miRNA z *A. thaliana*.

Na podstawie uzyskanych danych statystycznych, zaproponowane zostały cztery pozycje w sekwencji targetu, zawierające (lub nie) określone nukleotydy. W związku z tym, że sekwencje targetów są odwrotnie komplementarne do oddziałujących z nimi sekwencji miRNA określenie ich pozycji wymaga przeliczenia za pomocą wzoru: **pozycja w targecie = 22 – pozycja w miRNA**. Pozycje ustalonych nukleotydów stanowią pierwsze sito do filtrowania fragmentów potencjalnych miejsc docelowych w zadanym genie. Powyższa analiza dla sekwencji miRNA z *Arabidopsis thaliana* pozwoliła na określenie następujących filtrów dla potencjalnych miejsc docelowych wycisanego transkryptu:

1. obecność A w pozycji 21
2. obecność U w pozycji 12
3. nieobecność G w pozycji 9
4. nieobecność C w pozycji 2

<sup>1</sup> Pochodzenie wszystkich sekwencji opisano w rozdziale Materiały i Metody, zaś same sekwencje znajdują się w rozdziale Załączniki na końcu niniejszej rozprawy.

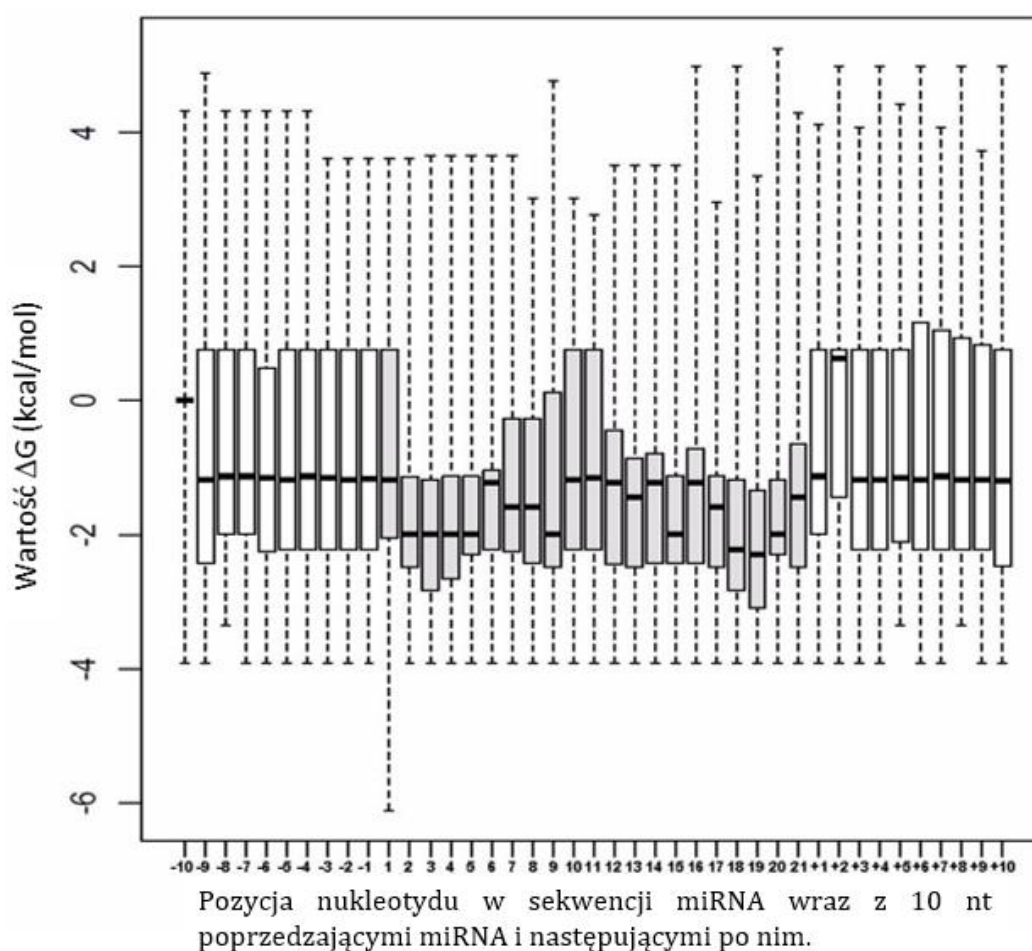
Parametry te są opcjonalne i użytkownik może zdecydować, które z nich będą stosowane, lub też zdefiniować własne odmienne od wyżej wymienionych, oparte na analizie miRNA pochodzących z organizmu dla którego mają zostać zaprojektowane nowe sekwencje amiRNA i amiRNA\*. Specyfikacja może także dotyczyć określonej tkanki, stadium rozwoju w jakim konkretne miRNA ulegają ekspresji lub dowolnego zestawu natywnych miRNA według wyboru użytkownika.

#### 4.1.2. Struktura drugorzędowa dupleksu miRNA:target

Do wyznaczenia pierwszego profilu termodynamicznego przez program AmiRNA Designer wykorzystano sekwencje oddziałujących ze sobą miejsc docelowych (target) oraz miRNA dla *A. thaliana*, dostępnych w bazie danych ASRP (Gustafson 2005). Zawarte są tam zarówno pary sekwencji miRNA:target, których oddziaływanie potwierdzono eksperymentalnie bezpośrednio u *A. thaliana* jak również te, przewidziane komputerowo na podstawie potwierdzonej homologii do par pochodzących z innych roślin. Ze wszystkich zdeponowanych tam par sekwencji do eksperymentu obliczeniowego wybrano miRNA o długości 21 nukleotydów, zatem tylko takiej długości sekwencje zostały wygenerowane przez program. Długość sekwencji miRNA jest parametrem ustawianym przez użytkownika programu - można wybrać dowolną długość fragmentu z zakresu 10-50 nt. Dodatkowo, aby uniknąć nadreprezentacji dużych rodzin miRNA oraz docelowych transkryptów, wyselekcjonowano jedynie unikalne (nieredundantne) sekwencje miRNA i targetów. Ostatecznie z 570 par podane kryteria spełniało 210 i one właśnie zostały wykorzystane do stworzenia profilu termodynamicznego miRNA:target.

Głównym parametrem w programie AmiRNA Designer, służącym do opisanie dupleksów miRNA:target jest profil termodynamiczny. W celu uzyskania danych termodynamicznych, niezbędnych do dalszej pracy, wykonywana jest predykcja drugorzędowych struktur dupleksów miRNA:target za pomocą programu UNAFold. Następnie analizowany jest rozkład struktury 2D na poszczególne elementy wyróżnione w modelu NN, dzięki czemu każdej parze przypisuje się zmiany energii swobodnej ( $\Delta G$ ). Jako, że  $\Delta G$  jest określona dla elementów zamkniętych przez pary zasad należało opracować sposób przypisywania wartości do każdej pozycji niezależnie od tego czy występuje w niej sparowanie,

czy też jest to pętla wewnętrzna lub wybrzuszenie jednostronne. Obliczone wartości  $\Delta G$  na każdej pozycji miRNA dla wszystkich analizowanych struktur – wraz z rozkładem statystycznym przedstawionym jako wykres pudełkowy – stanowią referencyjną bazę dla nowych sekwencji projektowanych w dalszych etapach działania programu, zwaną dalej „profilem termodynamicznym miRNA:target” (Rys. 4.3.).



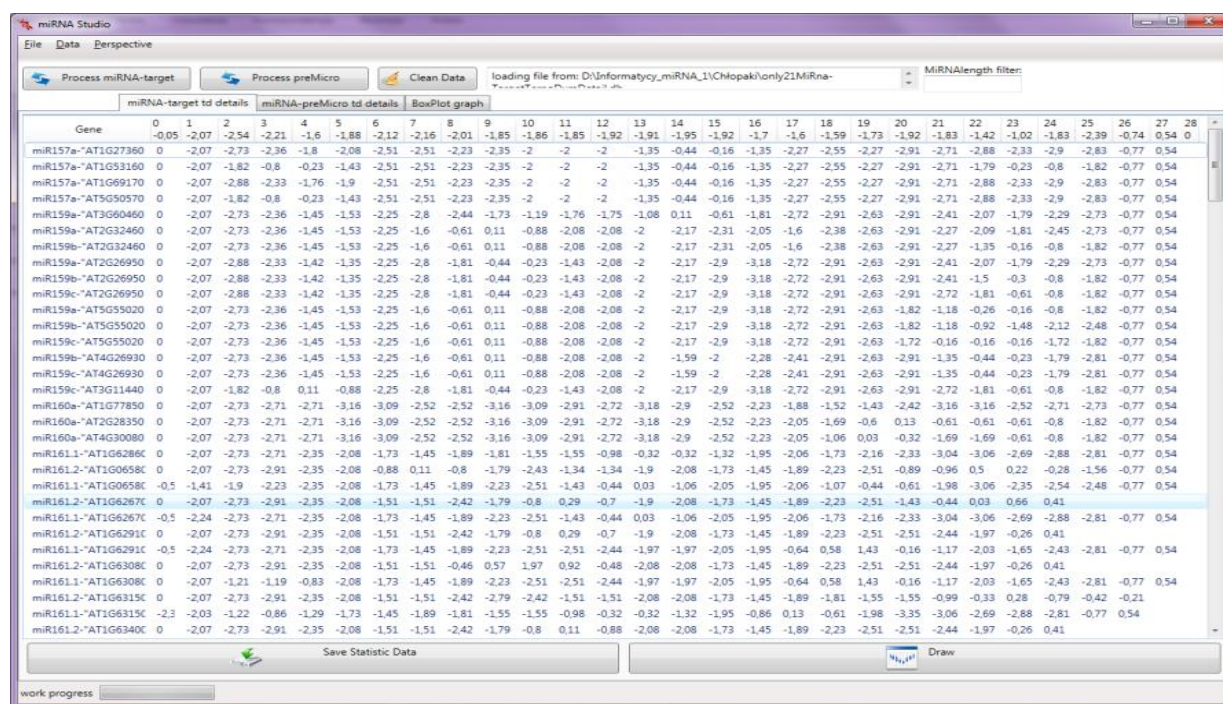
**Rys. 4.3.** Wykres profilu termodynamicznego dla dupletu miRNA:target wygenerowany przez program AmiRNA Designer dla analizowanego zestawu sekwencji z *Arabidopsis thaliana*. Ciemnoszarym kolorem oznaczono 11 nt miRNA, za którym następuje przecięcie sekwencji targetu.

Profile termodynamiczne przedstawiono na wykresie pudełkowym (ang. boxplot). Górna i dolna ścianka pudełka stanowią odpowiednio pierwszy i trzeci kwartył. Wewnątrz pudełka zaznaczono medianę. Końce wąsów odnoszą się do wartości minimalnych i maksymalnych.



Otrzymany profil wykazuje kilka istotnych cech, które zostały wykorzystane w dalszych etapach. Najistotniejszy wydaje się być centralny region miRNA, pomiędzy 10 a 11 nukleotydem, gdzie zachodzi cięcie transkryptu przez białko AGO1. W tym miejscu profil termodynamiczny wykazuje lokalne minimum  $\Delta G$ , co oznacza silniejsze oddziaływanie cząsteczek miRNA i targetu w środkowym regionie. Dodatkowo można wyróżnić jeszcze dwa lokalne minima położone odpowiednio w okolicy 3-4 nukleotydu miRNA oraz 16-17 nukleotydu. Ponadto, gdy rozpatrzmy uśrednioną energię swobodną sumaryczną dupletu dla połowy z końca 5' (fragment odpowiadający 1-10 nt miRNA) oraz osobno dla połowy z końca 3' (fragment odpowiadający 12-21 nt miRNA), widoczna jest między nimi różnica w intensywności oddziaływania. I tak odpowiednio wartość  $\Delta G$  dla połowy 5' wynosi -2,04 kcal/mol, natomiast dla połowy 3' więcej, bo -1,83 kcal/mol. Może to sugerować niewielką analogię do zwierzęcych miRNA, gdzie obserwowany jest region "seed" ok. 7 nt z 5' końca miRNA o dużym sparowaniu podczas gdy reszta miRNA wykazuje znacznie słabszą komplementarność do docelowej sekwencji transkryptu (Lim 2003).

Poniżej przedstawiono zrzut ekranu z programu AmiRNA Designer z przeliczonymi wartościami  $\Delta G$  przyporządkowanymi do każdej pozycji miRNA dla dupletu miRNA:target (Rys. 4.4.). W opisie zawarte są identyfikatory miRNA oraz numery genów z którymi tworzą parę.



Rys. 4.4. Wartości  $\Delta G$  przeliczone w programie AmiRNA Designer dla dupletu miRNA:target.

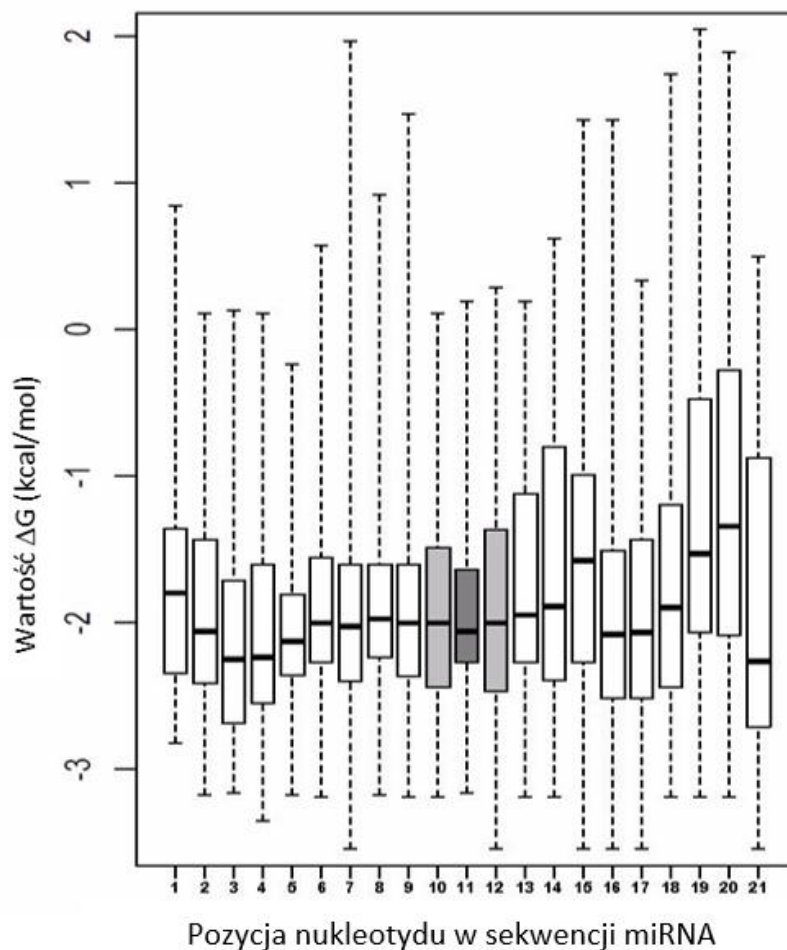
### 4.1.3. Struktura drugorzędowa prekursora miRNA

Do otrzymania drugiego profilu – analogicznego do opisanego w poprzednim rozdziale – wykorzystano sekwencje prekursorów miRNA dostępnych w bazie danych MiRBase (Griffiths-Jones 2008). Zdeponowane tam sekwencje nie stanowią pre-miRNA (charakteryzującego się tym, że jeden z jego końców (3' lub 5') jest także końcem miRNA), a często są przedłużone o kilka-kilkaset nukleotydów z obu końców pre-miRNA. Nie są to jednakże pełne sekwencje pri-miRNA – które stanowią dojrzały transkrypt zaczynający się czapczką na 5' końcu i ogonem poliadenylowym na 3' końcu. Predykcji struktury 2D dokonywano dla całej dostępnej sekwencji prekursora miRNA, a następnie w dalszej analizie uwzględniany był fragment zawierający miRNA:miRNA\* przedłużony o ok. 10 par nukleotydów poprzedzających miRNA oraz następujących za nim. Podobnie jak w poprzednim etapie, zestaw sekwencji dla naszego eksperymentu został ograniczony do prekursorów, z których pochodzą miRNA o długości 21 nukleotydów, i składał się ze 157 dupleksów miRNA:miRNA\* pochodzących z *A. thaliana*. Jednakże program AmiRNA Designer umożliwia analizę fragmentów o długości od 10 do 50 nukleotydów, a użytkownik może zdecydować czy chce ograniczyć się do jednej długości, czy przeanalizować ich kilka.

Otrzymany profil termodynamiczny wykazuje cechy odmienne do otrzymanego w poprzednim kroku profilu miRNA:target. Podobnie jak to zostało wykazane w badaniach dla innych organizmów zaobserwowano wyższą średnią wartość  $\Delta G$  dla pierwszej pozycji miRNA aniżeli dla ostatniej (Khvorova 2003, Krol 2004). Sugeruje to mniejszą stabilność 5' końca miRNA, co uważa się za cechę umożliwiającą odróżnienie nici miRNA od miRNA\* i pozostawienie właściwej cząsteczki w kompleksie efektorowym RISC. W odróżnieniu od profilu miRNA:target, nie występuje tu w części centralnej lokalne minimum, ale wartość energii swobodnej trzech środkowych pozycji jest wyższa niż sąsiednich. Ponadto lokalne minima występują na obu końcach miRNA.

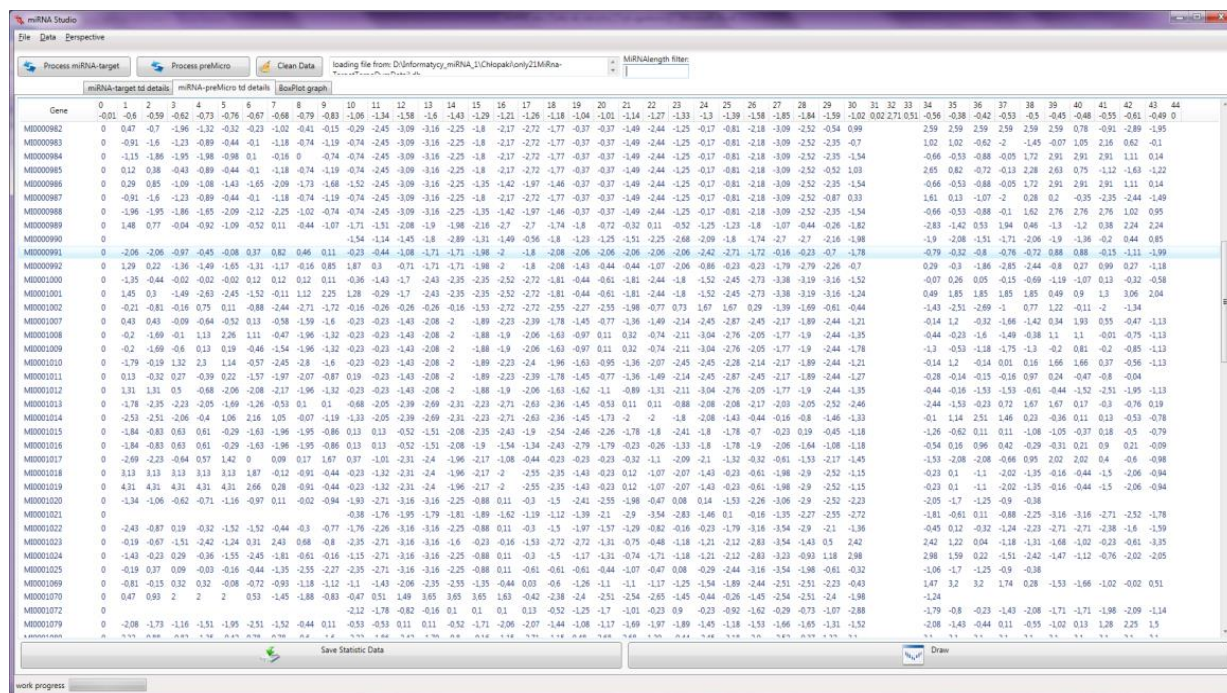
Obliczone wartości  $\Delta G$  dla wszystkich dostępnych struktur – wraz z rozkładem statystycznym przedstawionym jako wykres pudełkowy – stanowią referencyjną bazę dla nowych sekwencji projektowanych w dalszych etapach działania programu AmiRNA Designer, zwaną dalej „profilem termodynamicznym prekursora”. Rys. 4.5 przedstawia wykres pudełkowy wygenerowany przez program AmiRNA Designer zawierający wartości  $\Delta G$

przyporządkowane do każdej pozycji miRNA dla pre-miRNA (Rys. 4.5.). Górna i dolna ścianka pudełka stanowią odpowiednio pierwszy i trzeci kwartyl. Wewnątrz pudełka zaznaczono medianę. Końce wąsów odnoszą się do wartości minimalnych i maksymalnych.



**Rys. 4.5.** Wykres profilu termodynamicznego dla prekursora miRNA wygenerowany przez program AmiRNA Designer dla analizowanego zestawu sekwencji z *Arabidopsis thaliana*. Szarym kolorem zaznaczono pozycje 1-21 w sekwencji miRNA.

Poniżej przedstawiono zrzut ekranu z programu AmiRNA Designer zawierający przeliczone wartości  $\Delta G$  przyporządkowane do każdej pozycji miRNA dla struktury prekursora miRNA (Rys. 4.6.).

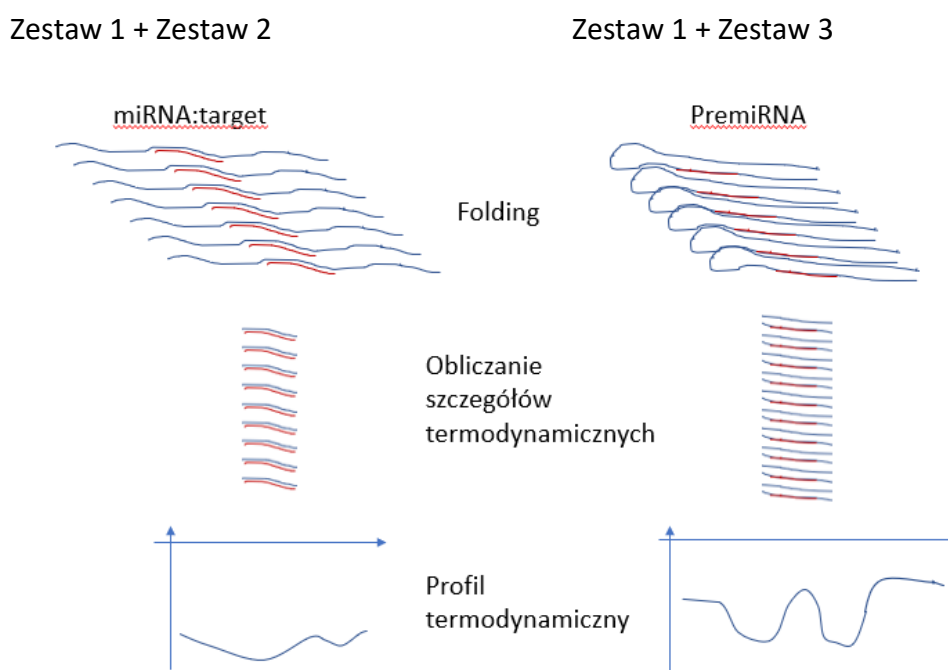


**Rys. 4.6.** Wartości  $\Delta G$  przeliczone w programie AmiRNA Designer dla prekursora miRNA w analizowanym zakresie -10 nt, miRNA, +10 nt. Pozycje 22-24 nie zawierają wartości ponieważ w analizowanym zestawie uwzględniono wyłącznie miRNA o długości 21 nt.

#### 4.1.4. Opis działania programu

##### 4.1.4.1. Etap I Analiza danych

Pliki wejściowe są przetwarzane w następujących parach: zbiór danych 1 (Tab. 4.1) ze zbiorem danych 2 (Tab. 4.2), w celu stworzenia profilu termodynamicznego miRNA:target, zaś zbiór danych 1 ze zbiorem danych 3 (Tab. 4.3), do utworzenia profilu termodynamicznego prekursora. Proces przedstawiono graficznie na Rys. 4.7.



**Rys. 4.7.**Schemat przedstawiający działanie Etapu I AmiRNA Designer.

W pierwszej parze (zestaw 1 + zestaw 2) identyfikowane są miRNA występujące w obu plikach: w pliku z zestawu 1 nazwy miRNA wyszukiwane są w kolumnie "ID", w pliku z zestawu 2 przeszukiwana jest kolumna „Locus”. Znaleziona sekwencja miRNA i jej target są łączone w parę oraz wydłużane o sekwencje stabilizujące GCG i CGC na obu końcach - zapobiega to hybrydyzacji miRNA:target w inny niż natywny sposób oraz ich przesunięciu się względem siebie podczas predykcji struktury drugorzędowej. Zbiór wszystkich dopasowanych par miRNA:target stanowi dane wejściowe narzędzia AmiRNA Designer. W kolejnym kroku program UNAFold, zintegrowany z AmiRNA Designerem, przewiduje drugorzędową strukturę duplexu miRNA:target oraz wyznacza parametry energetyczne przewidzianego modelu cząsteczki. W skład UNAFolda wchodzi szereg skryptów, które są uruchamiane sekwencyjnie za pomocą następujących poleceń:

```
hybrid-2s.pl, "--temperature=22 seq1 seq2"
```

```
ct-energy.exe, "--verbose --temperature=22 seq1-seq2.ct"
```

```
ct-energy-det.pl
```

Skrypt **hybrid-2s.pl** oblicza minimalną energię hybrydyzacji dwóch sekwencji wejściowych zapisanych w dwóch oddzielnych plikach w formacie FASTA. Dodatkowym

argumentem wejściowym jest temperatura, która stanowi punkt odniesienia dla energii hybrydyzacji. Skrypty **ct-energy.exe** oraz **ct-energy-det.pl** służą do wyznaczania energii wiązań. Zwracają one dane wyjściowe w formacie ASCII lub HTML. Otrzymane dane wyjściowe zostają przekazane do parsera, który transformuje kolejne elementy sparowanej sekwencji zgodnie z przyjętym w algorytmie modelem teoretycznym.

W drugiej parze plików (zestaw 1 + zestaw 3) sekwencje z pliku microRNA są dobierane na podstawie dopasowania danych z kolumny „Accession” z zestawu 1 z kolumną „miRBase accession number” z pliku z zestawu 3. Umożliwia to precyzyjną lokalizację miRNA w jego prekursorze. Komputerowa predykcja struktury drugorzędowej (program UNAFold) przeprowadzona jest dla całej dostępnej sekwencji prekursora miRNA. Następnie z otrzymanej struktury wycinany jest i poddawany dalszej analizie fragment zawierający miRNA, przedłużony o 10 nt (lub maksymalną liczbę nukleotydów, jaka jest dostępna w sekwencji z bazy danych) przed i za sekwencją miRNA. Tak przedłużony fragment miRNA jest zazwyczaj sparowany w dużym stopniu i rzadko występują w nim duże zaburzenia - jak wielonukleotydowe wybrzuszenia, czy rozgałęzienia, które często znajdują się w całkowitej strukturze prekursora.

Przetwarzanie danych z tego zestawu przebiega sekwencyjnie. Każda sekwencja z listy wejściowej jest zapisywana do pliku. Następnie uruchamiane są trzy skrypty programu UNAFold w kolejności: (1) **hybrid-ss-2s.pl**, (2) **ct-energy.exe**, (3) **ct-energy-det.pl**:

```
hybrid-ss-2s.pl, "--temperature=22 seq1"
```

```
ct-energy.exe, "--verbose --temperature=22 seq1.ct"
```

```
ct-energy-det.pl
```

Skrypt **hybrid-ss-2s.pl** przyjmuje jako argument wejściowy jedną sekwencję, którą jest cała dostępna sekwencja prekursora, a nie tylko 21-elementowy dupleks. Skrypty **ct-energy.exe** oraz **ct-energy-det.pl** działają tak samo jak dla pary zestaw 1 + zestaw 2.

Powyższa procedura uzasadniona jest tym, że predykcja struktury wyłącznie fragmentu sekwencji, mogłaby prowadzić do przesunięć i uzyskania szczegółów termodynamicznych dla innych dupleksów niż natywne.

Tab. 4.1. Struktura zestawu danych nr 1.

Accession	ID	Sequence	Mature1_Acc	Mature1_ID	Mature1_Seq	Mature2_Acc	Mature2_ID	Mature2_Seq
MI0005385	miR830	CAGAGUUCUGUCUAGUGUCUUUACUAAACACAAAUCUGUAAACGCCUCGAUCCUUCUCCAAUAGUUUAGGUUAGCUGACAUAAUAAUUAUUAAGACCCAGUUUCCAGAUAAACUAAUCUAAUUUGAAGAAAGUGAACGAGAGCGAUUACAAAGUUAAGGAGAGUUGAAUUUUAAGAACAUCC	MIMAT0004247	miR830*	UCUUUCUCAAUAGUUUAGGUU	MIMAT0004248	miR830	UAACUAAUUUUGAGAAGAAGUG
MI0005387	miR831	GCACAAUCCGACAGUCUACAAGAAGAGAUUUUAGAGUCUUAAGAGUCUCAAAGCGUACAAGGAGUAGGAGUGAAUCAAAAACAAGUUGUUUGGUUUUGUAAUCCAAAACAAGUGUUUUGGUUUUGGAAUCAGUUGUUUACCCAAAACACUUGUUUGGUUUUUAUCUUGAUUCUUCGUACUCUUCUGUACUCUUCUGGAUUCUUAUGACACUCUACAUGUAACUCCGCAUGGUUUUACCAAAGCCU	MIMAT0004249	miR831	UGAUUCUCUUGUACUCUUCUUG			
MI0005388	miR832	AAGAAAGAAAAAUAAGUGACAAGUACAACAACUAGAGAAAGUGAAACUUGCGGGAUCGGAUUGCAAGAUCCGAAAGCCAGUCUUUAGAGGAAUUCUGAACGGGCUAGUAAAGUAAACUUGUGAUUGACCAUUAUUGCAUUAUGAAUUAAGUUGGUGUGUGUCAUCACAGGUGCUAAGUACUUCACUUUUUUUUAACGAAAGCAUCACCCGUUUUAGCAUCAUCUUCUUGAAGCUUUUUUUUGCCAUUUAACAAAUUACCAAUUUUG	MIMAT0004250	miR832-5p	UGCUGGGAUCGGAAUCGAAA	MIMAT0004251	miR832-3p	UUGAUUCCCAAUCCAAGCAAG
MI0005389	miR833	GGACGUAGUUGUACUCUACAGAUGGUGUAGUCUGUAAAACAACGGAGGAGUUGUUUUGUACUCGGUCUAGUCAAACAAAACUUUUGCAAUUAUGUAUGUUGUAGCUAGACCGAUUACAACAACAGCUCUUGUUUGUAUUAACCAUCAACUUAAGAAACCGCACCACUC	MIMAT0004252	miR833-5p	UGUUUGUUGUACUCGGUCUAGU	MIMAT0004253	miR833-3p	UAGACCGAUUGCAACAACAAG
MI0005390	miR834	UACAACAUUAGUUAACAGUUGUCUAUCAUUUUUUCAGCGACAGCCAAAACACCACCGUUCUUCUACGAACAUUUAAGCCAUCCUUUAACCGUUGCUCACCAUUAUACGGCAUCACAAGA AAAUCAUGCAAGUGGAAACGGUAGUAGUAGCAGCGUGUUAU GACACGUUGUAGCAGUAGCGUGUUAUUGCCGAUAGCGUG AGCGGCAUUGAAGCAUUGGUGUUCAGAAAGCAUGUA	MIMAT0004254	miR834	UGGUAGCAGUAGCGUGGUA			
MI0005391	miR835	UGCUCGAUUCUUCAGAUAGAAAUUCUGUCUUCUUGUUGUUAAGCCUUAUUAAGAACUCUUCUUAUUAAGCUGUUUACCAACAUAUAACUUCUGUUGUAGUACUUCUCUGCUCACCAUCCACAAAGCUUAUCAUGCAAGAUUAAACCCUUUCUGAU GGGUUGCAAGAGAACACUAAAAUAGAAAGAAUAUCUUGG GGGAAACGUCACAAUCCAGGUUAUAGAGAAUAGGAUGCU GAUCGUUGAAAUAACUUAUAAGAUAGGAGAAUAGCC AAGAAAGAGUUUGUAACACAAGC	MIMAT0004255	miR835-5p	UUCUUGCAUUAUUCUUUAUC	MIMAT0004256	miR835-3p	UGGAGAAGAUACCGAAGAAAG
---	---	---	---	---	---	---	---	---

Informacja, zawarta w zestawie danych nr 3, o tym czy miRNA znajduje się w 3' czy też w 5' ramieniu prekursora ma szczególne znaczenie, ponieważ we wszystkich etapach analizy sekwencje są przetwarzane od 5' do 3' końca miRNA. W przypadku miRNA pochodzących z ramienia 3' prekursora analiza całej sekwencji w kierunku 5' do 3' prowadziłaby do ustalenia parametrów dla miRNA\* zamiast właściwego miRNA.

Tab. 4.2. Struktura zestawu danych nr 2.

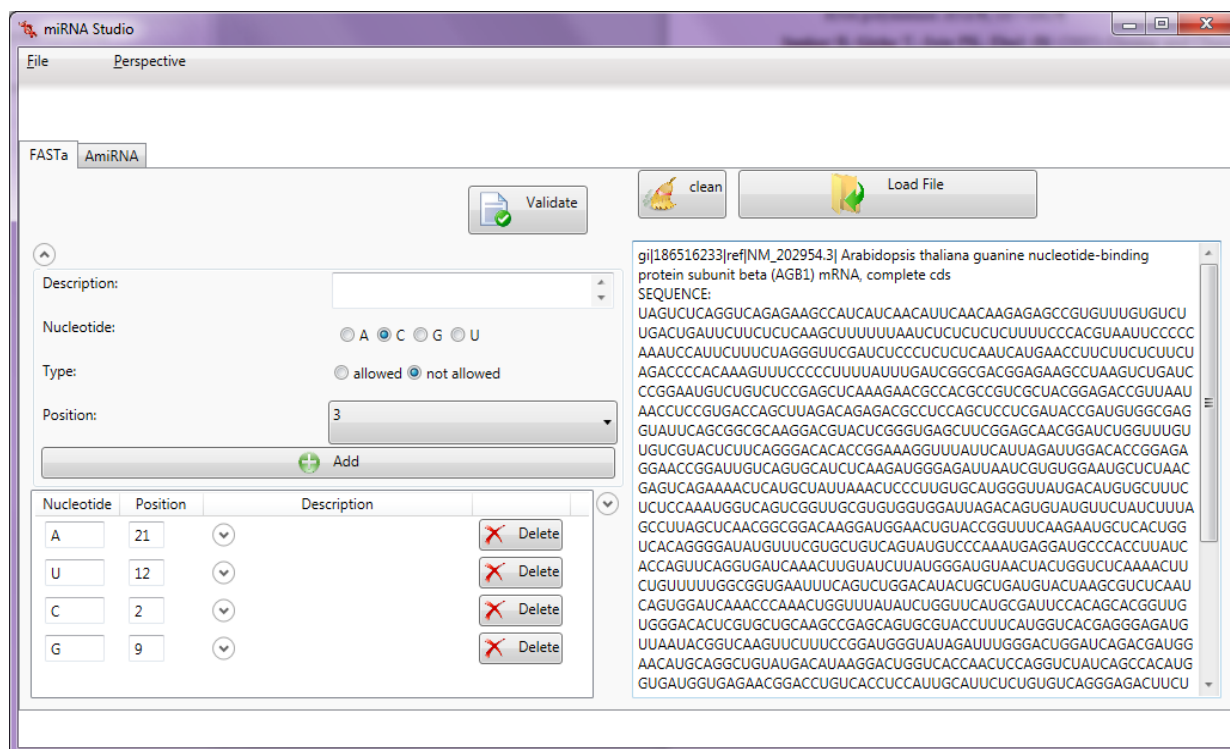
Gene	Gene Name	Locus	Sequence	Transcript Start	Transcript End	Chromosome	Genome Start	Genome End
AT2G33810	SPL3	miR156a,b,c,d,e,f,g	UUGCUUACUCUCUUCUGUCA	786	805	2	14312953	14312972
AT1G27360		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCA	1252	1272	1	9503604	9503624
AT1G27370		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCA	2367	2387	1	9508046	9508066
AT1G53160	SPL4	miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	CUGCUCUCUCUCUUCUGUCAU	591	611	1	19810757	19810777
AT1G69170		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCAG	1266	1286	1	26010393	26010413
AT2G42200		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCA	936	956	2	17596143	17596163
AT3G57920		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCA	784	804	3	21456419	21456439
AT5G43270	SPL2	miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCA	1185	1205	5	17380486	17380506
AT5G50570		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCAU	1035	1055	5	20601044	20601064
AT5G50670		miR156a,b,c,d,e,f,g,h:miR157a,b,c,d	GUGCUCUCUCUCUUCUGUCAU	1024	1044	5	20634654	20634674
AT3G15270	SPL5	miR156a,b,c,d,e,f,g:miR157d	CCGCUCUCUCUCUUCUGUCA	642	661	3	51411124	51411143
---	---	---	---	---	---	---	---	---

Tab. 4.3. Struktura zestawu danych nr 3.

miRbase accession number	Mature miRNA sequence	Precursor start	Precursor end	Polarity	Length of miRNA
MI0000184	UUGACAGAAGAUAGAGAGCAC	4	24	5prim	21
MI0000185	UUGACAGAAGAUAGAGAGCAC	19	39	5prim	21
MI0000186	UUGACAGAAGAUAGAGAGCAC	24	44	5prim	21
MI0000189	UUUGGAUUGAAGGGAGCUCUA	163	183	3prim	21
MI0000190	UGCCUGGCUCCCUGUAUGCCA	4	24	5prim	21
MI0000191	UGCCUGGCUCCCUGUAUGCCA	5	25	5prim	21
MI0000192	UGCCUGGCUCCCUGUAUGCCA	5	25	5prim	21
MI0000193	UCAAUGCAUUGAAAGUGACUA	38	58	5prim	21
MI0000193_47_67	UGAAAGUGACUACAUCCGGGU	47	67	5prim	21
MI0000194	UCGAUAAACCUCUGCAUCCAG	89	109	3prim	21
MI0000195	UCGAUAAACCUCUGCAUCCAG	79	99	3prim	21
MI0000197	UGGAGAAGCAGGGCAGUGCA	19	39	5prim	21
MI0000198	UGGAGAAGCAGGGCAGUGCA	3	23	5prim	21
MI0000199	UCGGACCAGGCUUCAUCCCC	86	106	3prim	21
MI0000200	UCGGACCAGGCUUCAUCCCC	137	157	3prim	21
...	...	...	...	...	...

W eksperymencie obliczeniowym wykonanym w programie AmiRNA Designer na potrzeby tej pracy wykorzystano zestawy danych jak opisano powyżej. Użytkownik może jednakże wprowadzić własne dane pochodzące z innego organizmu. Zachowanie formatu wszystkich plików wejściowych jest konieczne do tworzenia nowych profili termodynamicznych.

Poniżej przedstawiono zrzut ekranu z programu AmiRNA Designer (Rys. 4.8.).

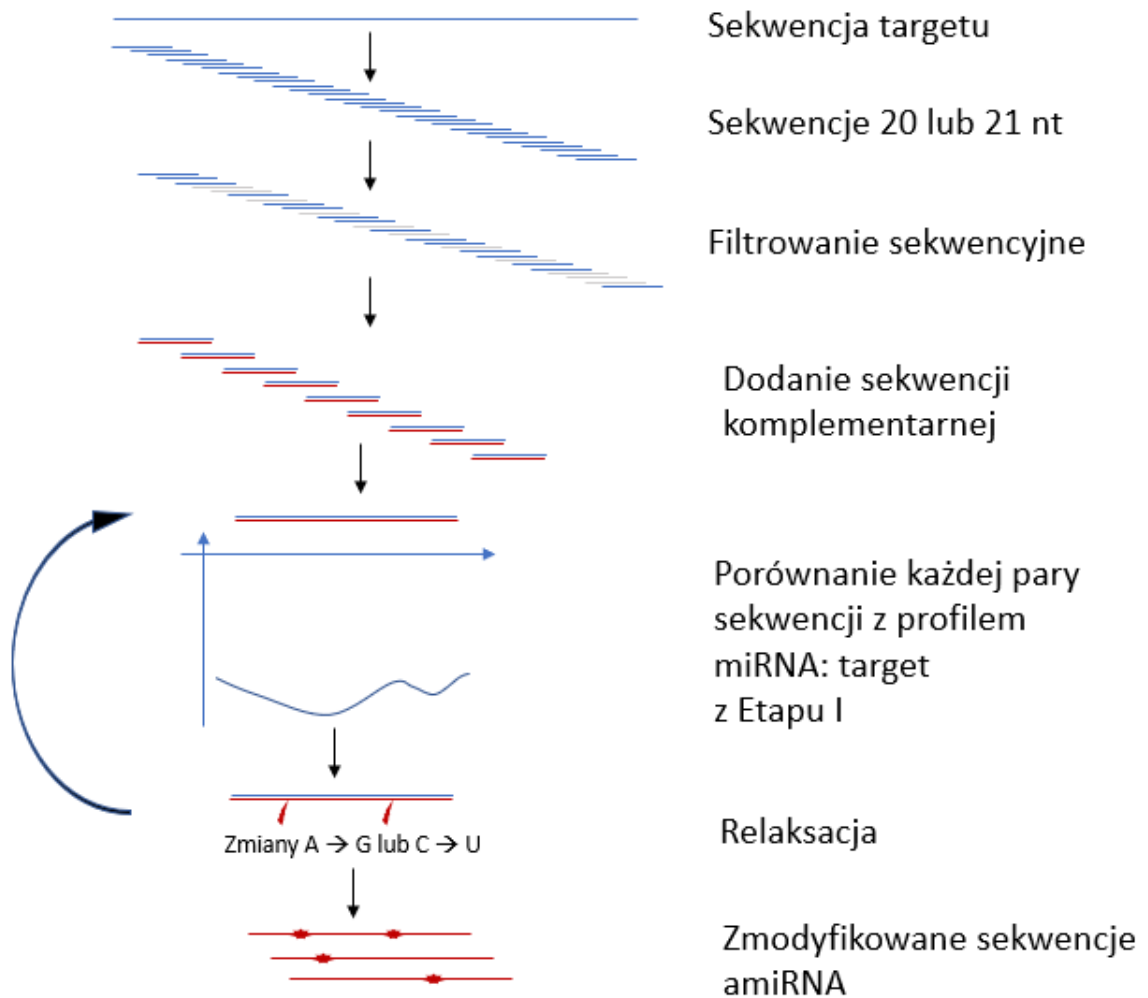


Rys. 4.8. Zrzut ekranu z programu AmiRNA Designer. Parametry sekwencyjne oraz sekwencja wyciskanego genu, dla której zaprojektowane zostały amiRNA i amiRNA\*..



#### 4.1.4.2. Etap II Projektowanie amiRNA

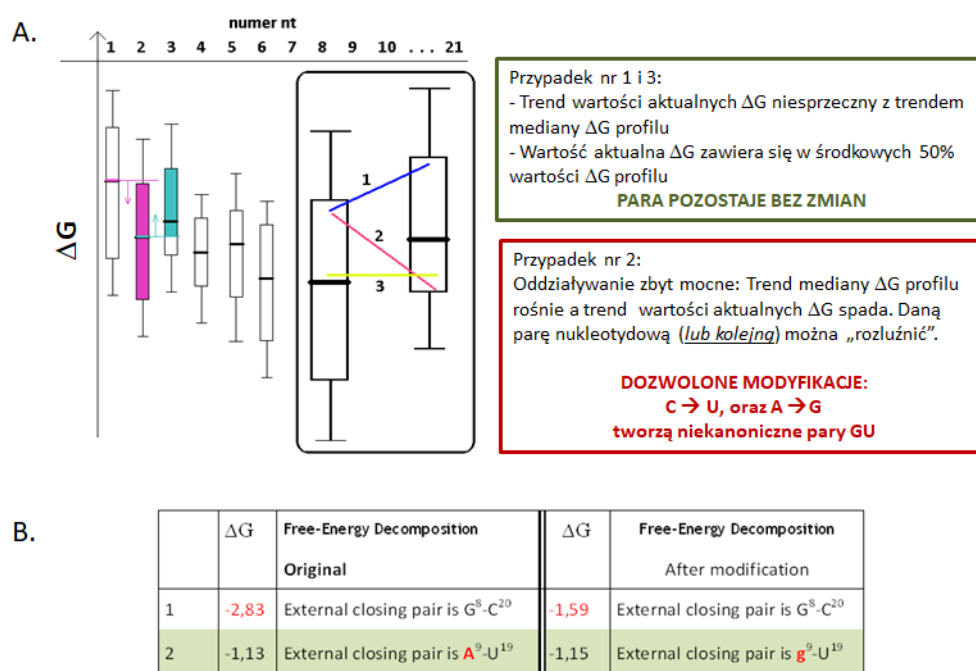
W drugim etapie – projektowania amiRNA – do wyciszenia docelowego genu, sekwencja transkryptu zostaje przedstawiona w postaci fragmentów o długości 21 lub 20 nukleotydów, pokrywających ją 20-krotnie z przesunięciem o jeden nukleotyd. Sekwencje te są traktowane jako potencjalne miejsca docelowe dla amiRNA. Ich liczba może być zmniejszona przez zastosowanie filtrów sekwencji uzyskanych w etapie statystycznej analizy dostępnych sekwencji miRNA. Jednakże nie jest zalecane stosowanie zbyt wielu lub zbyt restrykcyjnych filtrów, gdyż z bardzo niewielkiej puli sekwencji trudno jest otrzymać wystarczającą ilość wyników (przekraczającą 3 amiRNA). Kolejnym krokiem jest wygenerowanie sekwencji odwrotnie komplementarnej dla każdego z pozostałych po filtrowaniu fragmentów genu docelowego, w celu uzyskania pierwotnej sekwencji amiRNA. Zazwyczaj amiRNA są tak zaprojektowane tak by U znajdowało się na pozycji 1. Użytkownik może wybrać filtr A21 w miejscu docelowym lub skrócić je do długości 20 nukleotydów a w zamian dodać nukleotyd A na końcu. Wprowadza to U na pierwszej pozycji amiRNA (opcja U-force). Reguły programu pozwalają ustawić dowolny nukleotyd na każdej, wybranej pozycji amiRNA jako wymuszony lub zabroniony. Następnie przeprowadzana jest predykcja struktury drugorzędowej dupleksu, z której dokonany zostaje rozkład na parametry termodynamiczne modelu NN. Wartości energii swobodnej zostają przeliczone zgodnie z ustalonymi zasadami identycznie jak przy generowaniu profilu termodynamicznego na etapie analizy danych wejściowych. Wyniki porównywane są z profilem termodynamicznym miRNA:target wygenerowanym poprzednio. Proces przedstawiono graficznie na Rys. 4.9.



**Rys. 4.9.** Ilustracja etapu projektowania amiRNA.

Następujące warunki muszą być spełnione by pojedyncza sekwencja amiRNA została akceptowana: (i) wszystkie wartości  $\Delta G$  na każdej pozycji muszą zawierać się w zakresie pierwszego i trzeciego kwartyla; (ii) zmiana aktualnej wartości energetycznej nie jest ujemna podczas gdy trend mediany profilu na odpowiadających pozycjach jest rosnący (lub też nie jest dodatnia gdy trend mediany profilu jest malejący); (iii) jeżeli energia swobodna co najmniej jednej pary zasad analizowanego dupleksu przekracza trzeci kwartyl, sekwencja jest odrzucana. Zdarza się, że na niektórych pozycjach w analizowanej strukturze wiązanie jest zbyt silne w stosunku do profil - wówczas trend mediany profilu rośnie, jednak aktualna wartość energii  $\Delta G$  jest zbyt niska (Rys. 4.10.A). W takich przypadkach nukleotyd projektowanego amiRNA na tej pozycji powinien być wymieniony. Stworzyliśmy termin "relaksacja", aby opisać tę sytuację. Relaksacja pozwala na tworzenie par GU, które często występują w strukturach RNA. Modyfikacja kanonicznej pary zasad do pary GU powoduje mniejszą destabilizacją

dupleksu RNA niż wprowadzenie innych par niekanonicznych – uznawanych tu za niedopasowanie (np. UU). Zatem dozwolone substytucje w sekwencji amiRNA, które mogą być wprowadzane są następujące:  $A \rightarrow G$  i  $C \rightarrow U$ , co powoduje odpowiednio powstanie par GU i UG w duplesie amiRNA:target. Jeśli analizowaną parą jest GC lub UA i zmiana nie jest możliwa, to sprawdzane jest, czy w następującej po niej parze zasad może zostać wprowadzona relaksacja energii (Rys. 4.10.B). Zarówno liczba dozwolonych par GU (domyślnie 3), jak i minimalna odległość między nimi (domyślnie 5) są definiowane przez użytkownika.

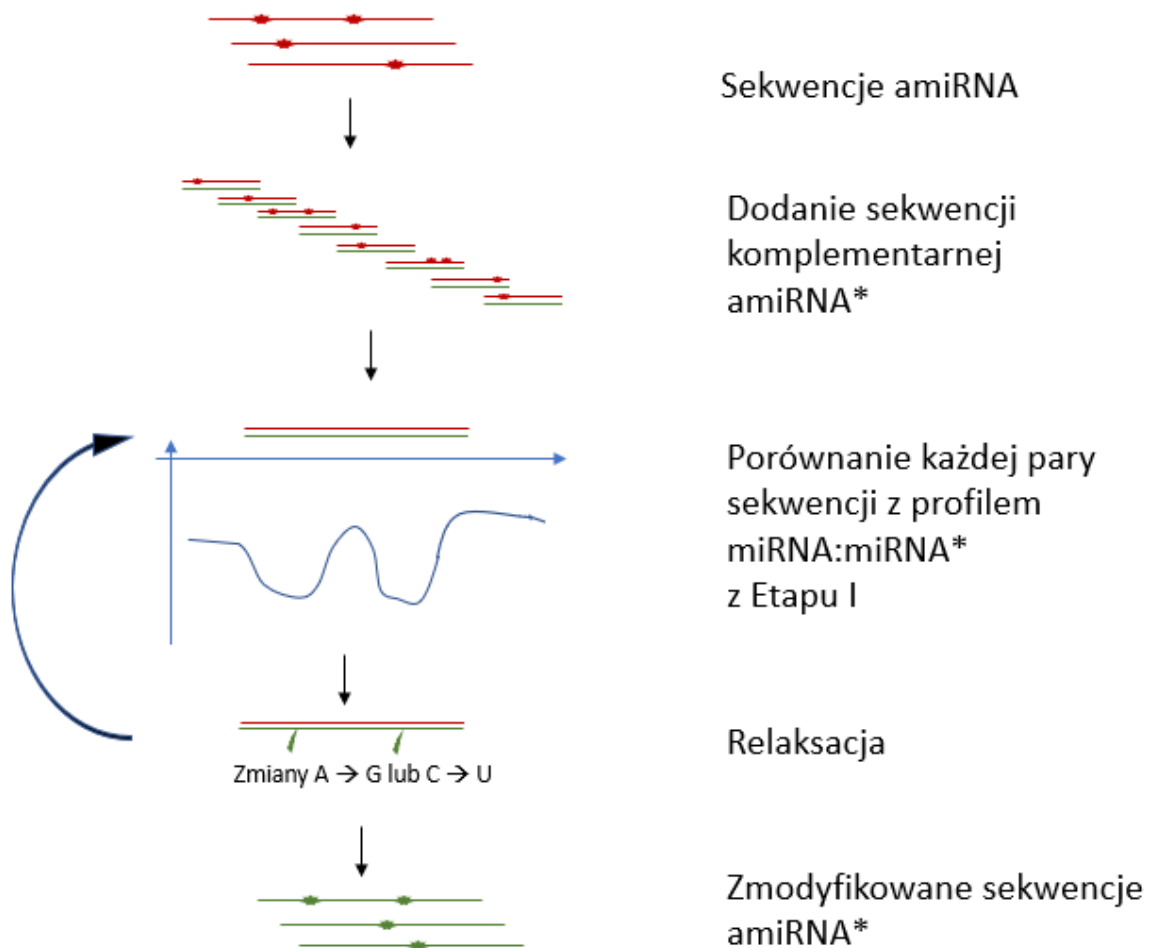


**Rys. 4.10.** Ilustracja sposobów określania miejsc, w które mogą zostać wprowadzone substytucje nukleotydowe w projektowanych sekwencjach.

Po wprowadzeniu wszystkich niezbędnych i dopuszczalnych relaksacji przeliczane są parametry termodynamiczne zmodyfikowanego dupleksu amiRNA:fragment i porównywane z profilem odniesienia miRNA:target. Jeśli powyższe wymagania są spełnione zmodyfikowane amiRNA jest akceptowane.

W następnym etapie analogicznie zostaje zaprojektowana sekwencja amiRNA\*. Po pierwsze tworzona jest sekwencja odwrotnie komplementarna do amiRNA. Następnie parametry termodynamiczne dla każdego dupleksu amiRNA:amiRNA\* porównuje się z uprzednio utworzonym profilem prekursora. Zasady relaksacji są takie same jak opisano

poprzednio. Na tym etapie modyfikowana jest tylko sekwencja amiRNA\*, podczas gdy sekwencja amiRNA pozostaje stała. Ostatecznie jeśli profil energetyczny dla duplexu amiRNA:amiRNA\* także spełnia wszystkie określone warunki, sekwencja amiRNA\* także zostaje zaakceptowana. Proces przedstawiono graficznie na Rys. 4.11.



**Rys. 4.11.** Ilustracja etapu projektowania amiRNA\*.

Wszystkie kroki są wykonywane kolejno dla każdego potencjalnego miejsca docelowego. Kończącym wynikiem procesu projektowania jest tabela zawierająca pary amiRNA:amiRNA\* spełniające wszystkie zdefiniowane warunki (Przykładowy wynik pokazano w Tab. 4.4).

**Tab. 4.4.** Przykładowy wynik programu AmiRNA Designer: amiRNA i amiRNA\* zaprojektowane dla jednego z miejsc w genie AGB1. Na czerwono zaznaczono zmiany relaksacyjne wprowadzone w etapie projektowania amiRNA a na niebiesko zmiany relaksacyjne wprowadzone w etapie projektowania miRNA\*.

Id amiRNA	Sekwencja amiRNA	Id amiRNA*	Sekwencja amiRNA*
amiR_AGB_410_5U_18G	UGUCUCUGAAGAGUACGGCAA	antymiR_AGB_410_5U_18G_15G	UUGCCGUACUCUUCGGAGACA
amiR_AGB_410_5U	UGUCUCUGAAGAGUACGACAA	antymiR_AGB_410_5U_15G	UUGUCGUACUCUUCGGAGACA
amiR_AGB_410_18G	UGUCCUGAAGAGUACGGCAA	antymiR_AGB_410_18G_15G	UUGCCGUACUCUUCGGGGACA

#### 4.1.4.3. Weryfikacja wyników

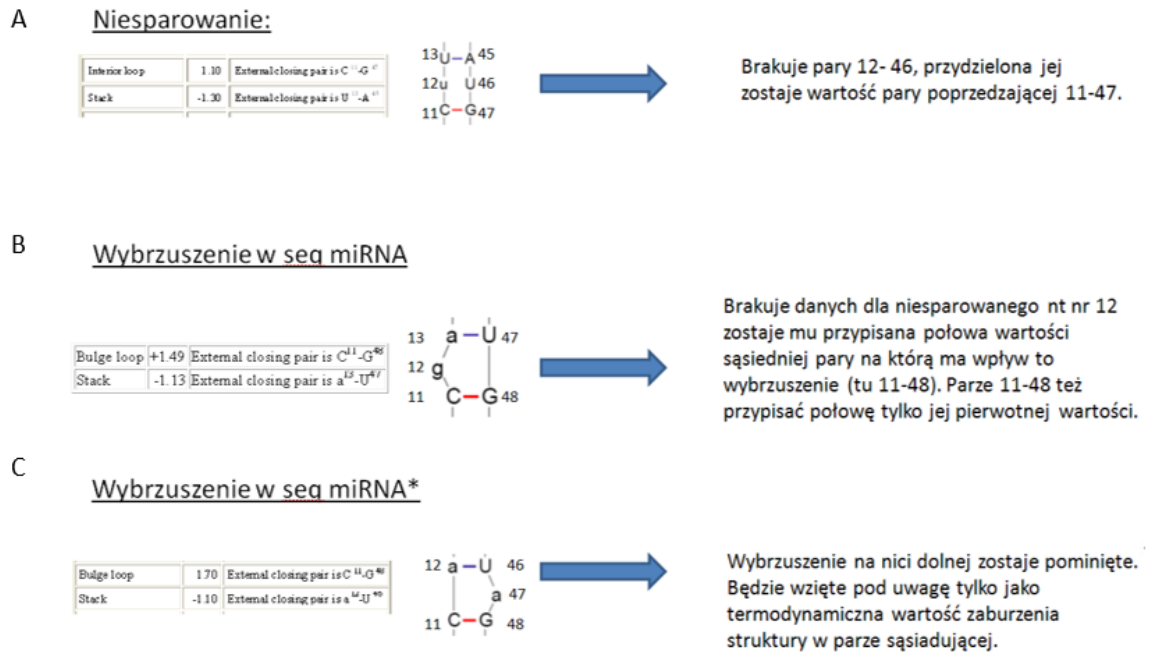
Wysoka komplementarność amiRNA do miejsca docelowego umożliwia zaprojektowanie sekwencji niemal unikalnej. Jednakże, wprowadzenie kilku niedopasowań w sekwencji amiRNA wymaga sprawdzenia, czy taka modyfikacja amiRNA nie może prowadzić do wyciszenia genów innych niż genu będącego przedmiotem zainteresowania.

Do weryfikacji ostatecznych sekwencji wygenerowanych przez AmiRNA Designer przygotowany został czwarty zestaw danych (dataset 4) (Rys. 4.12.). Stanowi on zbiór sekwencji wszystkich transkryptów obecnych u analizowanego organizmu (w tym przypadku *Arabidopsis thaliana*) w formacie fasta. Służy on do sprawdzenia zaprojektowanych sekwencji zarówno amiRNA jak i amiRNA\* pod kątem ich niespecyficznego oddziaływania z innymi, niż pierwotnie planowano, miejscami docelowymi.



niesparowanym. Wykorzystany tutaj został fakt, iż występowanie nieregularności w strukturze drugorzędowej ma wpływ na wartości zmiany energii swobodnej sąsiadujących z nim sparowanych nukleotydów. Na podstawie przewidzianych struktur drugorzędowych sekwencji odpowiednio – pary miRNA:target oraz prekursora - profile termodynamiczne obliczono w następujący sposób:

Dla każdego wybruszenia dwustronnego (ang. interior loop), które stanowi pojedyncze niesparowanie,  $\Delta G$  otrzymuje wartość taką samą jak poprzedzająca je para (określana w szczegółowym opisie rozkładu zmiany energii swobodnej dokonywanym przez programy typu mfold/UNAFold jako interior loop external closing pair) (Rys. 4.13.A). W przypadku gdy wybruszenie dwustronne jest większe niż 1 nukleotyd po każdej stronie nici – każdemu nukleotydowi odpowiadającemu konkretnej pozycji w miRNA zostaje przyporządkowana taka sama wartość. Jeśli mamy do czynienia z jednostronnym wybruszeniem (ang. bulge), postępowanie różni się w zależności czy występuje ono w sekwencji miRNA, czy też przeciwległej do niej (target lub miRNA\*). Dla pojedynczego (1 nt) wybruszenia na nici miRNA do pozycji o brakującej wartości  $\Delta G$  przypisana zostaje połowa wartości zmiany energii swobodnej pary poprzedzającej (Rys. 4.13.B). Tej parze przypisana zostaje taka sama wartość, czyli również połowa pierwotnej wartości  $\Delta G$ . W przypadku wystąpienia wybruszenia o większej liczbie nukleotydów zmiana energii swobodnej pary poprzedzającej zostaje podzielona przez liczbę nukleotydów w wybruszeniu plus jeden, a następnie przyporządkowana do każdej z wymienionych pozycji. Wybruszenie występujące na nici niebędącej miRNA zostaje uwzględnione tylko jako  $\Delta G$  poprzedzającej pary a dodatkowa pozycja w nici miRNA\* lub miejsca docelowego zostaje pominięta, aby utrzymać ciągłość sekwencji miRNA (Rys. 4.13.C). W przypadku wybruszenia w nici miRNA taka strategia spowodowałaby zmianę w sekwencji miRNA oraz przesunięcie wartości w profilu. Odmienne są traktowane miRNA pochodzące z ramienia 3' i 5' aby zachować wyżej wymienione reguły zawsze dla sekwencji miRNA oraz miRNA\*.



**Rys. 4.13.** Ilustracja sposobów przeliczania wartości zmian energii swobodnej dla różnych przypadków nieregularności występujących w strukturze drugorzędowej dupleksów miRNA:target oraz prekursora miRNA. Po lewej stronie parametry termodynamiczne oraz odpowiadający im fragment sekwencji drugorzędowej pochodzące z programu UnaFold. Wielkość liter nie jest istotna.

Następnym krokiem jest uśrednienie wartości zmian energii swobodnej dla każdej z kolejnych pozycji wraz z wartościami najbliższych sąsiadujących po obu stronach nukleotydów w sekwencji miRNA – okienko skanujące = 3 (za Król 2004). Ma to na celu „wygładzenie” profilu przy jednoczesnym odwzorowaniu rzeczywistej właściwości dwuniciowego RNA, gdzie na energię przypisaną do każdej pozycji mają wpływ pary sąsiadujące.

Uzyskane w opisany powyżej sposób wartości otrzymane dla obu zestawów danych (miRNA:target oraz fragmentów prekursora miRNA) stanowią podstawę do obliczeń statystycznych. Ponieważ rozkład energii na poszczególnych pozycjach nie jest rozkładem normalnym, do opisu profili termodynamicznych i porównywania z nimi nowych wartości przyjęta została analiza kwartyli. Metoda zastosowana w programie AmiRNA Designer opiera się na założeniu, że istotne statystycznie różnice w wartościach  $\Delta G$  na niektórych pozycjach mogą sugerować istnienie termodynamicznych parametrów będących wyznacznikami poprawnej biogenezy i funkcjonalności miRNA. Głównym porównywanym parametrem jest drugi kwartyl, czyli mediana oraz lokalny trend. Dodatkowym zaś jest trzeci kwartyl



wyznaczający najwyższą na danej pozycji wartość  $\Delta G$ , jaka nie może zostać przekroczona przy projektowaniu sekwencji amiRNA i amiRNA\*.

Analiza kwartyli jest częścią programu i jest generowana automatycznie dla wstępnych danych użytkownika. Uzyskane wyniki (w tym wypadku dla *A. thaliana*) są spójne z opisaną wcześniej obserwacją, że większa energia swobodna na końcu 5' miRNA oznacza mniejszą trwałość tej części dupleksu (Khvorova 2003, Król 2004). Konsekwencją tego jest wybór przez komórkowy system biogenezy srRNA właściwej nici miRNA jako „guide” i skierowanie przeciwległej nici „passenger” do degradacji. Również w tym aspekcie program AmiRNA Designer projektuje sekwencje zgodne z danymi literaturowymi, co zwiększa prawdopodobieństwo właściwego przetwarzania zaprojektowanych sekwencji w komórce.

Wszystkie wyniki otrzymane z opisanego powyżej przeliczenia danych mogą być zapisane jako arkusz w formacie CSV odpowiednio dla każdego profilu. Mogą być także używane do dalszego projektowania amiRNA – dla innych sekwencji docelowych – bez powtarzania obliczeń profili.

#### 4.1.6. Analiza przykładowych wyników

Gotowy program AmiRNA Designer wykorzystany został do zaprojektowania amiRNA dla dwóch następujących genów docelowych: AGB1 (ID: AT4G34460, występujących w literaturze również jako: ATAGB1, ELK4 (ang. Erecta-Like 4), białko wiążące GTP BETA 1, oraz Immutans (ID: AT4G22260, występujące w literaturze również jako: IM, IM1, IMMUTANS, plastydowego TERMINAL OXIDASE, PTOX). Dla tych genów nie odkryto u *Arabidopsis* żadnych naturalnych miRNA. Dodatkowa analiza *in silico* przeprowadzona dla tych genów z użyciem programu psRNATarget (Dai 2011) wykazała, że zestaw sekwencji miRNA wykorzystywany do tworzenia termodynamicznego profilu odniesienia nie zawiera miRNA skierowanych na badane sekwencje. Procedurę projektowania amiRNA i amiRNA\* przeprowadzono przy następujących parametrach: występowanie nukleotydu U w pozycji 12 w obrębie miejsca docelowego oraz wymuszonego U na pierwszym miejscu miRNA. W przypadku genu docelowego ABG1, AmiRNA Designer wygenerował 22 pary amiRNA:amiRNA\*, natomiast w przypadku Immutans – 20 par amiRNA:amiRNA\*. Wszystkie zaprojektowane sekwencje są przedstawione w tabelach Tab. 4.5. – AGB1 i Tab. 4.6 – IMM1. Analiza specyficzności

oddziaływania (off-target) wykazała, że żadne z amiRNA\* nie było komplementarne do żadnej z sekwencji kodującej mRNA *Arabidopsis thaliana*. Dla genu AGB1 tylko jeden amiRNA wskazywał potencjalne niespecyficzne oddziaływanie z hipotetycznym białkiem o numerach identyfikacyjnych: gi: 110740355, gb: ak230271 (Tab. 4.7.). W przypadku genu Immutans, potencjalnych niespecyficznych oddziaływań było więcej, jednakże można znaleźć co najmniej 3 pary amiRNA:amiRNA\* bez hipotetycznego efektu off-target (Tab. 4.8.). Większość sekwencji amiRNA mogących oddziaływać niespecyficznie była komplementarna do genów hipotetycznych, domniemanych lub nieznanymi białek. Ponadto sekwencje amiRNA\* nie są do tych genów w pełni komplementarne, a spełniają jedynie warunki opisane wcześniej (rodz. 4.1.4). Lista potencjalnych dodatkowych miejsc oddziaływania amiRNA wraz z określeniem miejsc niedopasowania dla każdego genu docelowego niebędącego pierwotnym targetem dla amiRNA są prezentowane użytkownikowi w celu podjęcia przez niego samodzielnej decyzji czy dany amiRNA jest odpowiedni do wykorzystania.

**Tab. 4.5.** Wyniki projektowania miRNA i miRNA\* dla genu AGB

amiRNA name	amiRNA sequence	amiRNA* name	amiRNA* sequence
amiR_AGB_132_6G_19U	UAGAGGGGGAGAUCAACUCU	antymiR_AGB_132_6G_19U_6U_1G	GGAGUUCGAUUCUCCCCUCUA
amiR_AGB_132_6G_19U	UAGAGGGGGAGAUCAACUCU	antymiR_AGB_132_6G_19U_11U_1G	GGAGUUCGAUUUCCCCUCUA
amiR_AGB_132_6G_19U	UAGAGGGGGAGAUCAACUCU	antymiR_AGB_132_6G_19U_6U	AGAGUUCGAUUCUCCCCUCUA
amiR_AGB_132_6G_19U	UAGAGGGGGAGAUCAACUCU	antymiR_AGB_132_6G_19U_11U	AGAGUUCGAUUUCCCCUCUA
amiR_AGB_132_6G_19U	UAGAGGGGGAGAUCAACUCU	antymiR_AGB_132_6G_19U_1G	GGAGUUCGAUUCUCCCCUCUA
amiR_AGB_132_6G_20U	UAGAGGGGGAGAUCAACUU	antymiR_AGB_132_6G_20U_6U	AAGGUUCGAUUCUCCCCUCUA
amiR_AGB_132_6G_20U	UAGAGGGGGAGAUCAACUU	antymiR_AGB_132_6G_20U_11U	AAGGUUCGAUUUCCCCUCUA
amiR_AGB_132_6G	UAGAGGGGGAGAUCAACCCU	antymiR_AGB_132_6G_6U_1G	GGGGUUCGAUUCUCCCCUCUA
amiR_AGB_132_6G	UAGAGGGGGAGAUCAACCCU	antymiR_AGB_132_6G_11U_1G	GGGGUUCGAUUUCCCCUCUA
amiR_AGB_132_6G	UAGAGGGGGAGAUCAACCCU	antymiR_AGB_132_6G_6U	AGGGUUCGAUUCUCCCCUCUA
amiR_AGB_132_6G	UAGAGGGGGAGAUCAACCCU	antymiR_AGB_132_6G_11U	AGGGUUCGAUUUCCCCUCUA
amiR_AGB_132_6G	UAGAGGGGGAGAUCAACCCU	antymiR_AGB_132_6G_1G	GGGGUUCGAUUCUCCCCUCUA
amiR_AGB_132_19U	UAGAGAGGGGAGAUCAACUCU	antymiR_AGB_132_19U_15U_1G	GGAGUUCGAUUCUUCUCUA
amiR_AGB_132_19U	UAGAGAGGGGAGAUCAACUCU	antymiR_AGB_132_19U_11U_1G	GGAGUUCGAUUUCCUCUCUA
amiR_AGB_132_19U	UAGAGAGGGGAGAUCAACUCU	antymiR_AGB_132_19U_15U	AGAGUUCGAUUCUUCUCUA
amiR_AGB_132_19U	UAGAGAGGGGAGAUCAACUCU	antymiR_AGB_132_19U_11U	AGAGUUCGAUUUCCUCUCUA
amiR_AGB_132_19U	UAGAGAGGGGAGAUCAACUCU	antymiR_AGB_132_19U_1G	GGAGUUCGAUUCUCCCCUCUA
amiR_AGB_132_20U	UAGAGAGGGGAGAUCAACUU	antymiR_AGB_132_20U_15U	AAGGUUCGAUUCUUCUCUA
amiR_AGB_132_20U	UAGAGAGGGGAGAUCAACUU	antymiR_AGB_132_20U_11U	AAGGUUCGAUUUCCUCUCUA
amiR_AGB_410_5U_18G	UGUCUCUGAAGAGUACGGCAA	antymiR_AGB_410_5U_18G_15G	UUGCCGUACUUCGGGAGACA
amiR_AGB_410_5U	UGUCUCUGAAGAGUACGGCAA	antymiR_AGB_410_5U_15G	UUGUCGUACUUCGGGAGACA
amiR_AGB_410_18G	UGUCCUGAAGAGUACGGCAA	antymiR_AGB_410_18G_15G	UUGCCGUACUUCGGGAGACA

Tab. 4.6. Wyniki projektowania miRNA i miRNA\* dla genu IMM

amiRNA name	amiRNA sequence	amiRNA* name	amiRNA* sequence
amiR_Imm_1149_19U	UCCUCAGGAACAACACACUCU	antymiR_Imm_1149_19U_15U_1G	GGAGUGUGUUGUUCUUGAGGA
amiR_Imm_1149_19U	UCCUCAGGAACAACACACUCU	antymiR_Imm_1149_19U_15U_21G	AGAGUGUGUUGUUCUUGAGGG
amiR_Imm_1149_19U	UCCUCAGGAACAACACACUCU	antymiR_Imm_1149_19U_15U	AGAGUGUGUUGUUCUUGAGGA
amiR_Imm_1149_19U	UCCUCAGGAACAACACACUCU	antymiR_Imm_1149_19U_1G	GGAGUGUGUUGUUCUUGAGGA
amiR_Imm_1149_19U	UCCUCAGGAACAACACACUCU	antymiR_Imm_1149_19U_21G	AGAGUGUGUUGUUCUUGAGGG
amiR_Imm_1149_20U	UCCUCAGGAACAACACACUU	antymiR_Imm_1149_20U_15U	AAGGUGUGUUGUUCUUGAGGA
amiR_Imm_1149_6G	UCCUCGGGAACAACACACCCU	antymiR_Imm_1149_6G_1G	GGGUGUGUUGUUCUUGAGGA
amiR_Imm_1149_6G	UCCUCGGGAACAACACACCCU	antymiR_Imm_1149_6G_21G	AGGUGUGUUGUUCUUGAGGG
amiR_Imm_1149_6G_19U	UCCUCGGGAACAACACACUCU	antymiR_Imm_1149_6G_19U_1G	GGAGUGUGUUGUUCUUGAGGA
amiR_Imm_1149_6G_19U	UCCUCGGGAACAACACACUCU	antymiR_Imm_1149_6G_19U_21G	AGAGUGUGUUGUUCUUGAGGG
amiR_Imm_283_19G	UCGACGAGAAGAGAGAGGGAG	antymiR_Imm_283_19G_6U_1U	UUCUCCUCUCUCUUCUUGUCGA
amiR_Imm_283_19G	UCGACGAGAAGAGAGAGGGAG	antymiR_Imm_283_19G_6U	CUCCUCUCUCUCUUCUUGUCGA
amiR_Imm_283_19G	UCGACGAGAAGAGAGAGGGAG	antymiR_Imm_283_19G_1U	UUCUCCUCUCUCUUCUUGUCGA
amiR_Imm_283_20G	UCGACGAGAAGAGAGAGGGAG	antymiR_Imm_283_20G_6U	CCUCCUCUCUCUUCUUGUCGA
amiR_Imm_283_7G	UCGACGGGAAGAGAGAGGAAG	antymiR_Imm_283_7G_6U	CUUCCUCUCUCUUCUUGUCGA
amiR_Imm_283_7G_19G	UCGACGGGAAGAGAGAGGGAG	antymiR_Imm_283_7G_19G_6U_1U	UUCUCCUCUCUCUUCUUGUCGA
amiR_Imm_283_7G_19G	UCGACGGGAAGAGAGAGGGAG	antymiR_Imm_283_7G_19G_6U	CUCCUCUCUCUCUUCUUGUCGA
amiR_Imm_283_7G_19G	UCGACGGGAAGAGAGAGGGAG	antymiR_Imm_283_7G_19G_1U	UUCUCCUCUCUCUUCUUGUCGA
amiR_Imm_283_7G_20G	UCGACGGGAAGAGAGAGGGAG	antymiR_Imm_283_7G_20G_6U	CCUCCUCUCUCUUCUUGUCGA
amiR_Imm_386_19U	UACCAGUAGAAGUCUCGGUUU	antymiR_Imm_386_19U_7G	AAACGGGACUUCUACUGGUA

Tab. 4.7. Wyniki analizy niespecyficznego oddziaływań dla genu AGB

Short regulatory RNA name	Hypothetical target	Short regulatory RNA sequence	Targeting sequence	Mismatched positions
amiR_AGB_132_6G_19U	no results	uagaggggagaucaacucu	no results	no results
amiR_AGB_132_6G_20U	gi 110740355 gb ak230271.1 ak230271.pln arabidopsis thaliana mrna for hypothetical protein, complete cds, clone: raf123-29-f17.	uagaggggagaucaacuu	aagcuuugaucuccaccucug	18 15 7
amiR_AGB_132_19U	no results	uagaggggagaucaacucu	no results	no results
amiR_AGB_132_19U	no results	uagaggggagaucaacucu	no results	no results
amiR_AGB_132_20U	no results	uagaggggagaucaacuu	no results	no results
amiR_AGB_132_20U	no results	uagaggggagaucaacuu	no results	no results
amiR_AGB_132_6G	no results	uagaggggagaucaacucu	no results	no results
amiR_AGB_132_6G	no results	uagaggggagaucaacucu	no results	no results
amiR_AGB_132_6G_19U	no results	uagaggggagaucaacucu	no results	no results
amiR_AGB_410_18G	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_18G	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_18G	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_5U	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_5U	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_5U	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_5U_18G	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_5U_18G	no results	uguccugaagagaucaacuu	no results	no results
amiR_AGB_410_5U_18G	no results	uguccugaagagaucaacuu	no results	no results
antymiR_AGB_132_6G_19U_6U_1G	no results	ggaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_19U_11U_1G	no results	ggaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_19U_6U	no results	agaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_19U_11U	no results	agaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_19U_1G	no results	ggaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_20U_6U	no results	aaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_20U_11U	no results	aaguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_6U_1G	no results	gggguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_11U_1G	no results	gggguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_6U	no results	aggguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_11U	no results	aggguucaucuccuccucua	no results	no results
antymiR_AGB_132_6G_1G	no results	gggguucaucuccuccucua	no results	no results
antymiR_AGB_132_19U_15U_1G	no results	ggaguucaucuccuccucua	no results	no results
antymiR_AGB_132_19U_11U_1G	no results	ggaguucaucuccuccucua	no results	no results
antymiR_AGB_132_19U_15U	no results	agaguucaucuccuccucua	no results	no results
antymiR_AGB_132_19U_11U	no results	agaguucaucuccuccucua	no results	no results
antymiR_AGB_132_19U_1G	no results	ggaguucaucuccuccucua	no results	no results
antymiR_AGB_132_20U_15U	no results	aaguucaucuccuccucua	no results	no results
antymiR_AGB_132_20U_11U	no results	aaguucaucuccuccucua	no results	no results
antymiR_AGB_410_5U_18G_15G	no results	uugccuacuucuccuccucua	no results	no results
antymiR_AGB_410_5U_15G	no results	uugccuacuucuccuccucua	no results	no results
antymiR_AGB_410_18G_15G	no results	uugccuacuucuccuccucua	no results	no results

Tab. 4.8. Wyniki analizy niespecyficznych oddziaływań dla genu IMM

Imm1\_12U\_Uforce\_result Off-Target analysis

Short regulatory RNA name	Hypotetical target	Short regulatory RNA sequence	Targeting sequence	Mismached positions
amiR_Imm_1149_19U	gi 26450514 gb ak117720.1 ak117720 pln arabidopsis thaliana mrna for unknown protein, complete cds, clone: raf17-40-e04.	uccucaggaacaacacacucu	agagucugguguccucagaggu	16 13 9
amiR_Imm_1149_20U	gi 17064829 gb ay062491.1 ay062491 pln arabidopsis thaliana zinc finger transcription factor-like protein (mzn1.8) mrna, complete cds.	uccucaggaacaacacacuu	aaggggucuuuccagaggu	17 14 6
amiR_Imm_1149_20U	gi 21655306 gb ay103314.1 ay103314 pln arabidopsis thaliana at5g58620/mzn1_70 mrna, complete cds.	uccucaggaacaacacacuu	aaggggucuuuccagaggu	17 14 6
amiR_Imm_1149_20U	gi 15809817 gb ay054176.1 ay054176 pln arabidopsis thaliana at5g58620/mzn1_70 mrna, complete cds.	uccucaggaacaacacacuu	aaggggucuuuccagaggu	17 14 6
amiR_Imm_1149_20U	gi 25083595 gb bt002083.1 bt002083 pln arabidopsis thaliana zinc finger transcription factor-like protein (at5g58620) mrna, complete cds.	uccucaggaacaacacacuu	aaggggucuuuccagaggu	17 14 6
amiR_Imm_1149_6G_19U	gi 67633515 gb dq056526.1 dq056526 pln arabidopsis thaliana putative e3 ubiquitin ligase scf complex subunit skp1/ask1 (at2g03170) mrna, complete cds.	uccucgggaacaacacacucu	cgacuguguuuaccgaggu	21 18 8
amiR_Imm_1149_6G	no results	uccucgggaacaacacacuu	no results	no results
amiR_Imm_283_19G	gi 2832401 gb y15055.1 y15055 pln arabidopsis thaliana mrna for dihydroxypolyprenylbenzoate methyltransferase.	ucgacgagaagagagaggag	cuucucucucugcugucuga	19 17 9
amiR_Imm_283_19G	gi 21403160 gb ay084450.1 ay084450 pln arabidopsis thaliana clone 108529 mrna, complete sequence.	ucgacgagaagagagaggag	cuucucucucugcugucuga	19 17 9
amiR_Imm_283_19G	gi 21407852 gb ay089078.1 ay089078 pln arabidopsis thaliana clone 2912 mrna, complete sequence.	ucgacgagaagagagaggag	gucucucucucucucucuga	21 18 5
amiR_Imm_283_19G	gi 20259178 gb ay091366.1 ay091366 pln arabidopsis thaliana putative dihydroxypolyprenylbenzoate methyltransferase (at2g30920) mrna, complete cds.	ucgacgagaagagagaggag	cuucucucucugcugucuga	19 17 9
amiR_Imm_283_19G	gi 22531063 gb ay136370.1 ay136370 pln arabidopsis thaliana unknown protein (at3g13380:at1g12760) mrna, complete cds.	ucgacgagaagagagaggag	aucucucucucucucucuga	21 18 5
amiR_Imm_283_19G	gi 222424259 gb ak317419.1 ak317419 pln arabidopsis thaliana at1g12760 mrna, complete cds, clone: raf123-01-e23.	ucgacgagaagagagaggag	aucucucucucucucucuga	21 18 5
amiR_Imm_283_19G	gi 67633363 gb dq056450.1 dq056450 pln arabidopsis thaliana dnaj chaperone c-terminal domain-containing protein (at1g11040) mrna, complete cds.	ucgacgagaagagagaggag	caccucucucucucucg	20 13 5
amiR_Imm_283_19G	gi 15027942 gb ay045828.1 ay045828 pln arabidopsis thaliana putative dihydroxypolyprenylbenzoate methyltransferase (at2g30920) mrna, complete cds.	ucgacgagaagagagaggag	cuucucucucugcugucuga	19 17 9
amiR_Imm_283_20G	gi 2832401 gb y15055.1 y15055 pln arabidopsis thaliana mrna for dihydroxypolyprenylbenzoate methyltransferase.	ucgacgagaagagagaggag	cuucucucucugcugucuga	20 17 9
amiR_Imm_283_20G	gi 19310724 gb ay079362.1 ay079362 pln arabidopsis thaliana putative m-type thioredoxin (at4g03520) mrna, complete cds.	ucgacgagaagagagaggag	ucucucucucucucggcg	21 18 4
amiR_Imm_283_20G	gi 21403160 gb ay084450.1 ay084450 pln arabidopsis thaliana clone 108529 mrna, complete sequence.	ucgacgagaagagagaggag	cuucucucucugcugucuga	20 17 9
amiR_Imm_283_20G	gi 21403206 gb ay084496.1 ay084496 pln arabidopsis thaliana clone 10950 mrna, complete sequence.	ucgacgagaagagagaggag	ucucucucucucucggcg	21 18 4

**Tab. 4.8.** Wyniki analizy niespecyficznego oddziaływań dla genu IMM c.d.

amiR_Imm_283_20G	gi 20259178 gb ay091366.1 ay091366 pln arabidopsis thaliana putative dihydroxypolyprenylbenzoate methyltransferase (at2g30920) mrna, complete cds.	ucgacgagaagagagaggagg	cuucucucucucugcucgca	20 17 9
amiR_Imm_283_20G	gi 62320064 gb ak221358.1 ak221358 pln arabidopsis thaliana mrna for putative m-type thioredoxin, complete cds, clone: raf125-30-j17.	ucgacgagaagagagaggagg	ucuucucucucucucgcg	21 18 4
amiR_Imm_283_20G	gi 15027942 gb ay045828.1 ay045828 pln arabidopsis thaliana putative dihydroxypolyprenylbenzoate methyltransferase (at2g30920) mrna, complete cds.	ucgacgagaagagagaggagg	cuucucucucucugcucgca	20 17 9
amiR_Imm_283_20G	gi 15028102 gb ay046001.1 ay046001 pln arabidopsis thaliana putative m-type thioredoxin (at4g03520) mrna, complete cds.	ucgacgagaagagagaggagg	ucuucucucucucgcg	21 18 4
amiR_Imm_283_7G	gi 26449367 gb ak117133.1 ak117133 pln arabidopsis thaliana at5g20540 mrna for unknown protein, complete cds, clone: raf116-67-a17.	ucgacgggaagagagaggagg	cuucuuucucuccgcg	17 15 4
amiR_Imm_283_7G_19G	no results	ucgacgggaagagagaggagg	no results	no results
amiR_Imm_386_19U	no results	uaccagugaagucucgguuu	no results	no results
antymiR_Imm_1149_19U_15U	no results	agaguguguuuuucugagg	no results	no results
antymiR_Imm_1149_19U_15U_1G	no results	ggaguguguuuuucugagg	no results	no results
antymiR_Imm_1149_19U_15U_21G	no results	agaguguguuuuucugagg	no results	no results
antymiR_Imm_1149_19U_1G	no results	ggaguguguuuuucugagg	no results	no results
antymiR_Imm_1149_19U_21G	no results	agaguguguuuuucugagg	no results	no results
antymiR_Imm_1149_20U_15U	no results	aagguguguuuuucugagg	no results	no results
antymiR_Imm_1149_6G_19U_1G	no results	ggaguguguuuuucgagg	no results	no results
antymiR_Imm_1149_6G_19U_21G	no results	agaguguguuuuucgagg	no results	no results
antymiR_Imm_1149_6G_1G	no results	ggguguguuuuucgagg	no results	no results
antymiR_Imm_1149_6G_21G	no results	aggguguguuuuucgagg	no results	no results
antymiR_Imm_283_19G_1U	no results	uuccucucucucugcga	no results	no results
antymiR_Imm_283_19G_6U	no results	cuccucucucucugcga	no results	no results
antymiR_Imm_283_19G_6U_1U	no results	uuccucucucucugcga	no results	no results
antymiR_Imm_283_20G_6U	no results	ccuccucucucucugcga	no results	no results
antymiR_Imm_283_7G_19G_1U	no results	uuccucucucuccgucga	no results	no results
antymiR_Imm_283_7G_19G_6U	no results	cuccucucucuccgucga	no results	no results
antymiR_Imm_283_7G_19G_6U_1U	no results	uuccucucucuccgucga	no results	no results
antymiR_Imm_283_7G_20G_6U	no results	ccuccucucucuccgucga	no results	no results
antymiR_Imm_283_7G_6U	no results	cuuccucucucuccgucga	no results	no results
antymiR_Imm_386_19U_7G	no results	aaacgggacucucugua	no results	no results

Dla dodatkowej weryfikacji oprogramowania AmiRNA Designer, zastosowano serwer psRNATarget aby potwierdzić czy sekwencje docelowe odpowiadające zaprojektowanym sekwencjom amiRNA i amiRNA\* są specyficzne. Przy użyciu domyślnych parametrów, serwer ten stworzył listę transkryptów genów docelowych, do których mogą potencjalnie wiązać się wygenerowane amiRNA, zasadniczo pokrywającą się z analizami przeprowadzonymi za pomocą przygotowanych na potrzeby tej pracy skryptów. Należy jednak zauważyć, że najlepsze wyniki wykazywały geny docelowe, do których właściwe amiRNA zostały zaprojektowane. Wyniki analizy psRNATarget zostały przedstawione w Tab. 4.9 i Tab. 4.10.

Do oceny komplementarności małych RNA i ich docelowego transkryptu w programie psRNATarget stosowany jest schemat punktacji opisany w publikacji Miru Zhang (Zhang 2005). Próg punktacji jest ustanowiony przez maksymalną wartość „Expectation (E)”. Para miRNA:target zostanie odrzucona, jeśli wynik jest większy od wartości progowej E. Domyślny próg odcięcia wynosi 3,0. Bardziej rygorystyczny próg odcięcia [0 - 2,0] powoduje zmniejszenie liczby przewidywanych wyników fałszywie dodatnich natomiast mniej rygorystyczny [4,0 - 5,0] dla uzyskania większej liczby wyników. W przeprowadzonej analizie pozostawiono domyślną wartość E = 3,0.

Dostępność transkryptu w miejscu docelowym dla małych RNA została określona w programie psRNATarget jako istotny czynnik, zaangażowany w rozpoznanie sekwencji docelowej, ponieważ także struktura drugorzędowa wokół niej może ograniczać hybrydyzację małych RNA. Serwer psRNATarget wykorzystuje program RNAup (Muckstein 2005) do obliczania wartości „Target Accessibility (UPE)” sekwencji docelowej. Jest ona reprezentowana przez energię potrzebną do rozplecenia struktury drugorzędowej wokół miejsca docelowego (zazwyczaj regionu komplementarnego z małym RNA oraz regionów flankujących – 17 nukleotydów od strony 5’ i 13 od strony 3’ targetu (Kertesz 2007)). Mniejsza energia swobodna oznacza większą łatwość hybrydyzacji miRNA z docelową sekwencją transkryptu.

Ponadto, w zależności czy występuje niesparowanie w centralnym regionie pary miRNA:target, przewidywany jest mechanizm regulacji jako represja translacji lub rozcięcie transkryptu (Brodersen 2008). Jeden z modeli działania srRNA (Axtell 2005) sugeruje, że małe regulatorowe RNA może mieć wiele miejsc docelowych (tzw. regiony uzupełniające) na określonym transkrypcie, co zwiększa efektywność srRNA. Serwer psRNATarget oblicza liczbę miejsc docelowych dla każdej pary srRNA:target określając czy spodziewana jest represja translacji, czy cięcie mRNA. Hipoteza o miejscach uzupełniających wymaga jeszcze dodatkowych badań i nie była brana pod uwagę na tym etapie projektowania amiRNA. Jest to jednak ciekawy wątek, który może być rozwijany w dalszych badaniach.

Tab. 4.9. Wyniki analizy psRNATarget dla amiRNA\* dla genu AGB1 ID: AT4G34460

miRNA Acc.	Target Acc.	Expectation (E)	Target Accessibility (UPE)	Alignment	Target Description	Inhibition	Multiplicity
amiR_AGB_132_19U	AT4G34460.1	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 155 AGGUUCGUAUCUCCUUCUCUC 175	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477106-16479528, REVERSE, LENGTH=1683	Cleavage	1
amiR_AGB_132_19U	AT4G34460.2	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 131 AGGUUCGUAUCUCCUUCUCUC 151	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479504, REVERSE, LENGTH=1868	Cleavage	1
amiR_AGB_132_19U	AT4G34460.3	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 133 AGGUUCGUAUCUCCUUCUCUC 153	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479506, REVERSE, LENGTH=1642	Cleavage	1
amiR_AGB_132_19U	AT4G34460.4	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 133 AGGUUCGUAUCUCCUUCUCUC 153	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479506, REVERSE, LENGTH=1616	Cleavage	1
amiR_AGB_132_19U	AT3G23633.1	2.5	17.119	miRNA 20 CUCAAGCUAGAGGGGAGAGAU 1 Target 686 GAGUUCGUAUCUCCUUCUCUA 705	Symbols:   F-box associated ubiquitination effector family protein   chr3:8496440-8497281, REVERSE, LENGTH=842	Translation	1
amiR_AGB_132_19U	AT3G23633.2	2.5	17.119	miRNA 20 CUCAAGCUAGAGGGGAGAGAU 1 Target 545 GAGUUCGUAUCUCCUUCUCUC 564	Symbols:   F-box associated ubiquitination effector family protein   chr3:8496440-8497281, REVERSE, LENGTH=701	Translation	1
amiR_AGB_132_19U	AT2G09865.1	2.5	24.855	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 808 AGAGUUCGUAUCUCCUUCUCUA 828	Symbols:   transposable element gene   chr2:3729543-3731042, REVERSE, LENGTH=1020	Cleavage	1
amiR_AGB_132_19U	AT2G30170.1	3.0	19.459	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 819 GGGUUUGAUGUUCUCCUUCUC 839	Symbols:   Protein phosphatase 2C family protein   chr2:12879675-12881498, REVERSE, LENGTH=1048	Translation	1
amiR_AGB_132_19U	AT2G30170.2	3.0	19.459	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 840 GGGUUUGAUGUUCUCCUUCUC 860	Symbols:   Protein phosphatase 2C family protein   chr2:12879675-12881503, REVERSE, LENGTH=1069	Translation	1
amiR_AGB_132_20U	AT4G34460.1	1.5	9.23	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 155 AGGUUCGUAUCUCCUUCUCUC 175	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477106-16479528, REVERSE, LENGTH=1683	Cleavage	1
amiR_AGB_132_20U	AT4G34460.2	1.5	9.23	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 131 AGGUUCGUAUCUCCUUCUCUC 151	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479504, REVERSE, LENGTH=1868	Cleavage	1
amiR_AGB_132_20U	AT4G34460.3	1.5	9.23	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 133 AGGUUCGUAUCUCCUUCUCUC 153	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479506, REVERSE, LENGTH=1642	Cleavage	1
amiR_AGB_132_20U	AT4G34460.4	1.5	9.23	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 133 AGGUUCGUAUCUCCUUCUCUC 153	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479506, REVERSE, LENGTH=1616	Cleavage	1
amiR_AGB_132_20U	AT2G27400.1	2.5	18.205	miRNA 20 UCUCAAGCUAGAGGGGAGAGAU 1 Target 234 AGGUUCGUAUCUCCUUCUCUA 253	Symbols: TAS1A   TAS1A; other RNA   chr2:11721539-11722468, REVERSE, LENGTH=930	Translation	1
amiR_AGB_132_20U	AT5G28350.1	2.5	24.245	miRNA 20 UCUCAAGCUAGAGGGGAGAGAU 1 Target 2231 AGGUUCGUAUCUCCUUCUCUA 2250	Symbols:   Quinoprotein amine de-hydrogenase, beta chain-like; RIC1-like guanyl-nucleotide exchange factor   chr5:10320139-10326946, FORWARD, LENGTH=3711	Translation	1
amiR_AGB_132_20U	AT5G28350.2	2.5	24.245	miRNA 20 UCUCAAGCUAGAGGGGAGAGAU 1 Target 2045 AGGUUCGUAUCUCCUUCUCUA 2064	Symbols:   Quinoprotein amine de-hydrogenase, beta chain-like; RIC1-like guanyl-nucleotide exchange factor   chr5:10320205-10326891, FORWARD, LENGTH=3470	Translation	1
amiR_AGB_132_20U	AT1G52170.1	2.5	24.814	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 51 GAGGUUCGACCCUUCUCUCUA 71	Symbols:   pre-tRNA   chr1:19424555-19424628, FORWARD, LENGTH=74	Translation	1
amiR_AGB_132_20U	AT2G18310.1	2.5	24.814	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 51 GAGGUUCGACCCUUCUCUCUA 71	Symbols:   pre-tRNA   chr2:7955677-7955750, REVERSE, LENGTH=74	Translation	1
amiR_AGB_132_20U	AT4G22635.1	2.5	24.814	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 51 GAGGUUCGACCCUUCUCUCUA 71	Symbols:   pre-tRNA   chr4:11910289-11910362, FORWARD, LENGTH=74	Translation	1
amiR_AGB_132_20U	AT2G30170.1	3.0	19.459	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 819 GGGUUUGAUGUUCUCCUUCUC 839	Symbols:   Protein phosphatase 2C family protein   chr2:12879675-12881498, REVERSE, LENGTH=1048	Translation	1
amiR_AGB_132_20U	AT2G30170.2	3.0	19.459	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 840 GGGUUUGAUGUUCUCCUUCUC 860	Symbols:   Protein phosphatase 2C family protein   chr2:12879675-12881503, REVERSE, LENGTH=1069	Translation	1
amiR_AGB_132_20U	AT2G30780.1	3.0	21.189	miRNA 20 UCUCAAGCUAGAGGGGAGAGAU 1 Target 1589 AGCUUCGUAUCUCCUUCUCUA 1608	Symbols:   Tetratricopeptide repeat (TPR)-like superfamily protein   chr2:13116474-13118808, FORWARD, LENGTH=2181	Cleavage	1
amiR_AGB_132_20U	AT5G54310.1	3.0	24.625	miRNA 21 UUCAAGCUAGAGGGGAGAGAU 1 Target 403 AAGGUUCGUAUCUCCUUCUCUC 423	Symbols: NEV, AGD5   ARF-GAP domain 5   chr5:22057052-22061270, REVERSE, LENGTH=1866	Translation	1
amiR_AGB_132_6G	AT4G34460.1	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 155 AGGUUCGUAUCUCCUUCUCUC 175	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477106-16479528, REVERSE, LENGTH=1683	Cleavage	1
amiR_AGB_132_6G	AT4G34460.2	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 131 AGGUUCGUAUCUCCUUCUCUC 151	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479504, REVERSE, LENGTH=1868	Cleavage	1
amiR_AGB_132_6G	AT4G34460.3	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 133 AGGUUCGUAUCUCCUUCUCUC 153	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479506, REVERSE, LENGTH=1642	Cleavage	1
amiR_AGB_132_6G	AT4G34460.4	1.5	9.23	miRNA 21 UCUCAAGCUAGAGGGGAGAGAU 1 Target 133 AGGUUCGUAUCUCCUUCUCUC 153	Symbols: AGB1, ELK4, ATAGB1   GTP binding protein beta 1   chr4:16477136-16479506, REVERSE, LENGTH=1616	Cleavage	1
amiR_AGB_132_6G	AT5G39390.1	2.5	15.876	miRNA 20 CCCAAGCUAGAGGGGAGAGAU 1 Target 258 CGGUUCGUAUCUCCUUCUCUA 277	Symbols:   Leucine-rich repeat protein kinase family protein   chr5:15763715-15765469, REVERSE, LENGTH=1509	Cleavage	1
amiR_AGB_132_6G	AT2G39630.1	3.0	15.998	miRNA 20 CCCAAGCUAGAGGGGAGAGAU 1 Target 53 GGUUCGUAUCUCCUUCUCUA 72	Symbols:   Nucleotide-diphospho-sugar transferases superfamily protein   chr2:16521744-16524687, REVERSE, LENGTH=1401	Cleavage	1

Tab. 4.10. Wyniki analizy psRNATarget dla amiRNA\* dla genu IMM1 ID: AT4G22260

miRNA Acc.	Target Acc.	Expectation (E)	Target Accessibility (UPE)	Alignment	Target Description	Inhibition	Multiplicity
amiR_imm_1149_19U	AT4G22260.1	0.5	20.822	miRNA 21 UCUCACACAACAAGGCCUCCU 1 :::..... Target 1150 AGGGUGUGUUGUCCUGAGGA 1170	Symbols: IM, IMM1   Alternative oxidase family protein   chr4:11769754-11772516, REVERSE, LENGTH=1435	Cleavage	1
amiR_imm_1149_19U	AT4G13215.1	1.5	20.266	miRNA 20 CUCACACAACAAGGCCUCCU 1 :::..... Target 161 GAGGGUGUUGUCCUGAGGG 180	Symbols:   pseudogene, leucine-rich repeat transmembrane protein kinase-related, similar to leucine-rich repeat transmembrane protein kinase, putative; chr4:7673224-7673760, REVERSE, LENGTH=537	Cleavage	1
amiR_imm_1149_19U	AT1G36580.1	3.0	13.992	miRNA 21 UCUCACACAACAAGGCCUCCU 1 :::..... Target 242 GGAUGAGUUGUCCUGAGGA 262	Symbols:   BEST Arabidopsis thaliana protein match is: short-chain dehydrogenase-reductase B (TAIR:AT3G12800.1); chr1:13816459-13818556, REVERSE LENGTH=875	Translation	1
amiR_imm_1149_19U	AT1G48175.1	3.0	18.299	miRNA 21 UCUCACACAACAAGGCCUCCU 1 .. :..... Target 169 GGGUGUGUUGUCCUGAGGA 189	Symbols: emb2191   Cytidine/deoxycytidylate deaminase family protein   chr1:1790507-1792331, FORWARD, LENGTH=890	Translation	1
amiR_imm_1149_20U	AT4G22260.1	0.5	20.822	miRNA 21 UUCACACAACAAGGCCUCCU 1 :::..... Target 1150 AGGGUGUGUUGUCCUGAGGA 1170	Symbols: IM, IMM1   Alternative oxidase family protein   chr4:11769754-11772516, REVERSE, LENGTH=1435	Cleavage	1
amiR_imm_1149_20U	AT4G14360.1	3.0	13.172	miRNA 21 UUCACACAACAAGGCCUCCU 1 :::..... Target 1810 AAGUAGUUGUCCUGAGGA 1830	Symbols:   S-adenosyl-L-methionine-Dependent methyltransferases superfamily protein   chr4:8267660-8271121, REVERSE, LENGTH=2418	Cleavage	1
amiR_imm_1149_20U	AT4G14360.2	3.0	13.172	miRNA 21 UUCACACAACAAGGCCUCCU 1 :::..... Target 1687 AAGUAGUUGUCCUGAGGA 1707	Symbols:   S-adenosyl-L-methionine-dependent methyltransferases superfamily protein   chr4:8267289-8270993, REVERSE, LENGTH=2666	Cleavage	1
amiR_imm_1149_20U	AT2G41610.1	3.0	16.816	miRNA 20 UCCACACAACAAGGCCUCCU 1 :::..... Target 684 UGGUGUUGUUGUCCUGAGGA 703	Symbols:   unknown protein; chr2:17347157-17348239, FORWARD, LENGTH=1083	Cleavage	1
amiR_imm_1149_20U	AT1G48175.1	3.0	18.299	miRNA 21 UUCACACAACAAGGCCUCCU 1 .. :..... Target 169 GGGUGUGUUGUCCUGAGGA 189	Symbols: emb2191   Cytidine/deoxycytidylate deaminase family protein   chr1:1790507-1792331, FORWARD, LENGTH=890	Translation	1
amiR_imm_1149_20U	AT4G38200.1	3.0	22.617	miRNA 21 UUCACACAACAAGGCCUCCU 1 :::..... Target 1717 GAGGUGUGUCCUGAGGA 1738	Symbols:   SEC7-like guanine nucleotide exchange family protein   chr4:17915293-17922732, FORWARD, LENGTH=5294	Cleavage	1
amiR_imm_1149_6G	AT4G22260.1	0.5	20.822	miRNA 21 UCCACACAACAAGGCCUCCU 1 :::..... Target 1150 AGGGUGUGUUGUCCUGAGGA 1170	Symbols: IM, IMM1   Alternative oxidase family protein   chr4:11769754-11772516, REVERSE, LENGTH=1435	Cleavage	1
amiR_imm_1149_6G	AT1G09930.1	1.5	20.047	miRNA 20 UCCACACAACAAGGCCUCCU 1 :::..... Target 1440 GGGUGUGUUGUACCGGGGA 1459	Symbols: ATOPT2, OPT2   oligopeptide transporter 2   chr1:3227490-3230043, REVERSE, LENGTH=2205	Cleavage	1
amiR_imm_1149_6G	AT1G48175.1	3.0	18.299	miRNA 21 UCCACACAACAAGGCCUCCU 1 .. :..... Target 169 GGGUGUGUUGUCCUGAGGA 189	Symbols: emb2191   Cytidine/deoxycytidylate deaminase family protein   chr1:1790507-1792331, FORWARD, LENGTH=890	Translation	1
amiR_imm_1149_6G_19U	AT4G22260.1	1.0	20.822	miRNA 21 UCUCACACAACAAGGCCUCCU 1 :::..... Target 1150 AGGGUGUGUUGUCCUGAGGA 1170	Symbols: IM, IMM1   Alternative oxidase family protein   chr4:11769754-11772516, REVERSE, LENGTH=1435	Cleavage	1
amiR_imm_1149_6G_19U	AT1G09930.1	2.0	20.047	miRNA 20 CUCACACAACAAGGCCUCCU 1 :::..... Target 1440 GGGUGUGUUGUACCGGGGA 1459	Symbols: ATOPT2, OPT2   oligopeptide transporter 2   chr1:3227490-3230043, REVERSE, LENGTH=2205	Cleavage	1
amiR_imm_1149_6G_19U	AT4G13215.1	2.0	20.266	miRNA 20 CUCACACAACAAGGCCUCCU 1 :::..... Target 161 GAGGGUGUUGUCCUGAGGG 180	Symbols:   pseudogene, leucine-rich repeat transmembrane protein kinase-related, similar to leucine-rich repeat transmembrane protein kinase, putative; chr4:7673224-7673760, REVERSE, LENGTH=537	Cleavage	1
amiR_imm_1149_6G_19U	AT2G03170.1	3.0	19.775	miRNA 20 CUCACACAACAAGGCCUCCU 1 :::..... Target 106 GACUGUGUUGUACCAGGAGU 125	Symbols: ASK14, SK14   SKP1-like 14   chr2:961322-961771, FORWARD, LENGTH=450	Cleavage	1
amiR_imm_283_19G	AT2G22830.1	1.5	6.87	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 91 UUCUCUCUCUCUCUCUCUGUGG 111	Symbols: SOE2   squalene epoxidase 2   chr2:9723680-9726270, REVERSE, LENGTH=2025	Translation	1
amiR_imm_283_19G	AT1G75180.2	2.0	9.366	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 41 UUGCCUCUCUCUCUCUGUGG 61	Symbols:   Erythrone-4-phosphate dehydrogenase family protein   chr1:28215841-28218801, REVERSE, LENGTH=1828	Cleavage	1
amiR_imm_283_19G	AT1G75180.3	2.0	9.366	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 41 UUGCCUCUCUCUCUCUGUGG 61	Symbols:   Erythrone-4-phosphate dehydrogenase family protein   chr1:28215836-28218801, REVERSE, LENGTH=1787	Cleavage	1
amiR_imm_283_19G	AT4G35250.1	2.0	9.647	miRNA 20 AGGGAGAGAGAAGGAGCAGCU 1 :::..... Target 71 UCUCUCUCUCUCUCUCUGGA 90	Symbols:   NAD(P)-binding Rossmann-fold superfamily protein   chr4:1677171-16773364, REVERSE, LENGTH=1513	Cleavage	1
amiR_imm_283_19G	AT1G20460.1	2.0	11.462	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 215 CUGUCUCUCUCUCUCUGGA 235	Symbols:   unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G76185.1); chr1:7091361-7092993, FORWARD, LENGTH=645	Cleavage	1
amiR_imm_283_19G	AT2G30920.1	2.0	14.176	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 145 CUUCUCUCUCUCUCUGGA 165	Symbols: ATCOQ3, EMB3002, COQ3   coenzyme Q 3   chr2:13157133-13159932, REVERSE, LENGTH=1353	Translation	1
amiR_imm_283_19G	AT5G45420.1	2.0	14.215	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 210 UUCUCUCUCUCUCUCUGCG 230	Symbols:   Duplicated homeodomain-like superfamily protein   chr5:18404069-18406606, REVERSE, LENGTH=1292	Translation	1
amiR_imm_283_19G	AT1G21920.1	2.5	11.747	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 124 CUCUCUCUCUCUCUCUGGA 144	Symbols:   Histone H3 K4-specific methyltransferase SET7/9 family protein   chr1:7704255-7706140, REVERSE, LENGTH=1727	Translation	1
amiR_imm_283_19G	AT3G07960.1	3.0	7.702	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 179 CUCUCUCUCUCUCUCUGGU 199	Symbols:   Phosphatidylinositol-4-phosphate 5-kinase family protein   chr3:2534887-2538372, FORWARD, LENGTH=2817	Translation	1
amiR_imm_283_19G	AT5G48120.1	3.0	10.167	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 2 CUCUCUCUCUCUCUCUGCG 22	Symbols:   ARM repeat superfamily protein   chr5:19507902-19514854, REVERSE, LENGTH=3793	Cleavage	1
amiR_imm_283_20G	AT2G30920.1	2.0	14.176	miRNA 21 GAGGGAGAGAGAAGGAGCAGCU 1 .. :..... Target 145 CUUCUCUCUCUCUCUGGA 165	Symbols: ATCOQ3, EMB3002, COQ3   coenzyme Q 3   chr2:13157133-13159932, REVERSE, LENGTH=1353	Translation	1



**Tab. 4.10.** Wyniki analizy psRNATarget dla amiRNA\* dla genu IMM1 ID: AT4G22260 c.d.

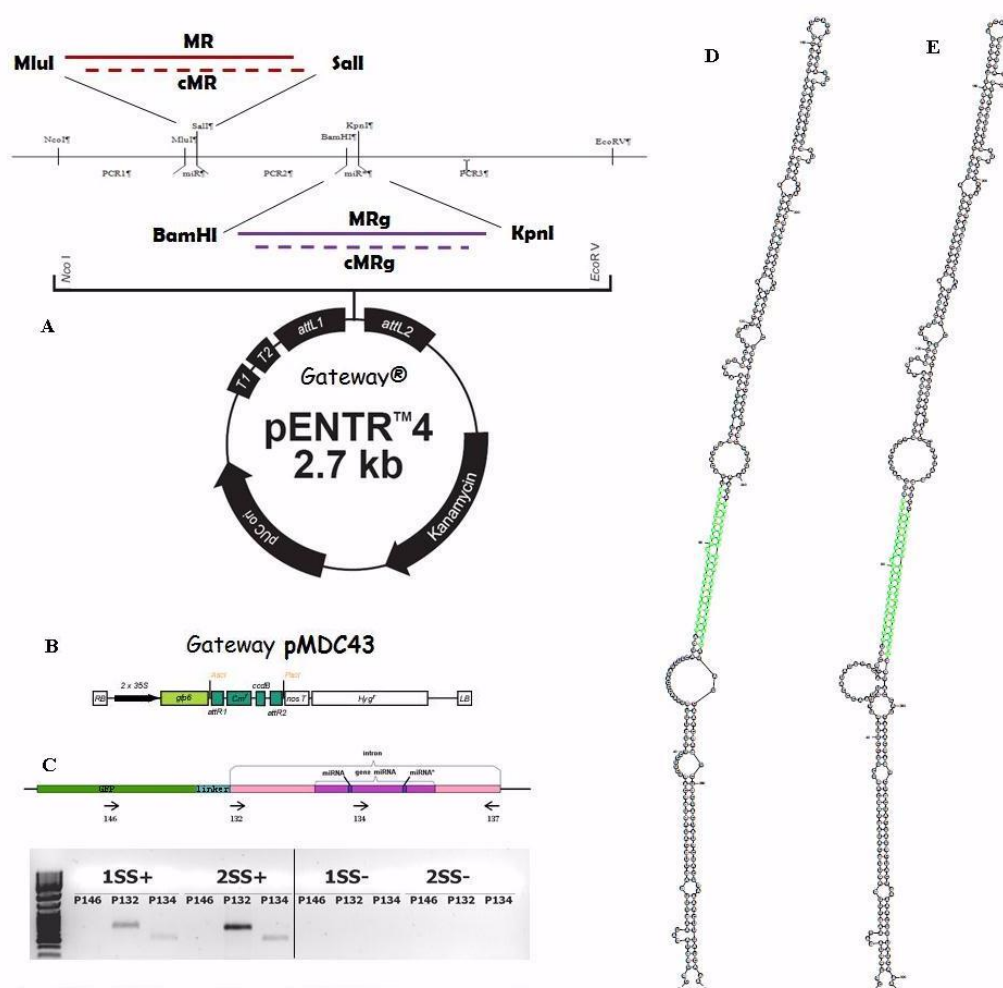
amiR_imm_283_20G	AT2G34500.1	2.5	7.055	miRNA 21 GGAGGAGAGAGAAGAGCAGCU 1 ..... Target 264 UCUCUCUUUCUCUCUCGUCGA 284	Symbols: CYP710A1   cytochrome P450, family 710, subfamily A, polypeptide 1   chr2:14539479-14541406, REVERSE, LENGTH=1928	Cleavage	1
amiR_imm_283_20G	AT5G61380.1	2.5	8.704	miRNA 21 GGAGGAGAGAGAAGAGCAGCU 1 ..... Target 36 UCUCUCACUCUCUCUCGCGG 56	Symbols: TOC1, APRR1, PRR1, ATOC1   CCT motif-containing response regulator protein   chr5:24675064-24678550, FORWARD, LENGTH=2707	Cleavage	1
amiR_imm_283_20G	AT1G20460.1	2.5	11.462	miRNA 21 GGAGGAGAGAGAAGAGCAGCU 1 ..... Target 215 CUCUCUCUCUCUCUCUCGUCGA 235	Symbols:   unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G76185.1)   chr1:7091361-7092993, FORWARD, LENGTH=465	Cleavage	1
amiR_imm_283_20G	AT1G07720.1	2.5	11.807	miRNA 20 GAGGAGAGAGAAGAGCAGCU 1 ..... Target 66 CUCCUCUCUCUCUCUCGUCUC 85	Symbols: KCS3   3-ketoacyl-CoA synthase 3   chr1:2390655-2392453, REVERSE, LENGTH=1799	Cleavage	1
amiR_imm_283_20G	AT1G34000.1	2.5	13.625	miRNA 21 GGAGGAGAGAGAAGAGCAGCU 1 ..... Target 614 CCUUUCUCUCUCUCUCUCGUCGA 634	Symbols: OHP2   one-helix protein 2   chr1:12357910-12358966, REVERSE, LENGTH=824	Cleavage	1
amiR_imm_283_20G	AT4G03520.1	2.5	14.569	miRNA 21 GGAGGAGAGAGAAGAGCAGCU 1 ..... Target 219 UCUCUCUCUCUCUCUCGCGCGG 239	Symbols: ATHM2   Thioredoxin superfamily protein   chr4:1562359-1564192, REVERSE, LENGTH=924	Cleavage	1
amiR_imm_283_20G	AT1G80580.1	3.0	5.234	miRNA 21 GGAGGAGAGAGAAGAGCAGCU 1 ..... Target 161 UCUCUCUCUCUCUCUCGUCUUA 181	Symbols:   Integrase-type DNA-binding superfamily protein   chr1:30293558-30294328, FORWARD, LENGTH=771	Cleavage	1
amiR_imm_283_20G	AT5G27925.1	3.0	8.601	miRNA 20 GAGGAGAGAGAAGAGCAGCU 1 ..... Target 4210 GUCCUUUCUCUCUCUCUCGUA 4229	Symbols:   transposable element gene   chr5:9949686-9954107, FORWARD, LENGTH=4422	Cleavage	1
amiR_imm_283_20G	AT1G75180.2	3.0	9.366	miRNA 20 GAGGAGAGAGAAGAGCAGCU 1 ..... Target 42 UGCCUCUCUCUCUCUCGUGG 61	Symbols:   Erythronate-4-phosphate dehydrogenase family protein   chr1:28215841-28218801, REVERSE, LENGTH=1828	Cleavage	1
amiR_imm_283_20G	AT1G75180.3	3.0	9.366	miRNA 20 GAGGAGAGAGAAGAGCAGCU 1 ..... Target 42 UGCCUCUCUCUCUCUCGUGG 61	Symbols:   Erythronate-4-phosphate dehydrogenase family protein   chr1:28215836-28218801, REVERSE, LENGTH=1787	Cleavage	1
amiR_imm_283_20G	AT3G14310.1	3.0	17.006	miRNA 20 GAGGAGAGAGAAGAGCAGCU 1 ..... Target 2072 CUUCUCUCUCUCUCUCGUCGUA 2091	Symbols: ATPME3, PME3   pectin methyltransferase 3   chr3:4771902-4775119, REVERSE, LENGTH=2115	Cleavage	1
amiR_imm_283_7G	AT2G30920.1	2.0	14.176	miRNA 21 GAAGGAGAGAGAAGAGCAGCU 1 ..... Target 145 CUUCUCUCUCUCUCUCGUCGA 165	Symbols: ATCOQ3, EMB3002, COQ3   coenzyme Q 3   chr2:13157133-13159932, REVERSE, LENGTH=1353	Translation	1
amiR_imm_283_7G	AT3G24160.1	2.5	5.812	miRNA 21 GAAGGAGAGAGAAGAGCAGCU 1 ..... Target 84 CUUCUCUCUCUCUCUCGUCGA 104	Symbols: PMP   putative type 1 membrane protein   chr3:8726122-8729389, FORWARD, LENGTH=1580	Translation	1
amiR_imm_283_7G	AT1G62610.1	2.5	6.286	miRNA 21 GAAGGAGAGAGAAGAGCAGCU 1 ..... Target 74 UUCUCUCUCUCUCUCGUCGA 94	Symbols:   NAD(P)-binding Rossmann-fold superfamily protein   chr1:23181463-23183227, REVERSE, LENGTH=1439	Translation	1
amiR_imm_283_7G	AT1G62610.2	2.5	6.286	miRNA 21 GAAGGAGAGAGAAGAGCAGCU 1 ..... Target 74 UUCUCUCUCUCUCUCGUCGA 94	Symbols:   NAD(P)-binding Rossmann-fold superfamily protein   chr1:23181463-23183227, REVERSE, LENGTH=1436	Translation	1

#### 4.1.7. Szkielet sztucznego miRNA

Gen At1g77235 koduje miR402, który pochodzi z intronu genu oznaczonego jako At1g77230 w tym samym kierunku, co pre-mRNA (Sunkar 2004). Badania wykazały, że ekspresja tego genu indukowana jest abiotycznym stresem (Kim 2010). Według komputerowych analiz nie ma żadnych elementów regulacyjnych powyżej na nici sekwencji intronu (począwszy od intronu do końca prekursora miRNA). Sugeruje to, że cała sekwencja intronu może być umieszczona pod kontrolą promotora, który może zostać zaindukowany, w celu uzyskania wysokiego poziomu ekspresji, lub też jest specyficzny dla określonych tkanek lub dla danego stadium rozwoju rośliny i wówczas ekspresję wykazuje tylko w specyficznych dla siebie warunkach. Trzy fragmenty pierwszego intronu genu At1g77230 osobno amplifikowano z użyciem starterów, wyposażonych w miejsca restrykcyjne na ich końcach 5'. Kolejne klonowania wszystkich fragmentów do wektora pENTR4 zaowocowały utworzeniem „pustej” konstrukcji genowej stanowiącej szkielet o nazwie pENTR\_V0 (Rys. 4.14.A), który jest gotowy do wprowadzenia docelowych sekwencji regulujących. Następujące kroki prowadzą do uzyskania funkcjonalnego wektora: utworzenie odwrotnie komplementarnych sekwencji

do zaprojektowanych zarówno amiRNA i amiRNA\* – odpowiednio: RevComp\_amiRNA i RevComp\_amiRNA\*. Dupleks amiRNA:RevComp\_amiRNA, o sekwencjach zawierających końce odpowiadające tym powstałym po rozcięciu wektora pENTR\_V0 przez enzymy restrykcyjne Sall i Mlul, natomiast amiRNA\*:RevComp\_amiRNA\* – o sekwencjach zawierających końce odpowiadające powstałym po rozcięciu wektora pENTR\_V0 przez enzymami BamHI i KpnI. Zapewnia to, że oba wstawiane do wektora dwuniciowe fragmenty są umieszczane w odpowiednim położeniu w wektorze do klonowania. Wszystkie te cztery sekwencje są dostępne komercyjnie w postaci jednoniciowego syntetycznego DNA, gotowego do hybrydyzacji – w celu uzyskania dwuniciowych fragmentów – a następnie ligacji do przygotowanego, przeciętego wektora. Utworzenie całej konstrukcji składa się z dwóch cykli podwójnego trawienia odpowiednimi enzymami restrykcyjnymi wektora oraz ligacji. Pierwszy cykl obejmuje cięcie wektora pENTR\_V0 enzymami Sall i Mlul i następczą ligację do niego duplesu AmiRNA:RevComp\_amiRNA. Skutkuje to otrzymaniem plazmidu pENTR\_V1, który jest amplifikowany w bakteriiach *E.coli*, izolowany i oczyszczany. Drugi cykl obejmuje trawienie wektora pENTR\_V1 enzymami BamHI i KpnI. Przygotowany uprzednio dupleks AmiRNA\*:RevComp\_amiRNA\* jest do niego ligowany prowadząc do uzyskania plazmidu pENTR\_V2 zawierającego gotową konstrukcję wyciszającą, która po namnożeniu i oczyszczeniu jest gotowa do dalszej procedury. Taki wektor może być zrekombinowany z wybranym wektorem binarnym z systemu Gateway® w celu transformacji roślin i uzyskania stabilnej lub przejściowej ekspresji amiRNA. Eksperyment pilotażowy przeprowadzono w celu sprawdzenia, czy konstrukcja samego intronu nie posiadająca sekwencji regulatorowej amiRNA oraz amiRNA\* ulega prawidłowej przejściowej ekspresji i jest właściwie przetwarzana w liściach *Nicotiana benthamiana*. Wektor pENTR\_V2 został zrekombinowany z wektorem Gateway® pMDC43 (Rys. 4.14.B). W rezultacie powstała konstrukcja genowa niosąca intron genu At1g77230 związany z białkiem GFP w kasecie ekspresyjnej pod kontrolą promotora 2x35S i terminatora Nos. Rekombinowanym wektorem transformowano *Nicotiana benthamiana* za pomocą agrobakterii *A.tumefaciens* i po 3 dniach izolowano całkowity RNA. Reakcja odwrotnej transkrypcji potwierdziła ekspresję wprowadzonego konstruktu, jak również usunięcie intronu - nieobecność prążka o wielkości odpowiadającej długości genu GFP z intronem na żelu agarozowym (Rys. 4.14.C).

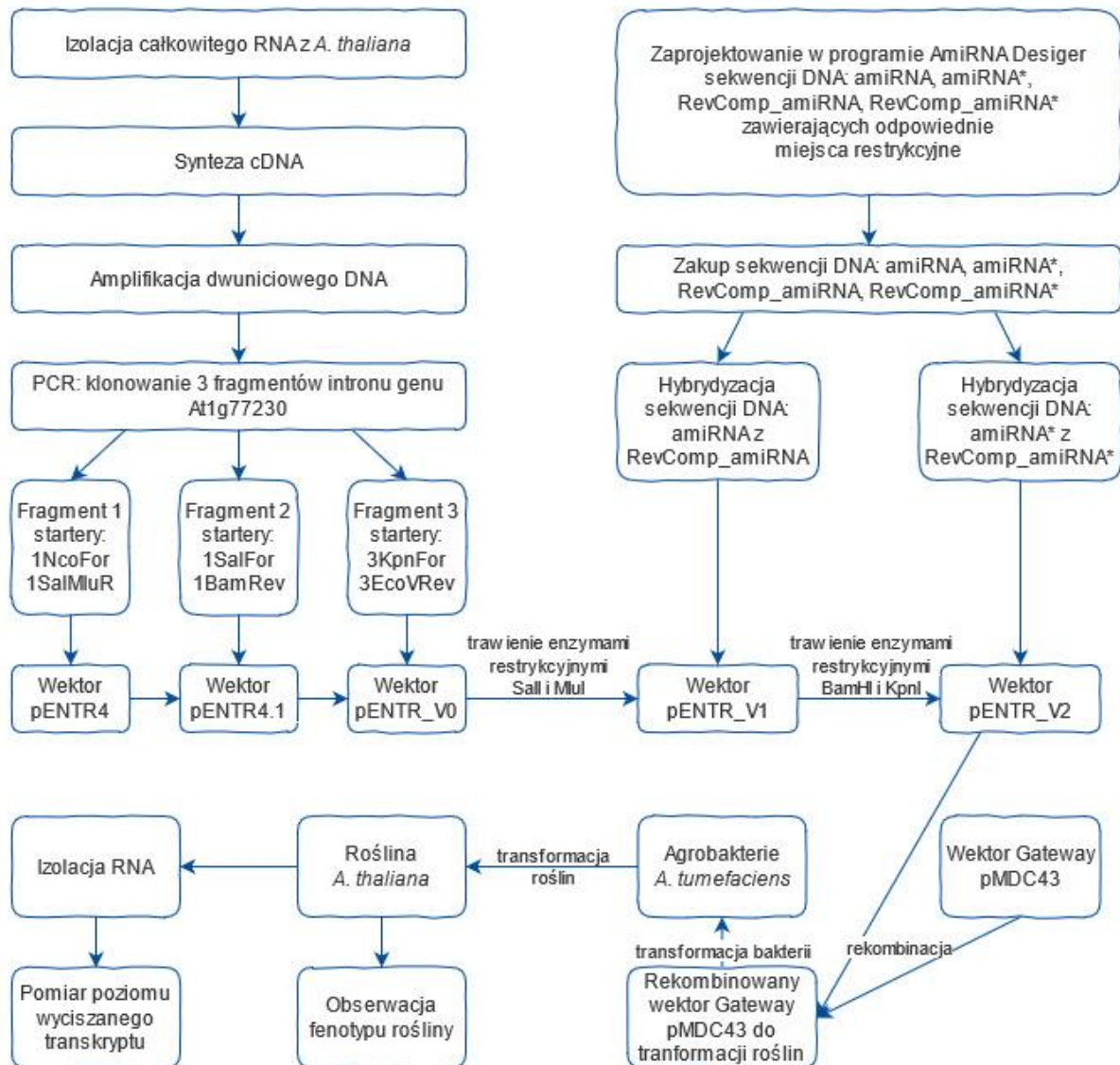
Dodatkowo przeprowadzono *in silico* analizę struktury drugorzędowej trzonu prekursora amiRNA. Obie sekwencje: oryginalnego intronu oraz przygotowanego szkieletu z wtórnie wprowadzonymi sekwencjami pierwotnie w nim występujących miRNA i miRNA\* zostały użyte do przewidywania struktur drugorzędowych. Ogólna struktura szkieletu pozostała zachowana zaś w pobliżu sekwencji miRNA:miRNA\* zaburzenia były podobne (Rys. 4.11.D, E). Sugeruje to, że w konstrukcji genowej niosącej zmodyfikowany szkielet intronu *A. thaliana* jest on utrzymywany w sposób umożliwiający przetwarzane go w szlaku RNAi w innych układach roślinnych, jak również zachowuje strukturę spinki do włosów.



**Rys. 4.14.** A. Schemat konstrukcji wektora wejściowego niosącego intron genu At1g77230 służącego do umieszczenia w nim zaprojektowanych sekwencji amiRNA oraz amiRNA\* wraz z odwrotnie do nich komplementarnymi i miejscami restrykcyjnymi. B. Kaseta ekspresyjna wektora binarnego pMDC43. Jego rekombinacja z wektorem wejściowym prowadzi do uzyskania wektora do ekspresji genu fuzyjnego z GFP w roślinach. C. Schemat gotowej konstrukcji ulegającej ekspresji w roślinie wraz z zaznaczonymi miejscami zaprojektowanych starterów do kontroli ekspresji i wycinania intronu w roślinie.

**Rys. 4.14. c.d.** Poniżej efekt reakcji PCR z użyciem tych starterów po reakcji odwrotnej transkrypcji na wyizolowanym RNA z transformowanych roślin. 1SS+ oznacza użycie odwrotnej transkryptazy, 1SS- stanowi kontrolę obecności DNA wektora. **D.** Przewidziana struktura drugorzędowa oryginalnego prekursora miRNA zawartego w intronie. **E.** Przewidziana struktura drugorzędowa po dodaniu miejsc restrykcyjnych i przywróceniu oryginalnej sekwencji wyciszającej.

Schemat zastosowania zaprojektowanych w programie AmiRNA Designer sekwencji do eksperymentu biologicznego wygląda następująco:



**Rys. 4.15.** Schemat postępowania z zaprojektowanymi sekwencjami w eksperymencie biologicznym

#### 4.1.8. Dyskusja

Jak dotąd powstało kilka narzędzi bioinformatycznych, które są w stanie zaprojektować amiRNA dla konkretnego organizmu lub grupy organizmów. Ostatnio opracowane narzędzia, takie jak miR-Synth (Laganà 2014), AmiRzyn (Baby 2012) są przeznaczone wyłącznie do ludzkich amiRNA. Z powodu znaczących różnic w biogenezie i sposobie działania zwierzęcego i roślinnego, miRNAe, narzędzia przeznaczone dla systemów zwierzęcych nie znajdują zastosowania w projektowaniu roślinnych sztucznych miRNA. Najnowszym z dostępnych narzędzi jest P-SAMS amiRNA Designer (Fahlgren 2016). Skupia się on na podobieństwie sekwencyjnym i nie bierze pod uwagę parametrów termodynamicznych. Jednym z najczęściej używanych narzędzi dedykowanych głównie do roślin jest internetowy MicroRNA Designer (WMD3) (Ossowski 2008, Schwab 2006). Ostatnie badania biologiczne weryfikujące niektóre amiRNA zaprojektowane przy jego pomocy wykazały, że parametry skutecznej sekwencji amiRNA nie są jeszcze w pełni zrozumiałe (Li 2013). Pewne sekwencje amiRNA określone przez WMD3 jako „zalecane” były mniej skuteczne, podczas gdy wyciszenie za pomocą sekwencji określonej przez WMD3 jako „wyciszająca w niewielkim stopniu” osiąga nawet 90%. Większość dostępnych narzędzi projektuje amiRNA na podstawie komplementarności sekwencji i innych parametrów, gdzie jednym z nich jest obliczenie całkowitej energii swobodnej dupleksu miRNA:target i/lub struktury drugorzędowej prekursora miRNA. W niniejszej pracy przedstawiono nową metodę projektowania sekwencji amiRNA. Bierze się tu pod uwagę rozkład energii swobodnej w strukturach znanych par miRNA:miRNA\* i miRNA:target na każdej z pozycji badanego dupleksu. Na tej podstawie, tworzone są termodynamiczne profile, z którymi są porównywane rozkłady energii swobodnej dla nowo opracowanych odpowiednich dupleksów zawierających amiRNA. Stworzony na potrzeby niniejszej pracy program AmiRNA Designer implementuje tę metodę do projektowania sekwencji potencjalnych amiRNA. Program udostępniono do powszechnego użycia (Mickiewicz 2016).

AmiRNA Designer może być stosowany do dowolnego zestawu danych oraz dowolnego organizmu, a nawet stadium rozwoju czy tkanki. Jego działanie opiera się na wyznaczeniu sekwencji, która tworzy strukturę drugorzędową podobną do profilu termodynamicznego, zdefiniowanego na podstawie zestawu sekwencji wskazanych jako wzorzec. Abstrahując od

problemu wyciszania genów można także zaprojektować takie sekwencje w innym dowolnie wybranym celu. Program jest na tyle elastyczny, że pozwala na wprowadzanie przez użytkownika dowolnie dużego i dowolnie różnorodnego zbioru danych. Należy jednak pamiętać, że może to spowodować iż różnice pomiędzy generowanymi strukturami zniwelują się wzajemnie, a otrzymany profil termodynamiczny będzie jedynie uśrednioną energią i nie będzie dla niego uzasadnienia funkcjonalnego. Może się również zdarzyć, że zadany zestaw sekwencji do stworzenia profilu uśredni się do niemal płaskiego histogramu, gdzie nie będzie podstaw do określenia zależności pomiędzy zmianą energii swobodnej przypisaną do poszczególnych pozycji w strukturze tworzonej przez sekwencję a sposobem funkcjonowania cząsteczki. Takie problematyczne sytuacje mogłyby zostać rozwiązane jedynie w trakcie szczegółowych i długotrwałych badań biochemicznych zaprojektowanych sekwencji. W niniejszej pracy, za pomocą programu AmiRNA Designer, wygenerowane zostały termodynamiczne profile dla miRNA:miRNA\* i miRNA:target dla *Arabidopsis thaliana*. Na podstawie uzyskanych danych zaprojektowano dwa niezależne zestawy potencjalnych amiRNA do wyciszenia dwóch przykładowych genów z *Arabidopsis thaliana*. Zaprojektowane sekwencje amiRNA i amiRNA\* analizowano za pomocą skryptów napisanych w języku Perl dla efektu niespecyficznego oddziaływania. Ponadto zaproponowano szkielet dla eksperymentalnego otrzymania amiRNA, który został wstępnie przetestowany pod kątem funkcjonalności.

Wyciszanie genów jest szeroko eksplorowanym obszarem badań naukowych, zarówno w badaniach podstawowych jak i aplikacyjnych. Projektowanie małych niekodujących RNA z wykorzystaniem narzędzi obliczeniowych znajduje zastosowanie w obu powyższych obszarach biologii. Program AmiRNA Designer to narzędzie bioinformatyczne, które umożliwia zaprojektowanie sekwencji sztucznych miRNA, przeznaczonych do umieszczenia w wektorze plazmidowym według opisanego powyżej schematu. Opracowany szkielet wektora oraz sposób jego użycia umożliwia wykorzystanie dostępnych komercyjnie syntetycznych fragmentów DNA, bez konieczności amplifikacji i klonowania genu będącego przedmiotem zainteresowania. AmiRNA Designer generuje sekwencje amiRNA oraz amiRNA\*, dla zadanej sekwencji genu, którego ekspresja ma zostać obniżona za pomocą mechanizmu RNAi. W celu zwiększenia podobieństwa termodynamicznego zaprojektowanych struktur do występujących w naturze została zaimplementowana metoda relaksacji przez wprowadzenie substytucji

nukleotydów, osłabiających wiązanie amiRNA:amiRNA\* oraz amiRNA:target w miejscach, gdzie było ono zbyt silne. Oparcie na zbiorze sekwencji wskazanych w etapie analizy pozwala na wyznaczenie profilu termodynamicznego dla danego zestawu i powoduje, że nowo zaprojektowane sekwencje tworzą struktury podobne do wzorca, co implikuje ich przetwarzane tak jak sekwencji natywnych, nawet jeśli nie znane są jeszcze wszystkie szczegóły mechanizmu RNAi. Prawdopodobieństwo oddziaływania zaprojektowanych amiRNA oraz amiRNA\* poza oczekiwanym miejscem docelowym jest weryfikowane za pomocą dedykowanych skryptów napisanych w języku Perl. AmiRNA Designer został opracowany dla roślin, ale pozwala użytkownikowi na ustawienie własnych parametrów i wykorzystanie danych z innych organizmów do generowania profili termodynamicznych dowolnych sekwencji.

#### 4.2. Modelowanie białka AtDCL4

Drugim aspektem rozważań na temat strukturalnych uwarunkowań procesu biogenezy małych regulatorowych RNA podjętych w ramach niniejszej pracy była analiza budowy białka AtDCL4. Przedmiotem badania było modelowanie struktury białka odpowiedzialnego za hydrolizę dwuniciowego RNA (w czego efekcie powstają cząsteczki miRNA:miRNA\*) w celu określenia strukturalnych cech dopasowania enzym-substrat dla tej reakcji.

Białko typu Dicer z *Arabidopsis thaliana* oznaczone jako AtDCL4 wycina głównie 21 nukleotydowe duplekisy RNA o niemalże idealnej komplementarności dopuszczając jednakże niekanoniczne pary zasad G-U, często występujące w strukturach RNA. Pomimo wielu prac eksperymentalnych dotyczących funkcji roślinnych białek DCL, nadal niewiele wiadomo o ich budowie. Poznanie struktury jednego z nich (a w dalszej przyszłości określenie także struktury pozostałych) może być pomocne w wyjaśnieniu wielu wyników badań biochemicznych. Pytanie dlaczego rośliny używają co najmniej czterech białek tego typu, podczas gdy u zwierząt odnajdujemy zwykle jedno – co najwyżej dwa – nadal pozostaje niewyjaśnione. Roślinne białka DCL składają się z 10 domen (por. Rys. 1.6), z których tylko dla kilku dostępne są poznane struktury. Nie dla wszystkich udało się także doświadczalnie określić ich funkcję, dlatego też w tej pracy skupiono się na części białka AtDCL4 odpowiadającej najmnijemu białku typu Dicer pochodzącemu z *Giardia intestinalis* (GiDCR). Białko GiDCR posiada jedynie 5 domen, które wystarczają by wycinało ono z długiego dsRNA fragmenty o długości ok 21 nt

odpowiadające cząsteczkom miRNA:miRNA\*. Analogiczny fragment cząsteczki białka AtDCL4 określić można jako „rdzeń katalityczny” (por. Rys. 4.16) - opracowanie jego modelu dałoby podstawy do ukierunkowania badań eksperymentalnych w przyszłości.

#### 4.2.1. Analiza sekwencji białek Dicer i DCL

Pierwszym krokiem w kierunku modelowania homologicznego AtDCL4 było przyrównanie wielu sekwencji (MSA, (ang. Multiple Sequence Alignment) z roślinnych białek DCL. Wyszukiwanie w bazie danych PlantGDB sekwencji białek DCL zakodowanych w genomach roślinnych dało ponad 100 wyników z wartością e-value poniżej  $1e^{-20}$ . Usunięto z tego zestawu sekwencje z silnym podobieństwem tylko krótkiego fragmentu (<500 aa), co sugeruje obecność jedynie domeny homologicznej, ale nie to, że sekwencja koduje całe białko typu Dicer. Pozostałe 75 rekordów opatrzone były jedynie numerem transkryptu, a nie przynależnością do konkretnego rodzaju DCL. Zatem za pomocą narzędzia NCBI BLAST i analizy filogenetycznej przeprowadzono przypisanie ich przynależności do poszczególnych klas rodziny RNazy III. Jako odniesienia użyto sekwencji białkowych *A. thaliana*: cztery białka DCL i dwa białka RTL (ang. *RNase III Like Proteins*, 1 i 2), które zawierają jedną domenę RNazy III i dwie domeny dsRBD. Są one pokrewne do DCL (Hiraguri 2005), lecz funkcjonują w innych szlakach biochemicznych. Analiza filogenetyczna wykazała, że wszystkie badane sekwencje były skupione w sześciu grupach odpowiadających sekwencjom odniesienia, 20 z nich stanowiło jednakże białka RTL i te zostały wyeliminowane z dalszej analizy. Ostatnim sitem dla pozostałych 55 sekwencji była ponownie ich długość. 15 z nich – pomimo dodatkowych domen sugerujących, że mogą one być białkami DCL – nie obejmowała rozważanego zakresu białka DCL od platformy do drugiej domeny RNazy. Ostatecznie analizowano 40 sekwencji z następujących gatunków: *Brachypodium distachyon* (stokłosa) (*Bdi*), *Glycine max* (soja) (*Gma*), *Medicago truncatula* (lucerna) (*MTR*), *Oryza sativa* (ryż) (*Osa*), *Populus trichocarpa* (topola) (*Ptr*), *bicolor sorgo* (sorgo) (*Sbi*), *Vitis vinifera* (winorośl) (*Vvi*) i *Zea mays* (kukurydza) (*Zma*). Lista ta została później uzupełniona sekwencjami białkowymi DCL1, DCL2 i DCL4 z *Nicotiana attenuata* (*Nat*), z danych literaturowych, które były analizowane i zdeponowane w bazach UniProt i NCBI już po przeprowadzeniu naszych analiz (Bozorov 2012). Drzewo filogenetyczne otrzymane z 43 analizowanych sekwencji znajduje się w załączniku nr 4 na końcu niniejszej rozprawy.



Ten sam zestaw sekwencji z roślinnych DCL analizowano pod kątem zidentyfikowania w dopasowaniu MSA motywów zakonserwowanych, insercji i delecji oraz określenia regionów pętli niezachowawczych. Wszystkie te informacje są bardzo istotne dla poprawy późniejszego dopasowania docelowej sekwencji AtDCL4 do potencjalnych szablonów strukturalnych. Dopasowanie MSA 43 sekwencji roślinnych obejmujących fragment odpowiadający modelowanej sekwencji AtDCL4 znajduje się w załączniku nr 5 na końcu niniejszej rozprawy. Otrzymane wyniki wykazały 42-48% podobieństwa sekwencji i 24-28% jej identyczności między AtDCL4 a pozostałymi AtDCL1-3 w obrębie badanego rdzenia katalitycznego. Największa zmienność cechuje domenę platformy, która wykazała 23-40% podobieństwa przy zaledwie 9-18% identyczności oraz 23-40% luk. Białko DCL1 ma dużą insercję około 50 aminokwasów, zlokalizowaną między platformą a domeną PAZ. Podobieństwo domeny PAZ wszystkich białek AtDCL była nieco większa i bliska 37-51% natomiast luki zostały zredukowane do 3-14%. Dla domeny RNazy III A (nazywanej dalej RIIIA, reszty 1089-1264 z AtDCL4) podobieństwo sekwencji wyniosło 51-58% przy 2-7% luk. Najbardziej konserwatywna była domena RNazy III B (analogicznie – RIIIB, reszty 1291- 1464 z AtDCL4), gdzie podobieństwo sekwencji białek AtDCL wynosiło 52-57% niemal bez luk.

Przygotowując dopasowanie MSA 43 sekwencji roślinnych pobrano z bazy danych NCBI sekwencję białka DCL2 z *Nicotiana attenuata* (Bozorov 2012) (Numer identyfikacyjny NCBI: JN032014). Sekwencja ta różniła się znacząco od pozostałych białek DCL2 na długości 45 aminokwasów, w regionie wysoce zachowawczym, stanowiącym fragment miejsca katalitycznego. Uzyskana została ona w procesie translacji *in silico* po sekwencjonowaniu kodującego DNA. Wiele metod, jakimi dokonuje się sekwencjonowania jest podatne na błędy w monotraktach – miejscach, gdzie ten sam nukleotyd występuje kilka razy – zdarza się to najczęściej w przypadku adeniny. Aminokwasy, zarówno rozpoczynający jak i kończący, różniącą się sekwencję pochodziły z miejsc gdzie zidentyfikowano monotrakt adeninowy. Ich korekta – skrócenie pierwszego (z 6 do 5) oraz wydłużenie drugiego z nich (z 3 do 4) (Rys. 4.15.A) spowodowały, że sekwencja białka stała się niemal identyczna z innymi białkami roślinnymi DCL2 (Rys. 4.15.B,C).

## A.

```

Nat_DCL2 (971-1033)   ATGCACTCGTATCAGTGCACGCAAAAA-CATTTTTTATCCAACCGCTAAGGTTCTGGAAGCAA
Nat_DCL2_mod         ATGCACTCGTATCAGTGCACGCAAAAAACATTTTTTATCCAACCGCTAAGGTTCTGGAAGCAA
*****
Nat_DCL2 (971-1033)   TTACAACAAGAAATGCGTCGAAAAGTTTCATTTGGAATCACTTGAGACACTGGAGACTCTTT
Nat_DCL2_mod         TTACAACAAGAAATGCGTCGAAAAGTTTCATTTGGAATCACTTGAGACACTGGAGACTCTTT
*****
Nat_DCL2 (971-1033)   TCTCAATATGCTGCGAGTATACAGCTGTTCAAGACTTATGAAAATACATCATGAGGGTCTA
Nat_DCL2_mod         TCTCAATATGCTGCGAGTATACAGCTGTTCAAGACTTATGAAA-TACATCATGAGGGTCTA
*****

```

## B.

```

*           20           *           40           *           60
Osa_DCL2a : LKVQLGPTV--QQFDVPLKILEALTTKKCOE EFSQESLETLGDSFLKYVTRHRLFSEYRLQHEGILT
Osa_DCL2b : LKVQLGPTV--QQFDVPLKILEALTTKKCOE EFSQESLETLGDSFLKYVTRHRLFSEYRLQHEGILT
Bdi_DCL2a : LKVQLGPRM--QQFIIPAVKILEAITTKCOE EFSQESLETLGDSFLKYVTTQHLYSKYKLLHHEGILT
Bdi_DCL2b : LKVQLGPRM--QQFSVPDLKILEALTTKCOE EFSQESLETLGDSFLKYVTTQHLTYTKYKLLHHEGTLT
Sbi_DCL2  : LKTQLGPSM--KQFDI PALKVLEALTAKKCKE EFSQESLETLGDSFLK-----A
Gma_DCL2a : LKKMHLDHCT--QNEIQTIKVLEAITTKRCNE AFHYESLETLGDSFLKYAASQQLFKTYKNHHEGLLS
Gma_DCL2b : LKKMHLDHCAR--NEIQTIKVLEAITTKRCNE AFHYESLETLGDSFLKYAASQQLFKTYHNHHEGLLS
Mtr_DCL2  : FKRMLHDYCPQ--NKIQTTFKVEAMTTKCKE TFFHYESLETLGDSFLKYAVSQQLFNMYQNHHEGLLS
Vvi_DCL2  : LKNIHLNYCKQ--NDIPTFKVLEAITTKHCOE GFHLESLETLGDSFLKYAASQQLFKTFQNHHEGLLS
Ptr_DCL2a : LKKMHS DHYLQ-HVNI PSMKVLEAITTKCOE NFNLESLETLGDSFLKYAASQQLFKIYQNHHEGLLS
Ptr_DCL2b : LKKMHS DHCMQ-NVDI PAMKVLEAITTKCOE KFHLESLETLGDSFLKYAASQQLFKIYQNHHEGLLS
Nat_DCL2  : LKRMHSYQCTQKHF LFP LRFWKQLQQNASKSFIWNHLRHL ETLFSNMLRVYSCSRLMKIHHHEGLLT
Ath_DCL2  : LKKS-----PKVNIPTIKVLEAITTKCEDQGFHLESLETLGDSFLKYAVCQQLFQHCHTHHEGLLS
LK                               4 lea6tt c e F esLetLgdsFlky 1 hheg 1

```

## C.

```

*           20           *           40           *           60
Osa_DCL2a : LKVQLGPTV--QQFDVPLKILEALTTKKCOE EFSQESLETLGDSFLKYVTRHRLFSEYRLQHEGILT
Osa_DCL2b : LKVQLGPTV--QQFDVPLKILEALTTKKCOE EFSQESLETLGDSFLKYVTRHRLFSEYRLQHEGILT
Bdi_DCL2a : LKVQLGPRM--QQFIIPAVKILEAITTKCOE EFSQESLETLGDSFLKYVTTQHLYSKYKLLHHEGILT
Bdi_DCL2b : LKVQLGPRM--QQFSVPDLKILEALTTKCOE EFSQESLETLGDSFLKYVTTQHLTYTKYKLLHHEGTLT
Sbi_DCL2  : LKTQLGPSM--KQFDI PALKVLEALTAKKCKE EFSQESLETLGDSFLK-----A
Gma_DCL2a : LKKMHLDHCT--QNEIQTIKVLEAITTKRCNE AFHYESLETLGDSFLKYAASQQLFKTYKNHHEGLLS
Gma_DCL2b : LKKMHLDHCAR--NEIQTIKVLEAITTKRCNE AFHYESLETLGDSFLKYAASQQLFKTYHNHHEGLLS
Mtr_DCL2  : FKRMLHDYCPQ--NKIQTTFKVEAMTTKCKE TFFHYESLETLGDSFLKYAVSQQLFNMYQNHHEGLLS
Vvi_DCL2  : LKNIHLNYCKQ--NDIPTFKVLEAITTKHCOE GFHLESLETLGDSFLKYAASQQLFKTFQNHHEGLLS
Ptr_DCL2a : LKKMHS DHYLQ-HVNI PSMKVLEAITTKCOE NFNLESLETLGDSFLKYAASQQLFKIYQNHHEGLLS
Ptr_DCL2b : LKKMHS DHCMQ-NVDI PAMKVLEAITTKCOE KFHLESLETLGDSFLKYAASQQLFKIYQNHHEGLLS
Nat_DCL2  : LKRMHSYQCTQKTF FFIPTAKVLEAITTKKCOE KFHLESLETLGDSFLKYAASIQLFKTYEIHHEGLLT
Ath_DCL2  : LKKS-----PKVNIPTIKVLEAITTKCEDQGFHLESLETLGDSFLKYAVCQQLFQHCHTHHEGLLS
LK                               6p K6LEA6Ttk C e F ESLETLGDSFLky 1 hheg 1

```

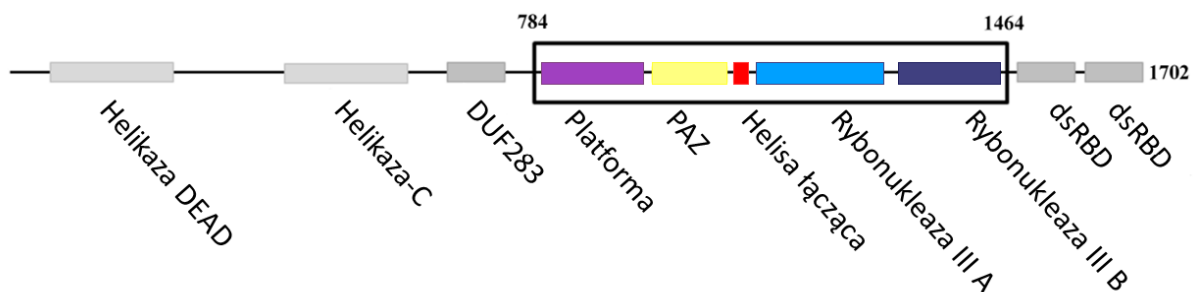
**Rys. 4.15.** A Zestawienie fragmentu sekwencji nukleotydowej Nat\_DCL2 oryginalnej i po modyfikacji ilości adenin w traktach (zaznaczone kolorem czerwonym). B. MSA roślinnych DCL2 z oryginalną sekwencją białka Nat\_DCL2. C. MSA roślinnych DCL2 z sekwencją białka Nat\_DCL2 otrzymaną po korekcie sekwencji genu.

Dopasowanie sekwencji GiDCR do profilu AtDCL1-4 pochodzących z MSA roślinnych białek DCL wykazało, że jego podobieństwo do AtDCL1 wyniosło 20% (10% identyczności) podczas gdy dla AtDCL2-4 podobieństwo oscyluje wokół 21% (przy 11% identyczności dla AtDCL2 i AtDCL4 oraz 9% identyczności dla AtDCL3). Najmniejsza liczba luk w dopasowaniu GiDCR i AtDCL4 dodatkowo uzasadnia słuszność wyboru właśnie tego DCL ze wszystkich

czterech białek typu Dicer pochodzących z *A. thaliana* jako celu modelowania homologicznego cząsteczki białka.

#### 4.2.2. Wyszukiwanie szablonów strukturalnych

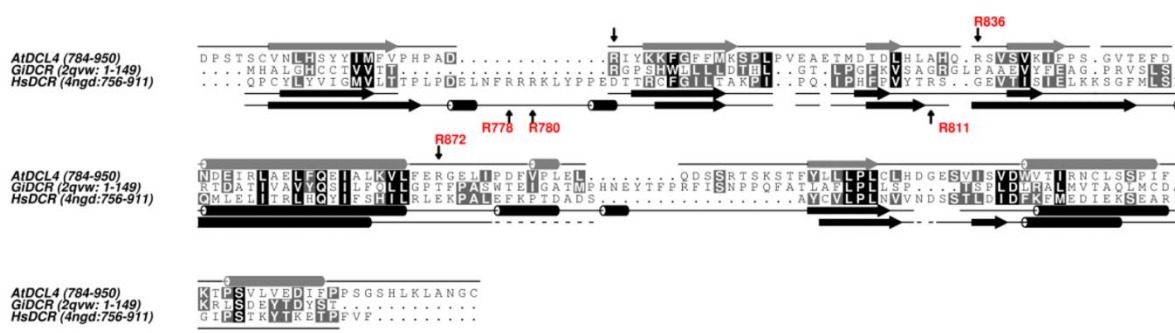
Ze względu na niezbyt duże podobieństwo sekwencji AtDCL4 do GiDCR (PDB ID: 2qvw), który początkowo był jedyną dostępną strukturą, mogącą służyć jako szablon dla modelowanego fragmentu, poszukiwano dodatkowych szablonów do modelowania poszczególnych domen. Zostaną one omówione w kolejności analogicznej do występującej w sekwencji białka (Rys. 4.16).



**Rys. 4.16.** Schematyczna reprezentacja budowy domenowej AtDCL4. Modelowany fragment zaznaczono prostokątem.

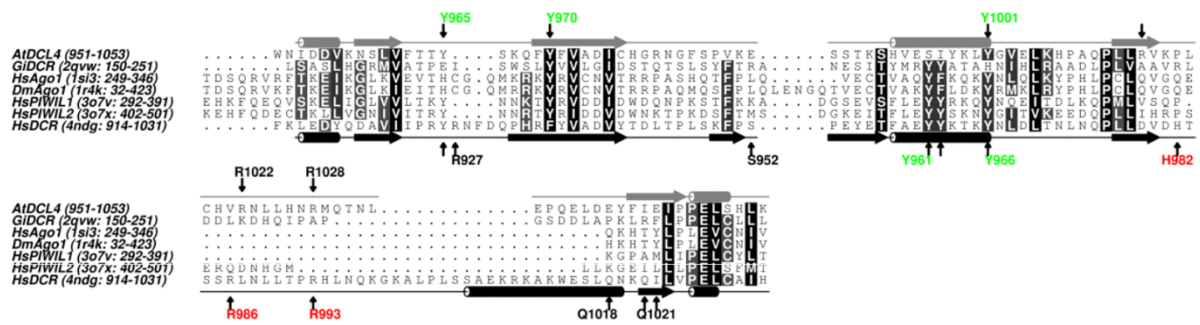
Wyszukiwanie w sekwencji platformy AtDCL4 (reszty 784-926) motywów strukturalnych lub sygnatur białkowych (ang. protein signature) za pomocą ogólnie dostępnych serwisów i oprogramowania nie dało zadowalających wyników. Stosunkowo od niedawna dostępne są struktury krystaliczne dla fragmentu obejmującego domeny PAZ, platformę i helisę łączącą białka HsDCR w kompleksach z dsRNA (PDB ID: 4ngb, 4ngc, 4ngd, 4ngf, 4ngg, 4nh3, 4nh5 i 4nh6) (Tian 2014). Porównanie domen platformy z GiDCR i HsDCR wykazuje, że struktury 3D są podobne (RMSD = 1,2 Å), pomimo niskiego podobieństwa sekwencji (25%). Ponadto przewidywania struktury drugorzędowej w tym regionie w czterech AtDCL wykazują charakterystyczny zwój  $\beta\beta\beta\beta\alpha\beta\alpha$ , podobnie jak dla struktur krystalicznych platformy GiDCR i HsDCR. Obserwacja ta sugeruje, że platforma roślinnych białek DCL, mimo niskiego stopnia podobieństwa sekwencji z platformą GiDCR i HsDCR, może mieć zachowaną strukturę, charakterystyczną dla białek Dicer i DCL. Dopasowanie strukturalne dwóch matryc

GiDCR (PDB ID: 2qvw) i HsDCR (PDB ID: 4ndg) oraz przewidzianych struktur drugorzędowych AtDCL4 zostało wykorzystane do skonstruowania dopasowania sekwencyjnego do modelowania homologicznego AtDCL4 (Rys. 4.17). W dostępnych strukturach krystalicznych, platforma otacza helisę łączącą tworząc z nią spójny blok strukturalny, podczas gdy w sekwencji obie domeny oddziela od siebie domena PAZ. Okazało się, że ważne jest jednak, aby modelować domenę platformy razem z helisą łączącą. Modele otrzymane w ten sposób wykazywały lepsze parametry w systemach oceniania niż takie, które powstały z prób modelowania każdej z tych dwóch domen osobno.



**Rys. 4.17.** Przyrównanie sekwencji aminokwasowych domen Platformy. Dwa dolne rzędy opisują motywy strukturalne obecne w poznanych strukturach białek GiDCR i HsDCR odpowiednio. Górny rząd symbolizuje predykcję struktury sekwencji AtDCL4. Poziome strzałki wskazują strukturę  $\beta$ -kartki, walce zaś  $\alpha$ -helisy. Pionowymi strzałkami zaznaczono miejsca wiązania RNA, reszty należące do kieszeni wiążącej 5' koniec RNA oznaczono czerwonym kolorem.

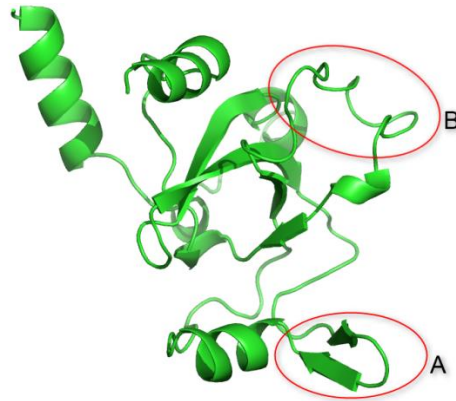
Domena PAZ, jako moduł wiążący dwuniciowe RNA jest stosunkowo często spotykana w wielu białkach. W niektórych z nich struktura PAZ została poznana czy to w formie samodzielnej domeny, czy w większych strukturach białkowych zawierających również inne domeny. Dodatkowym atutem rozwiązanych struktur jest to, że część z nich występuje jako kompleks z dsRNA, co stanowi wskazówkę w jaki sposób oddziałuje białko z RNA w przypadku modelowanego białka. Jednakże problemem jaki się ujawnił w projekcie zastosowania dostępnych struktur domeny PAZ, jako szablonu dla analogicznej domeny białka AtDCL4 było stosunkowo niskie podobieństwo ich sekwencji (Rys. 4.18).



**Rys. 4.18.** Przejście sekwencji aminokwasowych domen PAZ modelowanej sekwencji i rozwiązanych struktur jako szablonów. Helisy są oznaczone jako cylindry,  $\beta$ -kartki jako strzałki. Dolna linia wskazuje strukturę drugorzędową szablonu HsDCR. Miejsca wiązania RNA są zaznaczone pionowymi strzałkami, reszty z kieszeni wiążącej 5' są zaznaczone na czerwono, reszty z kieszeni wiążącej 3' – na zielono.

Za pomocą narzędzia do predykcji struktury konsensusowej PCONS5 udało się określić strukturę pochodzącą z GiDCR (PDB ID: 2qvw) jako najlepszy szablon do modelowania AtDCL4-Paz (927-1054 aa) (identyczność 14%, punktowany na 1,82). Kolejne szablony określone przez program PCONS5 pochodziły z domen PAZ białka Argonaute, czyli domeny PAZ z ludzkiego AGO1 (PDB ID: 1s13, identyczność 16%, punktowany na 1,77) i z muszki owocówki (*Drosophila melanogaster*) AGO1 (PDB ID: 1r4k, identyczność 14%, punktowany na 1,75). Wyniki predykcji z wykorzystaniem PCONS5 różnią się od siebie w zależności od długości sekwencji zadanej do analizy. Dlatego też ponowiono wyszukiwanie szablonów dla krótszego fragmentu AtDCL4-Paz (953-1054 aa), które potwierdziło wybór struktury GiDCR, jako najlepszego szablonu (identyczność 17%, punktowany na 2,02). Największą identyczność (28%) wykazywała struktura PAZ z ludzkiego białka-typu-PIWI-1 (Hs PIWIL1 PDB ID: 3o7v), jednakże jego punktacja wynosiła zaledwie 1,68. Analizy wykonane za pomocą metaserwerów nie uwzględniły struktur PAZ białka HsDCR. Podobieństwo ich sekwencji z sekwencją PAZ z AtDCL4 wynosiło 35% (podobnie z GiDCR), jednakże 14% sekwencji pozostawało bez szablonu (tylko 6% w przypadku GiDCR). Aby zwiększyć poprawność szablonu dla modelowania domeny PAZ, najpierw wykonano dopasowanie strukturalne wybranych struktur krystalicznych domen PAZ białka AGO i białek typu PIWI pochodzących z GiDCR oraz HsDCR (za pomocą VMD), a następnie porównano je z dopasowaniem wielu sekwencji (MSA) roślinnych domen PAZ. Takie porównanie ujawniło dwa regiony, które odróżniają od siebie poszczególne rodzaje białek wiążących dwuniciowe RNA. Pierwszy obszar znajduje się w przybliżeniu w środku sekwencji PAZ i stanowi dwie anty-równoległe beta kartki. W strukturze znajduje się on w dolnej części

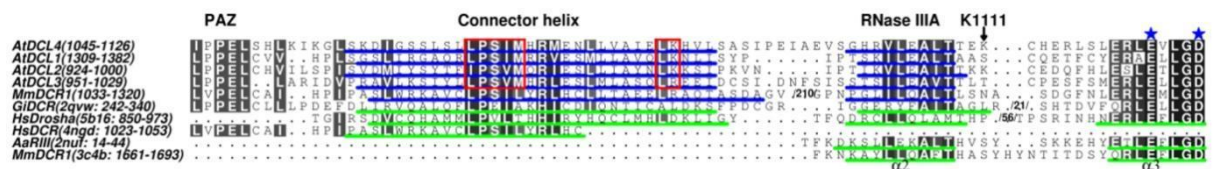
kieszeni wiążącej wystający 3' koniec dwuniciowego RNA (Rys. 4.19.A). Drugi rejon znajduje się w pobliżu C-końca i stanowi pętlę, która jest obecna w białkach typu Dicer, zarówno zwierząt jak i roślin. Pętla ta jest bardzo krótka w białkach typu PIWI i AGO (Rys. 4.19.B) (Mukherjee 2013, Simon 2011).



**Rys. 4.19.** Struktura domeny PAZ pochodząca z białka GiDCR. Zaznaczono charakterystyczne motywy różniące poszczególne rodzaje białek zawierających domenę PAZ. **A.** Element kieszeni wiążącej 3' koniec dsRNA. **B.** Dodatkowa pętla o nieznannej funkcji.

Biorąc pod uwagę strukturę GiDCR, można zauważyć, że odległość pomiędzy miejscami wiązania końca 3' RNA w domenie PAZ a miejscami aktywnymi w domenach RNazy III odpowiada długości wycinanych krótkich dsRNA. Tak więc sekwencja łącząca domeny PAZ i RIIIa była analizowana szczególnie starannie (Rys. 4.16). Co ciekawe, ten fragment różni się długością u poszczególnych taksonów. Na przykład: u DCL pochodzących z *Arabidopsis thaliana* występuje tu mniej niż 100 aa, natomiast w zwierzęcych białkach Dicer jest to ok. 300 aa, gdzie wyjątkiem jest DmDCR1 z ponad 500 aa. Struktura GiDCR w tym regionie zawiera helisę łączącą (ang. connector helix). Sugerowano, że mogłaby ona stanowić swego rodzaju „linijkę molekularną” do mierzenia efektywnej długości wytwarzanych dupleksów RNA (Lau 2012). W roślinnych białkach typu Dicer, odpowiedni region jest dobrze zachowany (ok. 80% podobieństwa). Motyw pochodzący z profilu roślinnych DCL (L/I)-PS-(I/L/V)-M odpowiada analogicznemu motywowi 265-LPEIA-269 w helisie GiDCR. Wysoce konserwatywna Prolina 266 (co odpowiada P1069 w AtDCL4) została wcześniej opisana jako swego rodzaju „zawias”, na którym występuje lekkie zgięcie helisy, istotnie ułatwiające elastyczność konformacyjną w GiDCR (MacRae 2006). Jednakże, dopasowanie sekwencji z GiDCR całej helisy łączącej do profilu roślinnych białek DCL wykazało dość słabe powinowactwo sekwencji z wyjątkiem pięciu położonych centralnie reszt. W celu lepszego wymodelowania struktury i określenia długości

tej helisy w AtDCL4 w pracy wykorzystane zostały programy JPRED i JUFO. Przewidziano za ich pomocą długość helisy z AtDCL1-3 oraz białka Dicer pochodzącego z myszy, aby znaleźć – jeśli istnieje – korelację pomiędzy długością helisy łączącej a długością wycinanego dupleksu RNA. Przewidywana długość helisy wynosi 30-31 aa dla różnych typów DCL. Nie zaobserwowano jej związku z długością fragmentów powstających dsRNA (Rys. 4.20). Ponadto przewidziana długość helisy łączącej z AtDCL4 (28 aa) jest podobna do analogicznej długości GiDCR (30 reszt aminokwasowych), chociaż dwuniciowy RNA, otrzymywany jako produkt GiDCR, ma 27 nukleotydów długości podczas, gdy AtDCL4 generuje fragmenty głównie 21 -nukleotydowe (Dunoyer 2010). Helisa łącząca jest poprzedzona pętlą, której sekwencja jest dobrze zachowana w obrębie każdego typu DCL, ale różni się znacznie pomiędzy tymi typami. Sugeruje to, że nie sama sztywna helisa stanowi molekularną liniijkę ale istotny w tym udział ma również struktura jaką przyjmują pętle ją otaczające.

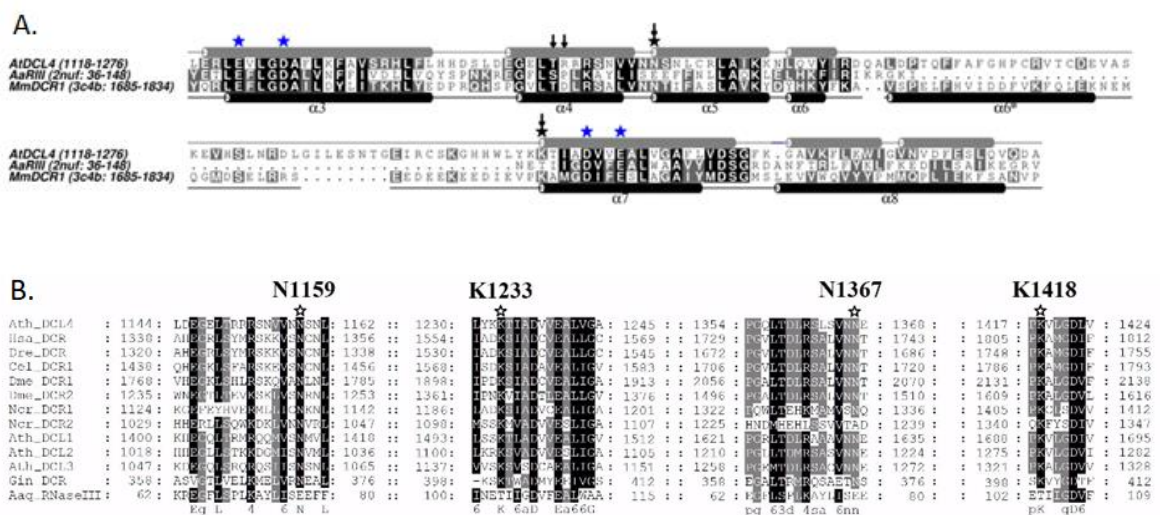


**Rys. 4.20.** Przyrównanie sekwencji aminokwasowych helis łączących z białkami roślinnymi AtDCL oraz dostępnymi szablony strukturalnymi. Niebieskim kolorem zaznaczono przewidywane długości helis, zielonym – obserwowane w strukturach szablony, gwiazdkami – motyw SM1A.

Helisa łącząca w GiDCR od strony C-końca jest połączona przez zwój strukturalny "pętla – krótka helisa – pętla" z "antyrownoległą" helisą stanowiącą pierwszy motyw strukturalny domeny RIIIA (SM1A, ang. signature motif) w miejscu aktywnym. Fragment o długości 9 reszt aminokwasowych, który odpowiada krótkiej helisie ( $\alpha 2$ ), jest dobrze zakonserwowany w obrębie całego dopasowania DCL i MmDCR (Rys. 4.20.).

Strukturalne dopasowania domeny GiDCR RIIIA do AaRNase III (a także do RIIIB pochodzących z HsDCR i MmDCR) (przy dokładnym nałożeniu „krótkiej helisy” we wszystkich tych strukturach) wykazały, że obie pętle otaczające krótką helisę różnią się zarówno między gatunkami jak i rodzajami DCL. Pierwsza pętla, poprzedzająca krótką helisę, jest nieco dłuższa u AtDCL niż odpowiedni fragment w GiDCR (o 9 aa). Pozostałe zwierzęce białka Dicer mają w tym miejscu długą sekwencję, których brakuje u roślinnych DCLi oraz GiDCR. Sugeruje się, że może ona pełnić funkcję "linijki", która oddziela domeny PAZ i RNazy III (Lau 2012) (patrz

rozdział 3.3). W naszym modelu ten region pełni rolę zawiasu zmieniając wzajemne położenie przylegających domen. Druga pętla, która następuje po „krótkiej helisie” ma taką samą długość we wszystkich typach roślinnych białek DCL. Jednakże, istnieją różnice w szablonach strukturalnych: w GiDCR pętla ta jest dłuższa o 13 reszt niż w AaRNase III. Te dodatkowe reszty w GiDCR tworzą krótkie wstęgi  $\beta$ , która oddziałują z platformą, powiększając jej arkusz  $\beta$ , co służy stabilizacji całej struktury białka. Dopasowanie sekwencji pochodzącej z AtDCL4 do profilu skonstruowanego z trzech szablonów (AaRNase III, GiDCR i MmDCR) wykazało większe podobieństwo do RNazy III z bakterii (Rys. 4.18.B), jednakże brak pętli w domenie bakteryjnej spowodował, że z lepszym wynikiem domena ta była modelowana wraz z domeną RIIIA na podstawie szablonu MmDCR (PDB ID: 3c4b). Ostatecznie modelowanie doprowadziło do uzyskania pętli (pomiędzy  $\alpha 6$  i  $\alpha 7$ , Rys. 4.21.A) nieznacznie dłuższej niż w MmDCR ale krótszej niż w GiDCR.



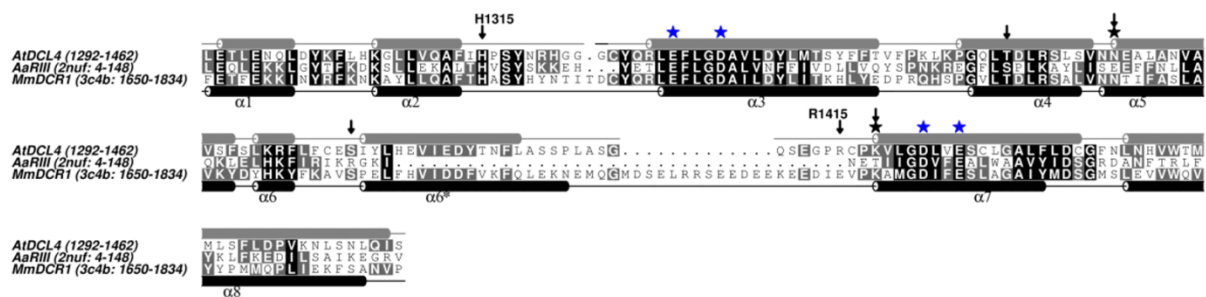
**Rys. 4.21.** A. Przycięcie sekwencji aminokwasowych domen RNazy III A modelowanego białka oraz szablonów pochodzących z białek GiDCR ora MmDCR. Helisy są oznaczone jako cylindry. Dolne linie wskazują strukturę drugorzędową szablonów. Miejsca wiązania RNA są zaznaczone pionowymi strzałkami. Gwiazdki wskazują reszty katalityczne. B. Przycięcie sekwencji aminokwasowych białek roślinnych DCL oraz bakteryjnej RNazy III w regionach oddziałujących z regionem katalitycznym. Tło w odcieniach szarości oznacza stopień zakonserwowania danej reszty na określonej pozycji w sekwencji.

Pierwsza domena RNazy AtDCL4 (reszty 1090-1274, RIIIA) wykazuje 23% podobieństwa i 30% luk do odpowiedniego fragmentu GiDCR, natomiast porównanie z domenami RIIIA pozostałych kręgowców wykazuje ok 33% podobieństwa i ok 30% luk. Dopasowanie sekwencji domen AtDCL4 RIIIA i MmDCR1 RIIIB wykazuje 37% podobieństwa i 16% luk. Również analiza



PCONS5 potwierdziła, że najlepszym szablonem strukturalnym dla AtDCL4 RIIIA jest struktura MmDCR1 RIIIB (PDB ID: 3c4b) (wynik PCONS5: 3,6177). Drugim najlepszym szablonem jest struktura rybonukleazy III z *Thermotoga marine* (PDB ID: 1o0w) z wynikiem PCONS5 3,4626, podczas gdy trzeci najlepszy szablon to struktura RNazy III z *Aquifex aeolicus* (PDB ID: 2nuf, PCONS5 zdobył wynik 3,2859). AtDCL4 RIIIA zawiera insercję umieszczoną w N-końcowym fragmencie drugiego motywu katalitycznego (SM2A, reszty 1190-1232). Sekwencja ta nie jest zachowana pomiędzy różnymi typami roślinnych DCL, a jej długość w tym regionie różni się do 23 reszt aminokwasowych. W AtDCL4 fragment ten zawiera 54 aminokwasów, co zbliża je najbardziej do DCL3, natomiast białko DCL1 posiada najdłuższą sekwencję w tym regionie (63 aminokwasów), podczas gdy najkrótsza sekwencja występuje u DCL2 (40 aminokwasów). Ten fragment AtDCL4 nie ma swoich odpowiedników w żadnej ze znanych struktur RNaz bakteryjnych. Ponadto jest dłuższy i wykazuje znacznie mniejsze podobieństwo do fragmentu występującego w analogicznym miejscu w sekwencji pochodzącej z myszy (MmDCR1). Analiza dokonana narzędziami przewidującymi uporządkowanie sekwencji białkowej ujawniła, że region 1203-1218 raczej wykazuje tendencję do przyjmowania struktury nieuporządkowanej niż do uformowania jakiegokolwiek rodzaju zwoju. Próby odnalezienia wzorca strukturalnego tylko dla tego regionu (reszty 1175 - 1235) także nie przyniosły pozytywnych rezultatów. W GiDCR ten obszar odpowiada pętli (389 - 400), która nie jest widoczna w strukturze krystalicznej, co może oznaczać, że jest nieuporządkowana i dopiero w kompleksie z substratem RNA przyjmuje funkcjonalną uporządkowaną strukturę. Pętla ta została nazwana „pętlą pozycjonowania”, a badania biochemiczne ujawniły, że jest ona odpowiedzialna za utrzymanie prawidłowej wielkości generowanych dsRNA (MacRae 2007). Wykazano, także, że w AtDCL1 mutacje w tym regionie domeny RIIIA mają wpływ na wydajność przetwarzania pri-miRNA (Liu 2012b). Żadna z badanych (zmutowanych) reszt w AtDCL1 nie jest zachowana w DCL4. Trzy z nich są zakonserwowane tylko w roślinnych białkach DCL1, podczas gdy kolejna mutacja (numeracja R1495, DCL1), umieszczona najbliżej SM2A, jest zakonserwowana w DCL1, DCL2 i DCL3 ale nie w DCL4 (Liu 2012b). W HsDCR ten fragment ma około 127 aminokwasów i został zidentyfikowany jako "miejsce wiązania" białek AGO (Sasaki 2007). Jednakże zachowawczy rejon odpowiedzialny za oddziaływanie z białkami AGO jest obecne tylko u białek typu Dicer u kręgowców. Z powodu braku wiarygodnej informacji pochodzącej z szablonów strukturalnych sekwencję w obszarze reszt 1178-1232 zamodelowano jako pętlę, która była odcinana przed analizami oceniającymi poprawność końcowego modelu.

Druga domena RNazy AtDCL4 (reszty 1291-1364, RIIIB) jest najbardziej podobna do analogicznych domen obecnych u innych białek z rodziny Dicer. Wykazuje 34% podobieństwa i 23% luk w porównaniu do tej samej domeny z GiDCR, natomiast porównanie jej do domen RIIIB białek typu Dicer pochodzących od kręgowców wykazuje ponad 50% podobieństwa i 8% luk. Dopasowanie sekwencji domen AtDCL4 RIIIB i MmDCR1 RIIIB pokazuje 57% podobieństwa i 8% luk, tak więc tę część struktury wymodelowano stosując krystaliczną strukturę 3c4b jako szablon (Rys. 4.22).



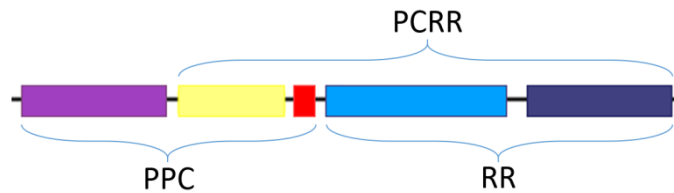
**Rys. 4.22.** Przyrównanie sekwencji aminokwasowych domen RNazy III B modelowanego białka oraz szablonów pochodzących z białek AaRNaseIII oraz MmDCR1. Helisy są oznaczone jako cylindry. Dolne linie wskazują strukturę drugorzędową szablonów. Miejsca wiązania RNA są zaznaczone pionowymi strzałkami. Gwiazdki wskazują reszty katalityczne. Tło w odcieniach szarości oznacza stopień zakonserwowania danej reszty na określonej pozycji w sekwencji.

#### 4.2.3. Modelowanie homologiczne białka

Jako że określiliśmy kilka szablonów do modelowania mniejszych fragmentów, ostatecznie dla fragmentu obejmującego obie domeny RNazy AtDCL4 RIIIA-RIIIB użyliśmy wzorca zbudowanego z dwóch jednakowych domen 3c4b (MmDCR1) ustawionych względem siebie w sposób, jaki występują w strukturze RNazy bakteryjnej 2nuf (AaRNaseIII). Jednocześnie struktura 2nuf stanowi kompleks RNazy III-RNA w postaci katalitycznej, została zatem użyta jako szablon także dla oddziaływań AtDCL4-RNA.

Podsumowując, modele homologiczne zostały uzyskane niezależnie dla poszczególnych bloków: platformy – PAZ – helisy łączącej (PPC) oraz dimeru RNazy: RIIIB RIIIA (RR). Dopiero w ostatecznym modelu wykorzystana została struktura 2qww (GiDCR), która

wraz z oddzielnie modelowanym modułem RIIIA-RIIIB została użyta do modelowania fragmentu PAZ-helisa łącząca-RIIIA -RIIIB (PCRR) (Rys. 4.23.).



**Rys. 4.23.** Modelowane fragmenty białka AtDCL4.

#### 4.2.4. Modelowanie kompleksu białko – RNA

Modelowanie AtDCL4 w kompleksie z dsRNA prowadzono przy następujących założeniach: (i) wiązanie dsRNA do domeny RNazy III zachowane jest w analogiczny sposób jak w kompleksie dimeru RNazy III z *A. aeolicus* z dsRNA (PDB ID: 2nuf (Gan 2008)), (ii) 2-nt wystający koniec 3' dsRNA jest wiązany w domenie PAZ, (iii) AtDCL4 generuje fragmenty idealnie dopasowanego dsRNA o długości 21-nt. Jako modelu dsRNA użyto dwóch struktur: dupletu o długości 32-par zasad zbudowanego w postaci kanonicznej A-RNA oraz fragmentu heliakalnego RNA pochodzącego z kompleksu RNazy III-RNA z *A. aeolicus* (PDB ID: 2nuf), wydłużonego fragmentem dupletu w kanonicznej formie A-RNA. Ułożenie wystających 2-nt na 3' końcu było wzorowane na podstawie RNA w kompleksie z domena PAZ białka HsAgo1 z HsAgo1 powiązanego z domeną PAZ (PDB ID: 1si3). Końce 5' otrzymały reszty fosforanowe.

Modelu strukturalnego PCRR użyto do budowy wstępnego kompleksu z RNA. Domeny RNaz w budowanym modelu AtDCL4 zostały nałożone na strukturę szablonu dimeru RNaz III w kompleksie z RNA (PDB ID: 2nuf) przez minimalizację RMSD jedynie reszt znajdujących się w miejscach aktywnych białka a nie całości cząsteczki. Miało to na celu zapewnienie lepszego dopasowania RNA do miejsc katalitycznych. Łańcuchy RNA z 2nuf zostały przeniesione do modelu PCRR, a następnie przedłużone strukturą A-RNA w kierunku domeny PAZ do momentu osiągnięcia pożądanej długości produktu AtDCL4, czyli dwóch nici o długości 21 nt. Jednakże w otrzymanym modelu wystający koniec 3' dsRNA nie trafiał do kieszeni PAZ (Rys. 4.24.), co oznaczało, że niezbędne było znaczne przesunięcie domeny PAZ, by RNA znalazło się

w kieszeni wiążącej białka. Jest to uzasadnione tym, że GiDCR, główny szablon modelu PCRR, generuje dłuższe fragmenty dsRNA (25-27 nt) (McRae 2006) niż AtDCL4 (21 nt). Ponadto, wcześniej wykazano, że wiązanie dsRNA do GiDCR wymaga elastyczności konformacyjnej zarówno białka jak i RNA (McRae 2006).

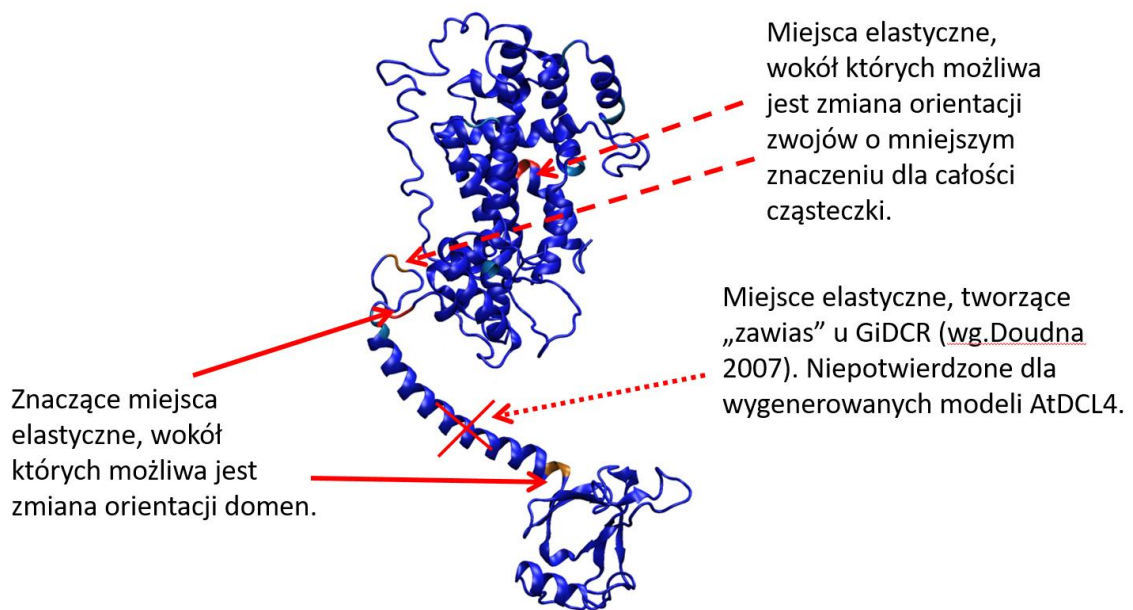


**Rys. 4.24.** Porównanie ułożenia domen PAZ i RNaz we wstępnej wersji modelu AtDCL4 (niebieski) nałożone na struktury krystaliczne kompleksów z RNA 1si3 (czerwony) oraz bakteryjnej RNazy III 2nuf (bordowy). Dwuniciowe RNA przedłużono fragmentem A-RNA do długości 21 nt z wystającym 3' 2-nt końcem (fioletowy). W takim układzie konieczna jest rearanżacja modelowanego białka, tak by RNA trafiła do kieszeni domeny PAZ – kotwiczącej 3' koniec RNA, natomiast w odległości 21 nt znajdowało się w pobliżu miejsca katalitycznego.

Użyliśmy dwóch podejść, aby umieścić domenę PAZ w pozycji pozwalającej na interakcję z wystającym 2nt 3'końcem RNA. Pierwsza metoda polegała na wykorzystaniu analizy drgań normalnych (NMA, ang. Normal Mode Analysis) zarówno cząsteczki dsRNA jak i modelu AtDCL4. Struktury z puli otrzymanej w wyniku analizy NMA dla PCRR były następnie nakładane na domenę RNaz z kompleksu 2nuf. Analogicznie potraktowane zostało dsRNA, z czego powstał zestaw struktur, który został nałożony na RNA z tego samego kompleksu. Uzyskane kompleksy AtDCL4-dsRNA, które wykazywały najlepsze kontakty pomiędzy końcem

3 'RNA i kieszeni wiążącej domeny PAZ wybrano do dalszej analizy (model PCRR1-RNA1). W drugim podejściu założyliśmy, że domena PAZ wiąże wystający koniec 2-nt RNA w podobny sposób jak PAZ pochodzący z HsAgo1. Struktura białka została poznana jako kompleks z RNA, którym był dupleks RNA o długości 9-nt (PDB ID: 1si3 (Ma 2004)). Dopasowanie modelu AtDCL4 do takiego stanu, by jednocześnie oba końce dsRNA znajdowały się w miejscach wiązania wymagało wykonania ruchu domeny PAZ oraz helisy łączącej. Do zmiany położenia domen w początkowym modelu PCRR w kompleksie z dsRNA zostały wykorzystane miejsca po dwóch stronach helisy łączącej. Położenie domeny PAZ wyznaczono przez superpozycję RNA pochodzącego z wzorca 1si3 na koniec dsRNA związanego z modelem PCRR w obrębie domen RNaz. Następnie domena AtDCL4 PAZ została nałożona na domenę PAZ z 1si3. W ostatnim kroku helisa łącząca została ręcznie umieszczona w taki sposób, by łączyć obie domeny wiążące RNA. Wymagało to odbudowania wiązań w resztach wykorzystanych jako „zawiasy” i doprowadziło do uzyskania modelu PCRR2-RNA2. To podejście ma swoje uzasadnienie w poprzednich badaniach symulacji dynamiki molekularnej białka GiDCR (Sarżyńska 2010), w których zidentyfikowano miejsca po obu stronach helisy łączącej jako szczególnie podatne na zmianę struktury (Rys. 4.25).

Badania te wykazały w symulacjach MD, że RIIIA-RIIIB i PAZ stanowią dwa ruchome bloki względem platformy przy analizie całego białka GiDCR, natomiast w przypadku białka bez platformy sama domena PAZ wykazywała znaczną ruchomość. Ponadto, badania biochemiczne ujawniły, że ludzkie białko Dicer rozdzielone na dwa fragmenty w obszarze pomiędzy PAZ i RIIIA wykazywało aktywność (zachodziło cięcie dsRNA na produkty o określonej wielkości), a te fragmenty tworzą stabilny kompleks, nawet w przypadku nieobecności RNA (Ma 2012).



**Rys. 4.25.** Analiza HingeProt miejsc elastycznych w modelu białka AtDCL4. Wskazano miejsca elastyczne odmienne od tych sugerowanych w literaturze.

Modele generowane z użyciem NMA (PCRR1-RNA1) i w sposób manualny (PCRR2-RNA2) zostały wykorzystane jako rusztowania do utworzenia końcowych modeli obejmujących także domenę platformy (odpowiednio PPCRR1-RNA1 i PPCRR2-RNA2). W obu modelach domena platformy została wbudowana przez strukturalnie dopasowanie modułu PPC przez superpozycję jej domen PAZ i helisy łączącej do modeli PCRR1 i PCRR2. Dodatkowo przeprowadzono drobne manualne modyfikacje w celu uniknięcia kolizji sterycznych między domeną platformy a RNA. W przypadku modułu PPC2 platforma została nieco przesunięta w stosunku do domeny PAZ w celu lepszego dopasowania do dsRNA. Dostępne struktury zawierające obszar platformy i PAZ z HsDicer (Tian 2014) oraz GiDCR (McRae 2006) różnią się orientacją PAZ w stosunku do domeny platformy. Badania wykonane metodą analizy głównych składowych (PCA, ang. Principal Component Analysis) dla czterech różnych łańcuchów dostępnych w strukturze pochodzącej z GiDCR także wskazują na elastyczność domeny PAZ. Te obserwacje sugerują, że orientacja przestrzenna PAZ względem domeny platformy może się różnić między białkami Dicer i DCL, a także różnych stadiach działania, w obrębie jednego białka. Tym samym uzasadnia to podejście do modelowania domeny platformy niezależnie, a następnie umieszczenia jej w wygenerowanym kompleksie białko-RNA.

W ostatniej analizie modeli wygenerowanych *in silico* – w formie kompleksu białka z dsRNA – została przeprowadzona krótka symulacja MD z więzami pozycyjnymi z zastosowaniem potencjału harmonicznego, założonymi na domenach RNaz oraz ciężkich atomach zasad RNA, aby umożliwić domenom platformy i PAZ dostosowanie się do struktury RNA. Symulacje przeprowadzono aż RMSD osiągnęła fazę *plateau* (2ns dla obu modeli). Podczas symulacji, drugorzędowa struktura domen platformy PAZ została dobrze zachowana. Dodatkowo tą samą metodą przeprowadzona została analiza trwałości białka w nieobecności dsRNA. Symulację MD o długości 1 ns przeprowadzono tylko na części białkowej po ekstrakcji ich kompleksu białko-RNA. Przygotowano struktury P1 i P2 pochodzące kompleksów PCRR1-RNA1 i PCRR2-RNA2, odpowiednio. Model białka P1 był nieco bardziej stabilny i po 1ns MD miał RMSD = 1,6 Å, natomiast w przypadku P2 RMSD wynosiło 2,2 Å, w odniesieniu do struktur w kompleksie z RNA. To, że globalna struktura białka pozbawiona RNA pozostaje dość stabilna nie oznacza, że przyłączenie RNA, nie ma wpływu na strukturę AtDLC4. Być może kolejne badania krystalograficzne oraz bioinformatyczne, pozwolą określić czy białka typu Dicer w stanie wolnym wykazują fluktuacje konformacji na tyle istotne, aby jedna z konformacji wykazywała znacznie większe powinowactwo do RNA, powodując jego związanie. Ponadto obecność dodatkowych domen białka lub partnerów może pomóc stabilizować wzajemne rozmieszczenie głównych domen AtDCL4. Konieczne są jednak dalsze badania, aby w pełni zrozumieć wiązanie dsRNA przez AtDCL4.

W celu oceny poprawności otrzymanych modeli AtDCL4 oraz metod zastosowanych w celu ich ulepszenia, modeli złożono do serwera analizującego jakość białka (ProQ, Wallner 2003). Według ProQ, najlepsze modele pochodzące wprost z Modellera zostały ocenione jako "dość dobry model" (przewidywane LGscore ok 2) po NMA i ręcznych manipulacjach przewidywany LGscore > 4 osiągnęła poziom "bardzo dobry model". Walidacja modeli przed i po symulacjach MD wykonana za pomocą programu ProQ pokazuje podobne parametry (Tab. 4.11), co oznacza, że modele mają wysoki stopień poprawności.

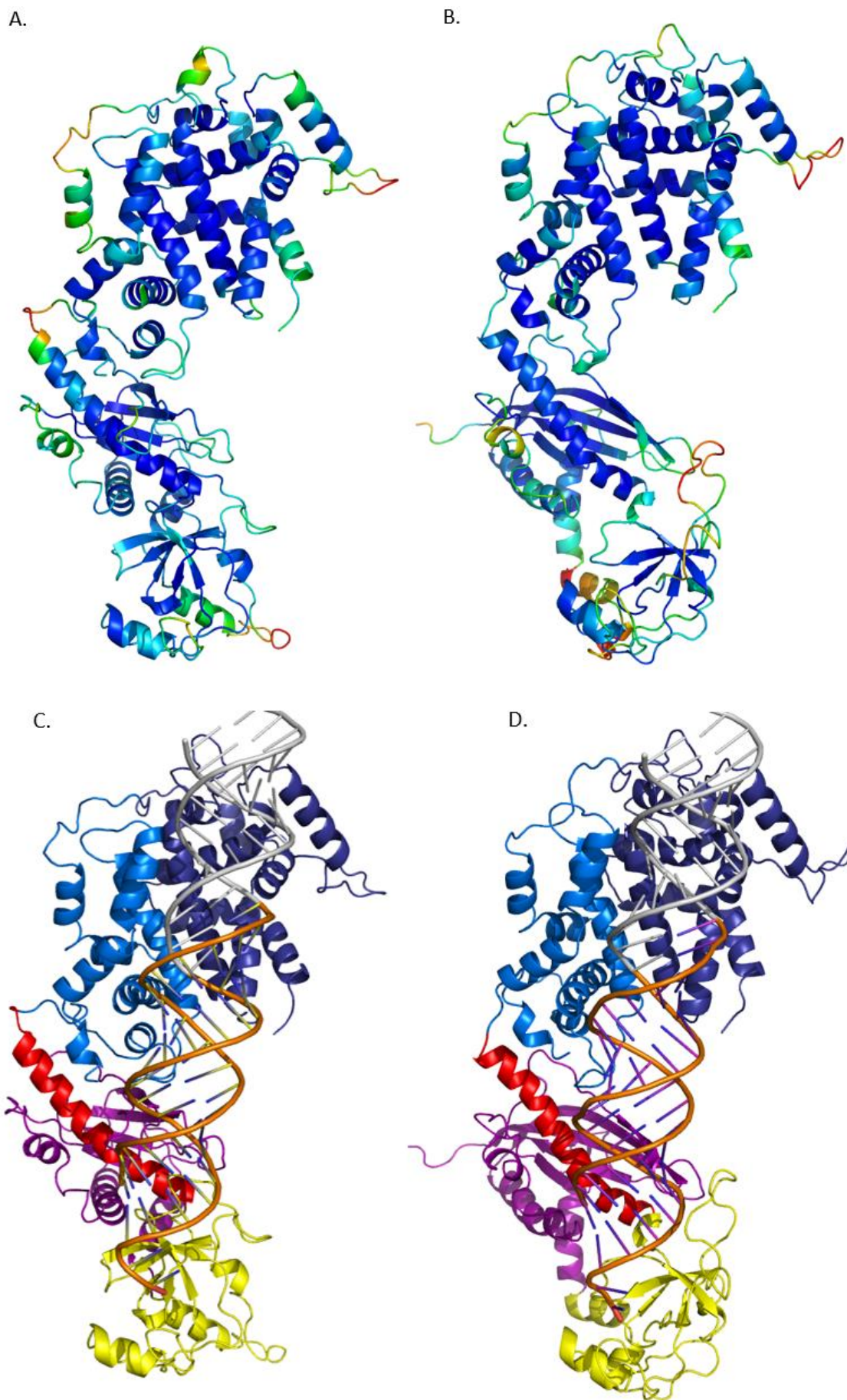
**Tab. 4.11.** Jakość modeli AtDCL4 określona przez program ProQ.

Model	LGscore	MaxSub
PPCRR1	5.101	0.209
PPCRR1_MD*	5.050	0.210
PPCRR2	4.436	0.210
PPCRR2_MD*	4.665	0.208

\* 2.0ns of MD

Innym sposobem sprawdzania jakości otrzymanych struktur było zastosowanie MetaMQAPII dostępnego na metaserverze Genesilico (Pawłowski 2008). Jest to narzędzie do oceny jakości struktur białka dla modeli teoretycznych. Oblicza ona globalną dokładność struktury białka, jak również dla poszczególnych reszt dodając obliczoną wartość „ $\beta$ -faktor” do pliku pdb. Rys. 4.26.A,B przedstawia modele PPCRR1 i PPCRR2 z uwzględnieniem obszarów o wyższych i niższych poziomach dokładności oznaczonych odpowiednio jako fragmenty niebieskie i czerwone (oraz odcienie zawarte pomiędzy tymi kolorami). Co ważne, regiony, które są funkcjonalnie istotne (na przykład reszty katalityczne czy reszty wiążące RNA oraz struktura we wnętrzu białka) charakteryzują się stosunkowo wysoką dokładnością, podczas gdy regiony określone jako mniej dokładne są usytuowane w mniej istotnych funkcjonalnie częściach struktury białka. Rdzenie strukturalne obydwu domen RNaz III jak i obszaru oddziaływań pomiędzy nimi mogą być traktowane jako pewne (przy szacowaniu dokładności poszczególnych reszt), co wskazuje na właściwe ułożenie w naszych modelach dwóch prostopadłych do siebie helis domen RNaz zaangażowanych w cięcie dsRNA. Wnętrze domeny PAZ zawierające "kieszonkę wiążącą" dwóch niesparowanych nukleotydów z 3' końca dupleksu dsRNA zostało ocenione jako korzystne, pomimo że zachowawcze reszty części tej kieszeni, a mianowicie: helisa i pętla (reszty 982-1002) wykazują niższy poziom dokładności. Może to jednakże wynikać z dozwolonej ruchomości tej części, która jest także widoczna w strukturach krystalicznych domen PAZ GiDCR, jak również w symulacjach MD i NMA. Struktury niektórych pętli i helis peryferyjnych należy traktować jako niepewne.

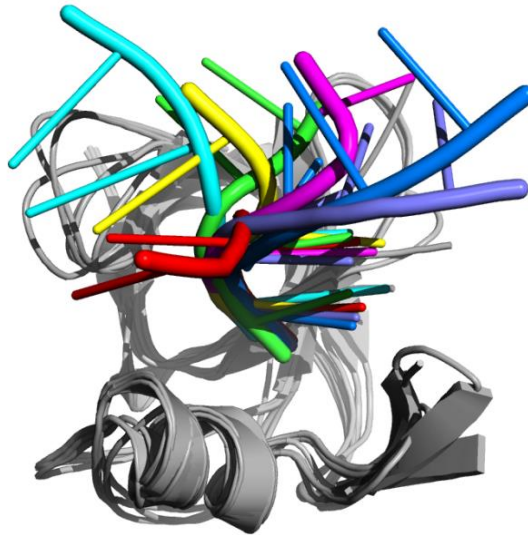




**Rys. 4.26. A, B.** Wyniki analizy jakości modeli AtDCL4 za pomocą serwera MetaMQAPI. Kolor niebieski wskazuje regiony o wysokiej jakości, natomiast czerwonej – niskiej. **C, D.** Kompleksy AtDCL4 z dsRNA, na szaro zaznaczono RNA pozostające po reakcji cięcia **A, C.** Model PPCRR1 – NMA, B, D. PPCRR2 – Model utworzony z zastosowaniem ręcznych manipulacji. **C, D.** Kolorowanie domen: platforma – fioletowy, domena PAZ – żółta, helisa łącząca – czerwona, RIIIA – błękitna, RIIIB – granatowa.

#### 4.2.5. Oddziaływanie białka z RNA

Oba modele białka, zarówno uzyskany z użyciem NMA jak i metodą ręcznej manipulacji, posiadają konformację umożliwiającą generowanie 21-nukleotydowych dupleksów dsRNA, zgodnie z danymi biochemicznymi, jednakże występują pomiędzy nimi pewne różnice. Główną różnicą pomiędzy modelami jest wzajemna orientacja dimeru RNazy III, platformy i domeny PAZ. Domeny te w otrzymanych modelach stanowią trzy ruchome bloki. Dostępne dane strukturalne nie są wystarczające do jednoznacznej orientacji przestrzennej domeny PAZ w stosunku do domen RNazy III. Chociaż położenie dsRNA w kieszeni katalitycznej może być określone jako wymodelowane z dużą pewnością, opartą na strukturach krystalicznych RNazy III, w postaci kompleksu z RNA (PDB ID: 2nuf, 4oog), istnieje wiele możliwości ułożenia dsRNA na powierzchni PAZ. Strukturalne nałożenie dostępnych domen PAZ w kompleksie z RNA pokazuje, że tylko pozycja 2-nt na 3' końcu w kieszeni wiążącej PAZ jest zachowana, natomiast pozostała część łańcucha RNA różni się znacznie pomiędzy strukturami (Rys. 4.27). Ostatnio opublikowana struktura fragmentu obejmującego fragment platforma-PAZ-helisa łącząca z HsDCR (Tian 2014) wykazała, że pętla PAZ może tworzyć spiralny zwój, który prawdopodobnie jest odpowiedzialny za regulację wiązania dsRNA do powierzchni HsDCR. Tak więc zaproponowano dwa modele (PPCRR1 i PPCRR2), które różnią się między sobą sposobem w jaki dsRNA jest umieszczone na powierzchni PAZ.



**Rys. 4.27.** Porównanie strukturalne różnych kompleksów PAZ-RNA. Błękitny TtAgo (3dlb), różowy TtAgo (3dlh), zielony HsAgo1 (1si3), żółty – TtAgo (3f73), czerwony HsPIWIL1 (307v), HsDCR (4nha) – fioletowy.

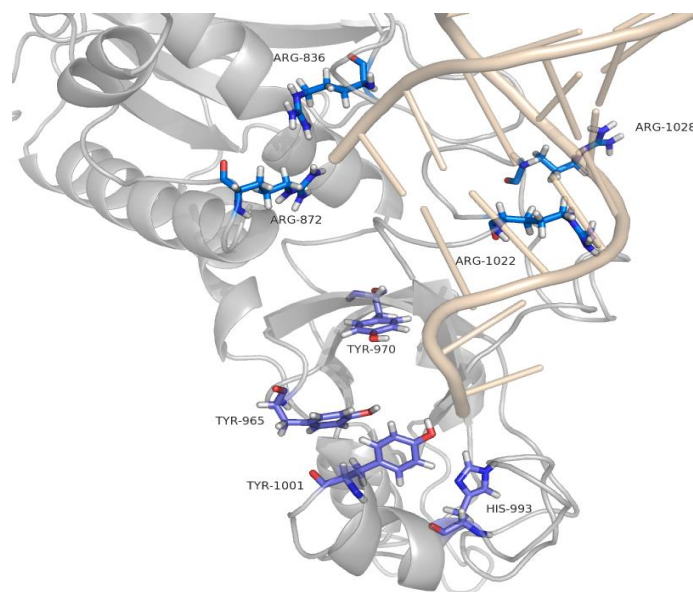
W modelach AtDCL4 miejsca katalityczne RNazy III są oddzielone od kieszeni PAZ wiążącej wystający koniec 3' RNA przez moduł platforma-helisa łącząca, podobnie jak w GiDCR. Badania strukturalne ludzkiego białka Dicer za pomocą mikroskopii elektronowej (EM) oraz nową metodą analizy lokalizacji domen, ujawniły, że domena platformy nie leży bezpośrednio pomiędzy PAZ i RNazami III, ale jest oddzielona od RNaz przez "domenę linijki" (Lau 2012). Ponadto, wykazano, że umieszczenie domen RNaz III i modułu platforma-PAZ w strukturze HsDCR na mapie EM jest odmienne niż w strukturze GiDCR (Lau 2012). Ponieważ AtDCL4, podobnie jak GiDCR, nie ma odpowiedniej domeny "pomiarowej", jest bardziej prawdopodobne, że pozycja domeny PAZ względem RNazy III jest jednak bardziej podobna do GiDCR niż do HsDCR. W każdym z modeli wzajemne położenia modułów platformy-helisy łączącej, PAZ i dimeru RNaz są nieco odmienne, jednakże nie można wykluczyć, że te elementy mają możliwość przemieszczania się względem siebie. W obu modelach pętla RNazy 1109-TEKCHER-1115 umożliwia tworzenie kontaktów z platformą. W modelu opartym na NMA (PPCRR1) powierzchnia interakcji pomiędzy platformą i RNazami jest bardziej rozbudowana niż w drugim modelu (PPCRR2) i obejmuje również reszty 1097-1100. W modelu PPCRR1 również domena PAZ jest bliżej połączona z modułem platforma-helisa łącząca niż w modelu PPCRR2. W modelu PPCRR1 dodatkowo reszty L960 i F969 z powierzchni domeny PAZ tworzą powierzchnię styku z platformą. Ponadto, położenie domeny PAZ stabilizuje pętlę PAZ, która

oddziałuje z platformą (reszty z pętli PAZ 1018-LCHVRNLLHNR-1028 są w kontakcie z resztami pochodzącymi z dwóch pętli platformy: 835-QR-836 oraz 802-HPADR-806).

Dwa proponowane modele kompleksów AtDCL4-dsRNA różnią się między sobą również strukturą dsRNA. Chociaż w obu części dsRNA, oddziałująca z domenami RNaz przyjmuje taką samą strukturę jak we wzorcu strukturalnym 2nuf – następujący po nim fragment w modelu zbudowanym z udziałem modyfikacji manualnych jest prosty, podczas gdy w drugim modelu – NMA – dsRNA jest lekko odchylony od osi RNA. Wartość RMSD wyznaczona dla atomów fosforu cząsteczek RNA w obu modelach wynosi 3.1 Å.

Mapowanie potencjału elektrostatycznego na powierzchni otrzymanych modeli PPCRR1 i PPCRR2 wykazuje wysoko dodatnio naładowane obszary w regionie kontaktu z dsRNA na obu domenach: platformy i PAZ (Rys. 4.25.). Umożliwia to prawidłowe oddziaływanie z ujemnie naładowanymi resztami fosforanowymi łańcucha RNA. Powierzchnia PAZ wygenerowanych modeli posiada mniejsze obszary naładowane dodatnio do wiązania wystającego 3' końca dsRNA, aniżeli domeny PAZ pochodzące z HsAgo1 i GiDCR (Ma 2004). W niedawnych badaniach wykazano, że pomimo iż domena PAZ jest wysoce konserwatywna u wszystkich białek typu Dicer, w białkach DCL4 może wykazywać znaczną odmienność w rozkładzie potencjału elektrostatycznego na powierzchni kieszeni wiązania RNA (Mukherjee 2013). Szczególnie u roślin jednoliściennych można zaobserwować całkowite odwrócenie potencjału elektrostatycznego kieszeni wiążącej RNA domeny PAZ białek DCL4 gdzie przyjmuje on głównie ujemne wartości. Ta radykalna zmiana potencjału elektrostatycznego na powierzchni domeny PAZ jest całkowicie nowa i nie jest charakterystyczna dla domen PAZ w ogólności (Lingel 2004, Wang 2009). Chociaż funkcje tych zmian we właściwościach elektrostatycznych kieszeni wiążącej dsRNA nie są znane, Mukherjee zaproponował, że DCL4 mogą wykazywać różne właściwości wiązania dsRNA w porównaniu z innymi białkami typu Dicer. W procesie modelowania wygenerowaliśmy kilka modeli domeny PAZ na podstawie nieco innego dopasowania do szablonów. Modele różniły się rozkładem potencjału elektrostatycznego na powierzchni kieszeni wiązania RNA, jednakże wyniki oceny jakości modeli jako najlepsze wskazywały właśnie te, których kieszeń wiązania 3' końca RNA była naładowana dodatnio.

Ogólne właściwości domeny PAZ z AtDCL4, niezbędne do rozpoznawania wystającego końca 3' dsRNA, są podobne jak w kompleksie PAZ-siRNA pochodzącym z HsAGO1 (PDB ID: 1si3) chociaż nie wszystkie reszty kieszeni wiążącej 3' koniec dsRNA są konserwowane (por Rys. 4.18.). Alfa-helisy na dnie kieszeni wiążącej HsAGO1-PAZ zawierają kilka zakonserwowanych reszt aromatycznych, które wydają się być istotne dla rozpoznawania dsRNA. W AtDCL4-PAZ tylko reszta Y1001 jest zachowawcza i odpowiada Y314 w HsAGO1-PAZ. W naszych modelach ta reszta może uczestniczyć w tworzeniu wiązań wodorowych z końcem 3' dsRNA. Reszty Y309 i F310 (numeracja wg HsAGO1), które zostały określone jako wymagane dla optymalnego wiązania RNA przez domenę PAZ z HsAGO1-PAZ i HsDCR (Gurtan 2012, Tian 2014) nie są zachowane w DCL4 PAZ. Reszta Y309 z HsAGO1-PAZ odpowiada S996 w AtDCL4 natomiast F310 – I997. W strukturze HsAGO1-PAZ siRNA F292 uczestniczy w oddziaływaniach warstwowych z 3' końcową zasadą. Chociaż ta reszta jest wysoce konserwowana wśród domen PAZ, w AtDCL4-PAZ w analogicznej pozycji występuje V985, a tym samym nie jest zachowana. Jednakże inne kluczowe interakcje między 2-nt wystającym 3' końcem RNA i domeną PAZ obserwowane w HsAGO1-Paz-siRNA zostały utrzymane w modelu AtDCL4-PAZ-dsRNA. Reszta fosforanowa ze szkieletu RNA z końca 3' tworzy wiązania wodorowe z resztą Y965 AtDCL4. Odpowiada to wiązaniu H269 z HsAGO1-PAZ z RNA. Natomiast zachowawcza reszta Y970 z AtDCL4, która odpowiada Y277 w HsAGO1-PAZ, może tworzyć wiązanie wodorowe z przedostatnią rybozą na 3' końcu RNA.



**Rys. 4.28.** Przestrzenne położenie reszt z domeny PAZ i platformy odpowiadających za wiązanie dsRNA. Ułożenie pokazuje potencjalne oddziaływanie z końcami 5' i 3' cząsteczki RNA.

W obu otrzymanych modelach 5' koniec nici dsRNA wchodzi do dodatnio naładowanej kieszeni wewnątrz platformy. W szczególności reszta R872, która znajduje się na C-końcu helisy łączącej, jest odpowiedzialna za tworzenie wiązań wodorowych z 5' końcem dsRNA i przyczynia się do utworzenia właściwego potencjału elektrostatycznego na powierzchni białka. Reszty, które wchodzi w interakcje z dsRNA znajdują się na platformie i w pętlach domeny PAZ, chociaż specyficzne interakcje są różne w obu modelach to ogólne dopasowanie dsRNA do powierzchni platformy jest zachowane. W obu modelach pętla platformy 830-LHLAHQRSVS-839 tworzy interakcje z RNA. W modelu PPCRR1 wspomniana pętla (a zwłaszcza jej fragment QR) stanowi płaszczyznę kontaktu ze szkieletem RNA w obrębie końca 5', podczas gdy w modelu PPCRR2 pętla ta wchodzi do mniejszej bruzdy ponad końcem 5' dsRNA i jest zaangażowana w bezpośrednie interakcje z dsRNA poprzez atomy szkieletu obu nici. Pętla ta nie jest ściśle zachowana w obrębie roślinnych DCLi. Sama reszta R836 jest zachowana w białkach DCL4, a w typie DCL3 jest zastąpiona przez elektrododatnią lizynę. Dodatkowo, N-końcowa część tej pętli, podobnie jak analogiczna pętla GiDCR, zaangażowana jest w kontakty z domeną RIIIA. W obu modelach może ona odgrywać rolę w umiejscowieniu przestrzennym modułu platformy-PAZ-helisy łączącej względem modułu heterodimeru RNazy III.

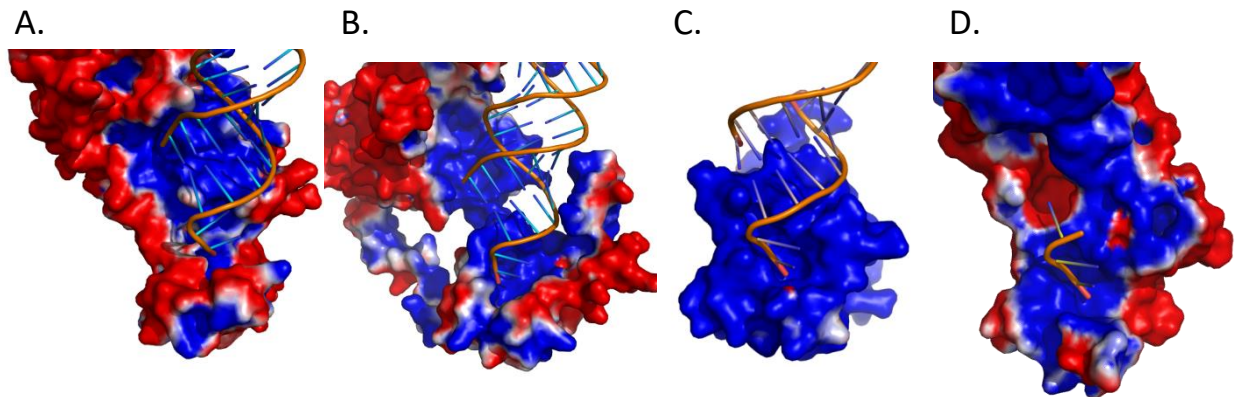
Rozległe interakcje z dsRNA również wykazuje N-końcowa część dużej pętli PAZ 1019-CHVRNLLHNR-1028. Interakcje te różnią się między modelami, ponieważ pętla ta nie ma identycznego położenia. Zatem dsRNA został zadokowany na powierzchni PAZ w nieco inny sposób. W modelu PPCRR1 pętla ta tworzy oddziaływania z dużą bruzdą RNA i stanowi fragment kieszeni wiążącej z 5' koniec, natomiast w modelu PPCRR2 pętla PAZ raczej wykazuje interakcje z małą bruzdą i atomami szkieletu RNA.

Długa pętla PAZ jest cechą charakterystyczną białek Dicer i DCL, podczas gdy w domenach PAZ białek AGO i PIWI pętla ta jest znacznie krótsza (Simon 2011). Predykcje z wykorzystaniem metaservera Genesilico pozwoliły określić C-końcową część tej pętli (fragment 1031-TNLEPQELD-1039) jako nieuporządkowaną. Motyw 1023-NNL-1025 na N-końcu dużej pętli PAZ jest dobrze zachowany zarówno wśród białek DCL jak i Dicer (Mukherjee 2013). Pozostała sekwencja tej pętli PAZ jest zachowana jedynie wśród roślinnych DCL. Istnieje jednak dla niej swego rodzaju odpowiednik w szablonach struktur HsDCR lub

GiDCR (Rys. 4.18.). Struktura krystaliczna fragmentu białka HsDCR obejmującego platformę-PAZ-helisę łączącą, pokazała że wspomniana pętla może tworzyć helisę zbudowaną z 14 aminokwasów, lub pozostaje nieuporządkowana (Tian 2014). Te dwie formy strukturalne mogą sugerować dwa różne etapy oddziaływania białka z dsRNA. Rozpleciona helisa umożliwia kontakt dsRNA z powierzchnią białka, co jest związane z etapem cięcia. Uformowanie helisy może służyć odsunięciu przeciętego dsRNA od powierzchni białka pod kątem ok.60°, co wskazuje na etap uwolnienia produktu (Tian 2014). Różne kształty pętli PAZ w naszym modelu odzwierciedlają jej możliwy dynamiczny charakter. Jest prawdopodobne, że pętla jest częściowo nieuporządkowana w białku niezwiązanym z RNA, przyjmuje natomiast inną formę po przeprowadzonej przez DCL reakcji cięcia.

Wcześniejsze prace nad HsDCR i DmDCR1 wykazały specjalistyczną kieszeń, która wiąże resztę fosforanową 5' końca prekursora dsRNA (Park 2011, Tian 2014). Struktura krystaliczna fragmentu białka HsDCR obejmującego platformę-PAZ-helisę łączącą pokazuje pięć reszt argininy (R788, R790, R821, R996, R1003, numeracja według HsDCR), które mogą być wymagane dla rozpoznania 5' końca RNA (Tian 2014). Reszty R788, R790 należą do pierwszej pętli platformy (pętla  $\beta 1$ - $\beta 2$ ) i nie są zachowawcze w białkach DCL4. W otrzymanych modelach AtDCL4, odpowiadająca jej pętla jest krótsza i nie oddziałuje bezpośrednio z 5' końcem RNA jednakże jest zaangażowana w inne oddziaływania z RNA. W modelu NMA (PPCRR1), reszta R806 z pętli jest zaangażowana w oddziaływania ze szkieletem nici RNA zawierającej niesparowany 3' koniec. Reszta R821 w HsDCR znajduje się w innej pętli platformy – ta reszta nie jest również zachowana w DCL4 jednak reszty w odpowiadającej jej lokalizacji w pętli AtDCL4 także tworzą interakcje z dsRNA i przyczyniają się do powstania *obszarów dodatnio naładowanych* na powierzchni platformy. Istnieją ponadto dwie argininy w pętli PAZ AtDCL4, które są usytuowane podobnie jak reszty R996 i R1003 w HsDCR ale nie są one określone jako zachowawcze (R1022 z AtDCL4 odpowiada R996 w HsDCR, natomiast R1028 - R1003) (Mukherjee 2013). Wykazano, że te dwie reszty rozpoznają 5' koniec dsRNA (w obu modelach reszta R1028 nie jest zaangażowana w kontakt z 5' końcem dsRNA). Poprzednie badania sugerują, że K1016 z AtDCL4 może być częścią kieszeni wiążącej reszty fosforanowe 3' końca RNA (Fukunaga 2014), ale w uzyskanych modelach raczej R1014 wskazuje na interakcję z dsRNA.

Oba modele AtDCL4 wykazują dodatnie naładowanie kieszeni wiążących zarówno 3' jak i 5' koniec RNA (Rys. 4.29.). Jednakże utworzone są one z reszt aminokwasowych, które nie są ściśle zakonserwowane u ssaków i w innych białkach zawierających domenę PAZ.



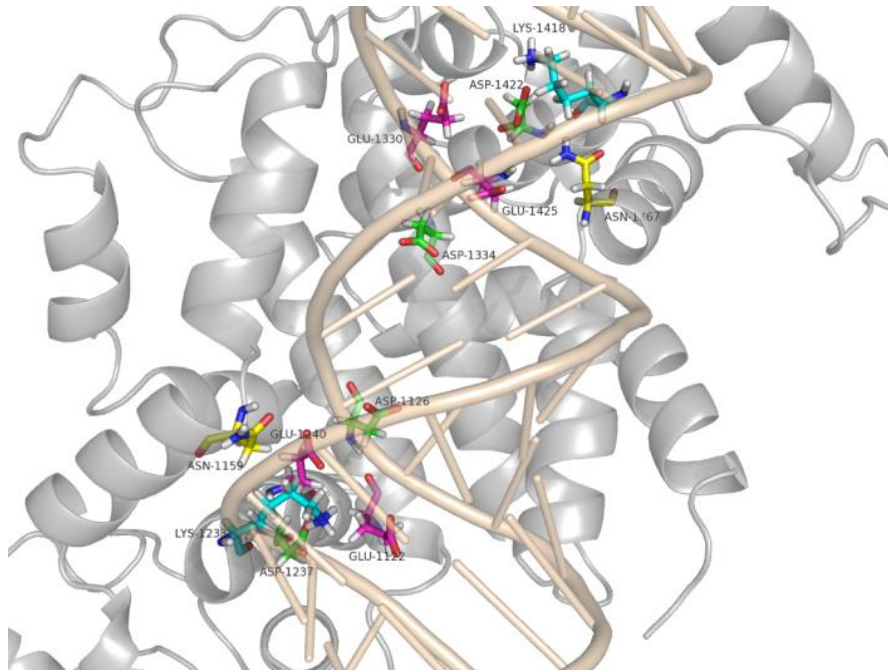
**Rys. 4.29.** Potencjał elektrostatyczny na powierzchni domeny PAZ i fragmentu platformy **A.** Model PPCRR1 **B.** Model PPCRR2, **C.** HsAGO1 w kompleksie z RNA (1si3); **D.** Giardia Dicer (2qvw), Nałożono domenę PAZ z HsAGO1 i przeniesiono fragment RNA, dla jasności rysunku jedynie 3nt. **A, B, C, D.** Kolor niebieski wskazuje potencjał pozytywny większy niż 1 kT/e, czerwony potencjał negatywny mniejszy niż -1 kT/e, biały – obojętny.

Moduł obejmujący RNazę A i B w obu wygenerowanych modelach jest identyczny, podobnie jego oddziaływanie z RNA nie różni się pomiędzy modelami, zatem omówiony zostanie jako model AtDLC4 bez rozróżniania jaką metodą został wygenerowany.

W przestrzeni katalitycznej dimeru RNaz AtDLC4 usytuowane są wysoce zakonserwowane reszty odpowiedzialne za hydrolizę wiązań fosfodiesterowych: E1122, D1126, D1237, E1240, w RIIIA oraz E1330, D1334, D1422, E1425 w RIIIB. Prekursor siRNA – długa cząsteczka dsRNA – został umieszczony analogicznie jak w przypadku wzorca strukturalnego 2nuf. Reszty katalityczne nie znajdują się w bezpośrednim kontakcie z dsRNA albowiem wykazano, że dwa jony  $Mg^{2+}$  są niezbędne dla tworzenia aktywnego katalitycznie kompleksu białko-RNA (Gan 2008). Dwie symetryczne pary reszt: N1159 i N1367 oraz K1233 i K1418 (odpowiednio w RIIIA i RIIIB) w obu modelach także biorą udział w interakcji z dsRNA. N1367 i K1418 odpowiadają N1726 i K1790 w mysim białku MmDCR1 (Du 2008). Reszty te są wysoce zachowawcze w domenach RIIIA i RIIIB w eukariotycznych białkach typu Dicer oraz innych endorybonukleazach. W niedawno rozwiązanej strukturze RNazy III z drożdży N278 (odpowiadający w AtDCL4 N1159 i N1367) współdziała z resztą fosforanową na 5' końcu i cząsteczkami wody podczas gdy K313 (odpowiadający w AtDCL4 K1233 i K1418) współdziała



zarówno z resztą fosforanową na 5' końcu jak i resztą D317 (odpowiadający w AtDCL4 D1237 i D1422) (Liang 2014). Obydwa modele AtDCL4 są zgodne z tymi danymi i sugerują, że te reszty także u roślinnych DCLi mogą odgrywać istotną rolę w reakcji rozszczepiania wiązania fosfodiesterowego, prowadzącego do utworzenia produktu – siRNA.



**Rys. 4.30.** Reszty katalityczne w domenach RNazy odpowiadające za cięcie dsRNA. Główne kolory wskazują poszczególne aminokwasy. Błękitny – lizyna, różowy – kwas glutaminowy, zielony – asparagina, żółty – kwas asparaginowy.

W przypadku modelowanego kompleksu AtDCL4-RNA, oba te kluczowe motywy wiążące RNA, znajdujące się w obszarach RNazy III A i RNazy III B, są zachowane w trakcie symulacji MD, co świadczy o ich stabilności. Ponadto, zidentyfikowano dodatkowe oddziaływania białko-RNA, które mogą pomóc w pozycjonowaniu dsRNA do miejsc cięcia (Gan 2006). W AtDCL4 reszty T1150 i T1358 znajdujące się w  $\alpha$ -helisach domen RIIIA i RIIIB (odpowiednio), są w stanie oddziaływać ze szkieletem dsRNA. Dodatkowo, reszta R1151 (w RIIIA) może wchodzić w interakcje ze szkieletem dsRNA. Odpowiada ona reszcie S68 z AaRNazy III, która oddziałuje z częścią cukrową reszty nukleotydowej w miejscu cięcia RNA (Gan 2006). Jednakże arginina R1151 jest zachowana jedynie w domenie RIIIA DCL4 i DCL1, ale w RIIIB nie występuje w tym miejscu. W domenie RIIIB występuje R1415, która może tworzyć kontakty z dsRNA. W tej pozycji reszta K lub R jest obecna w DCL1, oraz DCL3 i DCL4 ale nie DCL2.

W strukturze kompleksu bakteryjnej RNazy III z dsRNA została zidentyfikowana poprzedzająca helisę pętla zawierająca reszty katalityczne D1237, E1240 (w RIIIA) czy D1422, E1425 (w RIIIB) – numeracja wg AtDCL4 – jako dodatkowy region odpowiadający za rozpoznanie miejsca cięcia RNA (Gan 2006). W domenie RIIIA AtDCL4 odpowiednia pętla jest dłuższa i w otrzymanych modelach jej struktura nie jest pewna ze względu na brak szablonu matrycy strukturalnej. W AtDCL4 RIIIB odpowiadająca sekwencja jest krótsza niż w RIIIA o około 22 reszt aminokwasowych, a prócz krótkiej pętli zawiera helisę  $\alpha$ , która jest charakterystyczna dla eukariotycznych białek typu Dicer, ale nie występuje u bakteryjnych RNaz III (Du 2008). N-koniec tej helisy, zawierający reszty 1387-SIY-1389, może być potencjalnie zaangażowany w interakcje z dsRNA. Reszta S1387 jest dobrze zachowana zarówno wśród roślinnych DCLi jak i u MmDCR1. Nałożenie strukturalne AtDCL4-RIIIB z bakteryjną RNazą III wykazało, że S1387 odpowiada R97, które w bakteryjnej RNazie III tworzy wiązanie wodorowe z RNA.

Inny region w AtDCL4, który może tworzyć interakcje z dsRNA został przewidziany w pętli domeny RIIIA 1109-TEKCHER-1115. W obu modelach reszta K1111 może oddziaływać ze szkieletem dsRNA w małej bruździe. Dodatkowo naładowana reszta (K lub R) jest zachowana w tym położeniu w DCL2-4 ale nie DCL1. W GiDCR ta pętla jest dłuższa o 16-18 reszt niż u białek DCL i jest określana jako "pętla platformy". Wykazano, że mutacje w GiDCR R310A, R312A w tej pętli skutkują wpływem na właściwe przetwarzanie dsRNA (MacRae 2007).

W domenie RIIIB reszty 1315-HPSYN-1319 z pętli poprzedzającej helisę zawierającą motyw katalityczny E1330-D1334, podobnie jak w odpowiadającym jej regionie pętli w RIIIA, mogą tworzyć interakcje z nicią RNA zawierającą 3' koniec po reakcji cięcia. W szczególności reszta histydyny – H1315 – wydaje się być zaangażowana w wiązanie wodorowe ze szkieletem RNA. Interesujące jest, że histydyna ta jest zachowana zarówno u RNazy III z bakterii, jak również w domenie RIIIB roślinnych DCLi oraz białek typu Dicer pochodzących od zwierząt.

#### 4.2.6. Dyskusja

W pracy zaproponowano dwa modele białka DCL4 z *Arabidopsis thaliana* w regionie jego rdzenia katalitycznego w kompleksie z jego substratem - dsRNA. Modele AtDCL4

otrzymano wykorzystując metodę modelowania homologicznego, opierając się o szablony strukturalne poszczególnych domen oraz bloków domen, a następnie wykorzystując strukturę największego z dostępnych wzorców do modelowania wzajemnego ułożenia domen. Należy nadmienić, że roślinne białka DCL są na tyle odmienne od analogicznych białek zwierzęcych, których struktury częściowo poznano, oraz od białek mających pewne podobne funkcjonalne domeny, że proces ich modelowania nie był problemem trywialnym. Wszelkie próby uzyskania modelu metodami automatycznymi i wspomaganymi serwerami dedykowanymi otrzymywaniu struktur kończyły się niepowodzeniem. Dlatego też, wiele dopasowań zarówno sekwencyjnych, jak i strukturalnych było poprawianych ręcznie, tak aby ostateczny model był jak najbardziej zgodny z danymi biologicznymi opisującymi jego działanie.

Funkcja białka AtDCL4 jako "linijki" do odmierzenia 21-nt RNA i wycięcia odpowiedniego fragmentu z dłuższego prekursora wymaga nieco innego położenia modułu zawierającego dwie domeny RNazy III w stosunku do domeny PAZ niż w strukturze krystalicznej GiDCR. Do budowy kompleksu AtDCL4-dsRNA wykorzystano informacje strukturalne o interakcjach dwóch domen z dsRNA (domeny PAZ i RIIIA) oraz przewidywanej elastyczności AtDCL4. W szczególności, wymodelowana została domena platformy, która wykazuje niskie podobieństwo sekwencji z homologicznymi domenami o znanej strukturze. Przewidziana struktura drugorzędowa wskazuje na duże prawdopodobieństwo, że zachowana jest jej struktura trzeciorzędowa. Domena platformy w obu modelach AtDCL4, podobnie jak w GiDCR, łączy domenę PAZ oraz moduł domen RNaz III i odgrywa istotną rolę w umiejscowieniu przestrzennym tych domen względem siebie, w ten sposób przyczyniając się do "ustalenia wymiaru" ciętego dsRNA. Prezentowane modele wskazują reszty, które są ważne dla interakcji AtDCL4 z dsRNA. Modele ujawniły obecność dwóch strategicznych pętli należących do domen PAZ oraz platformy zawierających kluczowe reszty biorące udział w interakcjach pomiędzy tymi domenami jak i pomiędzy białkiem i dsRNA. Pętle te mogą ułatwiać wiązanie dsRNA do białka DCL lub też pomagać w oddysocjowaniu produktu po reakcji cięcia. Długa pętla domeny PAZ może odgrywać rolę w pozycjonowaniu domeny PAZ (w stosunku do platformy), co może pomóc w poprawnym kierowaniu nici 3' RNA do kieszeni wiążącej w domenie PAZ. W otrzymanych modelach ufosforylowany 5' koniec dsRNA znajduje się w elektrododatniej kieszeni. Reszty aminokwasowe, które tworzą kieszeń wiążącą nić 5' w naszym modelu AtDCL4 są tylko częściowo zachowane z tymi, które pełnią analogiczną

funkcję w HsDCR. Na podstawie analiz strukturalnych modeli możemy przewidzieć, że reszty R836 i R872 na powierzchni platformy są istotne dla rozpoznania białko-RNA. W szczególności, mutacje tych reszt mogą pomóc ustalić, jak ważna jest w przypadku DCL4 (lub roślinnych DCLi w ogólności) kieszeń wiążąca 5' koniec RNA. Innym interesującym regionem dalszych biochemicznych badań jest pętla PAZ (1017 – 1041), a zwłaszcza czy (i w jaki sposób?) mutacja reszt R1022 i/lub R1028 może prowadzić do zmian zdolności wiązania RNA przez AtDCL4 i jak przekłada się to na inne roślinne białka typu Dicer.

Całe białko AtDCL4 zawiera dodatkowe domeny, które także najprawdopodobniej przyczyniają się do właściwego umieszczenia dsRNA w miejscu katalitycznym (Cenik 2011, Liu 2012b). Są to: dwie domeny wiążące dsRNA (dsRBD), domeny helikazy i prawdopodobnie DUF283 (Qin 2010). Co więcej, funkcje AtDCL4 w kompleksie mogą być wspierane przez inne białka np. DRB4 (Fukudome 2011). Każdy z dwóch otrzymanych modeli może stanowić swoisty szkielet do stworzenia modelu całego białka DCL4 czy też funkcjonalnego kompleksu złożonego z białek z nim oddziałujących.

Prezentowane modele mogą być wykorzystane do planowania kolejnych eksperymentów zarówno biochemicznych jak i bioinformatycznych. Sugerowana struktura rdzenia katalitycznego jak i inne miejsca potencjalne odpowiadające za rozmaite interakcje mogą wspomóc proces projektowania eksperymentów np. w obszarze wprowadzanych do sekwencji białka mutacji punktowych, wyjaśnić obserwowane skutki mutacji występujących spontanicznie lub pomóc w interpretacji wyników badań biochemicznych.

## 5. Podsumowanie

Wpływ struktury cząsteczek biologicznych na ich funkcję jest niezaprzeczalny. Przedstawione w niniejszej rozprawie wyniki badań sugerują istnienie szeregu specyficznych cech strukturalnych w budowie zarówno białek jak i RNA biorących udział w biogenezie małych regulatorowych RNA u roślin na przykładzie *Arabidopsis thaliana*. Otrzymane modele białka AtDCL4 pokazały w jaki sposób jego struktura determinuje długości produktów prowadzonej przez nie reakcji hydrolizy dsRNA. Wymodelowano oddziałującą z dsRNA platformę stabilizującą, której obecność pomaga w utrzymaniu odległości wymaganej dla uzyskania określonej długości produktów. Prezentowane modele wskazują reszty, które są istotne dla oddziaływania AtDCL4 z dsRNA. Szczegółowo przeanalizowano domenę PAZ i zaobserwowano obecność kieszeni wiążącej 3' koniec RNA oraz potencjalną obecność drugiej kieszeni wiążącej 5' koniec dsRNA.

Badania nad strukturą drugorzędową natywnych prekursorów miRNA zaowocowały wyznaczeniem szeregu cech termodynamicznych, charakteryzujących ich budowę, niezbędnych do prawidłowego przetwarzania ich w ostateczny produkt – małe regulatorowe RNA. Opracowano na ich podstawie algorytm zaimplementowany w programie AmiRNA Designer przeznaczony do analizowania profili termodynamicznych drugorzędowych struktur RNA oraz projektowania na ich podstawie cząsteczek o podobnych właściwościach.

Omówione w niniejszej pracy wyniki zostały przedstawione w następujących publikacjach:

- 1) **Mickiewicz A**, Rybarczyk A, Sarzynska, J, Figlerowicz M, Błazewicz J. (2016) AmiRNA Designer – new method of artificial miRNA design. Acta Biochim. Pol. 63, 71–77. doi:10.18388/abp.2015\_989
- 2) **Mickiewicz A**, Sarzyńska J, Miłostan M, Kurzyńska-Kokorniak A, Rybarczyk A, Łukasiak P, Kuliński T, Figlerowicz M, Błazewicz J. (2017) Modeling of the catalytic core of Arabidopsis thaliana Dicer-like 4 protein and its complex with double-stranded RNA. Comput. Biol. Chem. doi: <http://dx.doi.org/10.1016/j.compbiolchem.2016.11.003>

- 3) Kurzynska-Kokorniak A, Koralewska N, Pokornowska M, Urbanowicz A, Tworak A, **Mickiewicz A**, Figlerowicz M. (2015) The many faces of Dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities. *Nucleic Acids Res.* 43, 4365–80. doi:10.1093/nar/gkv328 .
- 4) Sarzyńska J, **Mickiewicz A**, Miłostan M, Łukasiak P, Błażewicz J, Figlerowicz M, Kuliński T. (2010) Flexibility of Dicer Studied by Implicit Solvent Molecular Dynamics Simulations. *Comput. Methods Sci. Technol.* 16, 97–104. doi:10.12921/cmst.2010.16.01.97-104
- 5) Miskiewicz J, Tomczyk K, **Mickiewicz A**, Sarzynska J, Szachniuk M. (2017) Bioinformatics study of structural patterns in plant microRNA precursors. *Biomed Res Int.*; doi: 10.1155/2017/6783010.

Kontynuacją prac przedstawionych w niniejszej rozprawie zajmują się kolejne pokolenia doktorantów w poznańskim *Europejskim Centrum Bioinformatyki i Genomiki (ECBiG)*. Planowane jest wykonanie eksperymentów biochemicznych z wykorzystaniem sekwencji wyciszających otrzymanych przy użyciu programu AmiRNA Designer. Wykorzystując doświadczenie zdobyte w modelowaniu białka AtDCL4, zostanie wymodelowane białko AtDCL1. Ponadto wykonano analizę prekursorów miRNA dla struktur 3D pre-miRNA (Miśkiewicz 2018). W tym celu wykorzystując takie narzędzia obliczeniowe jak RNAComposer (Popenda 2012, Antczak 2016) i MCQ4Structres (Zok 2014). Wytworzone modele 3D RNA będą oceniane na podstawie ich dopasowania do modelu białka AtDCL1.

## 6. Streszczenie

Regulacja ekspresji genów jest dla roślin procesem dzięki któremu mogą bronić się przed egzogennymi patogenami, sterować procesami rozwojowymi, jak i reagować na zmieniające się warunki środowiska. Małe regulatorowe RNA, będące elementami szlaku potranskrypcyjnej regulacji ekspresji genów, to precyzyjny i szybki sposób na obniżenie poziomu ekspresji genów, co sprawia, że stanowią potencjalne narzędzie do badań biologii molekularnej a także celów aplikacyjnych na roślinach uprawnych.

Biogeneza małych regulatorowych RNA rozpoczyna się od długich RNA posiadających w swojej strukturze fragmenty dwuniciowe. Jedno z białek DCL wycina z nich duplekсы o długości 20-24 nt, które zostają następnie rozplecione w kompleksie Mikroprocesora, gdzie nici wiodąca pozostaje od niego przyłączona jako wzorzec do rozpoznania mRNA. W kolejnym etapie kompleks ten przyłącza się do mRNA i przecina je w miejscu hybrydyzacji srRNA. Każde z roślinnych białek DCL wycina fragmenty o charakterystycznej dla siebie, stałej długości, jednakże dotąd nie określono eksperymentalnie struktury żadnego z nich. Struktura duplexu miRNA:miRNA\* oraz oddziaływanie pomiędzy nicią wiodącą miRNA, a miejscem docelowym w mRNA wydaje się mieć istotny wpływ na efektywność potranskrypcyjnej regulacji ekspresji genów, jednakże nie są znane dokładne parametry strukturalne srRNA warunkujące skuteczność ich działania. W ramach niniejszej rozprawy zrealizowano następujące cele badawcze w kierunku poznania strukturalnych uwarunkowań biogenezy roślinnych srRNA:

1. Wyznaczono szereg cech, zarówno na poziomie sekwencji jak i struktury drugorzędowej dsRNA. Badano dwa zestawy danych: jeden obejmował duplekсы miRNA:target, zaś drugi prekursorы miRNA, z których, już po uzyskaniu struktury drugorzędowej, pobierane do dalszej analizy były duplekсы miRNA:miRNA\* przedłużone o 10 nt przed i za sekwencją miRNA. W ten sposób uzyskano referencyjne profile termodynamiczne, wykorzystane później jako parametr do weryfikacji potencjalnych amiRNA.

2. Opracowano nowy algorytm wykorzystujący cechy termodynamiczne natywnych struktur, które tworzą miRNA w swoich prekursorach, jak i oddziałujące z docelowym transkryptem wyciskanego genu. Stworzono narzędzie bioinformatyczne AmiRNA Designer umożliwiające projektowanie sekwencji wyciszających zadany gen. Wykorzystano

opracowany program do generowania sekwencji sztucznych miRNA (amiRNA) dla przykładowych genów. Po zaprojektowaniu każdej sekwencji program sprawdza jej zgodność z parametrami profili termodynamicznych. Opracowano sposób wprowadzania modyfikacji do sekwencji srRNA, tak by projektowane sekwencje lepiej pasowały do profili. Przebiega to wg. następującego planu:

1. Uzyskanie sekwencji komplementarnej do zadanego targetu=>amiRNA
2. Porównanie jej z profilem termodynamicznym dupleksu target:amiRNA
3. Modyfikacje sekwencji amiRNA, ponowne porównanie j.w.
4. Uzyskanie ostatecznej sekwencji amiRNA
5. Uzyskanie sekwencji komplementarnej do amiRNA=>amiRNA\*
6. Porównanie jej z profilem termodynamicznym dupleksu amiRNA:amiRNA\*
7. Modyfikacje sekwencji amiRNA\* ponowne porównanie j.w.
8. Uzyskanie ostatecznej sekwencji amiRNA\*

3. Wymodelowano strukturę przestrzenną białka AtDCL4 w kompleksie z jego substratem dsRNA, określając jej związek pomiędzy strukturą enzymu a długością wycinanych przez niego fragmentów dsRNA. Stosując metody *in silico* zidentyfikowano białkowe czynniki strukturalne – miejsca katalityczne wiążące RNA oraz istotne zwoje, domeny i poszczególne reszty aminokwasowe – oddziałujące z dsRNA i mające bezpośredni wpływ na powstawanie srRNA. W szczególności wykorzystano modelowanie porównawcze (w oparciu o homologię) oraz modelowanie kompleksu białko-RNA, metodę analizy drgań normalnych i symulacje dynamiki molekularnej kompleksów białek z RNA. Modelowanie struktury białka w oparciu o skromne dane – niewiele poznanych struktur oraz ich niskie podobieństwo sekwencyjne do modelowanego białka – okazało się zadaniem nietrywialnym. Zarówno fragmenty sekwencji nieposiadające żadnego odpowiednika mogącego stanowić szablon strukturalny, jak i te z nikłym podobieństwem do znanych struktur wymagały większej uwagi i dodatkowych zabiegów. Otrzymane modele zostały poddane szeregowi analiz komputerowych określających ich wiarygodność jako wysoką, a także analizie porównawczej łączącej poziomy sekwencji, struktury oraz danych pochodzących z eksperymentów biochemicznych.

W efekcie realizacji celów niniejszej pracy doktorskiej powstało pięć publikacji, przeglądowych, jak i opisujących wyniki badań własnych a także program komputerowy



AmiRNA Designer, który ma na celu projektowanie sekwencji wyciszających, skierowanych specyficznie do określonego genu, w szczególności roślinnego.

## 7. Abstract

For plants, the regulation of gene expression is a process through which they can defend themselves against exogenous pathogens, control development processes and respond to changing environmental conditions. Small regulatory RNAs, which are elements of the pathway of post-transcriptional regulation of gene expression, are a precise and fast way to reduce the level of gene expression, which makes them a potential tool for research of molecular biology and application purposes on cultivated plants.

The biogenesis of small regulatory RNAs starts with long RNAs with double-stranded fragments in their structure. One of the DCL proteins cuts 20-24 nt duplexes from them, which are then unraveled in the microprocessor complex where the leading strand remains attached therefrom as a template for mRNA recognition. In the next stage, this complex attaches the mRNA and cuts it at the site of srRNA hybridization. Each of the DCL plant proteins cuts fragments of its own characteristic, fixed length, however, the structure of none of them has yet been experimentally determined. The structure of the miRNA:miRNA\* duplex and the interaction between the miRNA guide strand and the target sequence in the mRNA seems to have a significant impact on the effectiveness of posttranscriptional gene expression regulation, however, the exact structural parameters of srRNAs determining their effectiveness are unknown. As part of this dissertation, the following research objectives have been achieved in order to learn about the structural conditions of srRNA plant biogenesis:

1. A number of features have been identified, both at the sequence and secondary structure level of the dsRNA. Two sets of data were investigated: one included miRNA: target duplexes, and the other, miRNA precursors, of which, after obtaining the secondary structure, miRNA duplexes: miRNA \* duplexes extended by 10 nt before and after the miRNA sequence were collected for further analysis. In this way, the reference thermodynamic profiles were obtained, which were later used as a parameter to verify potential amiRNAs.

2. A new algorithm was developed that uses the thermodynamic features of the native structures that make up the miRNAs in their precursors as well as interacting with the target transcript of the silenced gene. AmiRNA Designer – a bioinformatics tool to design sequences for silencing a specific gene was designed. The developed program was used to generate

artificial miRNA sequences - amiRNAs - for exemplary genes. After designing each sequence, the program checks its compliance with the parameters of the thermodynamic profiles. A method of introducing modifications to the srRNA sequence was developed so that the designed sequences better fit the profiles. It runs according to the following plan

1. Obtaining complementary sequence to given target=>amiRNA
2. Compare it with the thermodynamic profile of the target:amiRNA duplex
3. amiRNA sequence modifications, re-comparison as above
4. Obtaining the final sequence of amiRNA
5. Obtaining complementary sequence to amiRNA= >amiRNA\*
6. Compare it with the thermodynamic profile of the amiRNA:amiRNA\* duplex
7. amiRNA\* sequence modifications, re-comparison as above
8. Obtaining the final sequence of amiRNA\*

3. The spatial structure of the AtDCL4 protein in complex with its dsRNA substrate was modeled, determining its relationship between the structure of the enzyme and the length of generated dsRNA fragments. Using *in silico* methods, a few protein structural features were identified – RNA-binding catalytic sites and folds, domains and individual amino acid residues - responsible for dsRNA binding and directly impacted the formation of srRNA. In particular, comparative modeling (based on homology) and modeling of the protein-RNA complex, the method of normal mode analysis and molecular dynamics simulations of protein-RNA complexes were used. Modeling the protein structure based on scarce data - little known structures and their low sequence similarity to the modeled protein - turned out to be a non-trivial task. Both fragments of the sequence without any equivalent that could constitute a structural template and those with little resemblance to known structures required more attention and additional efforts. The obtained models were submitted to a series of computer analyzes verifying their reliability as high, as well as a comparative analysis combining the levels of sequence, structure and data from biochemical experiments.

As a result of the implementation of the objectives of this dissertation, five publications were created, both reviewing and describing the results of own research, as well as the AmiRNA Designer computer program, which aims to design silencing sequences specifically targeted at a specific gene, in particular for plants.

## 8. Bibliografia

- Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouche N, Gascioli V, Vaucheret H. (2006) DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr. Biol*, **16**: 927-932
- Agrawal N, Dasaradhi PV, Mohammed A, Malhotra P, Bhatnagar RK, Mukherjee SK. (2003) RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev*. **67**, 657-685.
- Ahmed W, Xia Y, Li R, Bai G, Siddique KHM, Guo P. (2019) Non-coding RNAs: Functional roles in the regulation of stress response in Brassica crops, *Genomics*, **18**: 30626-30628
- Ai T, Zhang L, Gao Z, Zhu CX, Guo X. (2011) Highly efficient virus resistance mediated by artificial microRNAs that target the suppressor of PVX and PVY in plants. *Plant Biol (Stuttg)*, **13**: 304-16
- Allen E, Xie Z, Gustafson AM, Carrington JC. (2005) microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell*, **121**: 207–221
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, **25**: 3389–3402
- Antczak M, Kasprzak M., Lukasiak P., Blazewicz J., (2016) Structural alignment of protein descriptors - a combinatorial model, *BMC Bioinformatics*, **17**: 383
- Axtell MJ, Jan C, Rajagopalan R, Bartel DP. (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565–577
- Axtell MJ, Westholm JO, Lai EC. (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, **12**: 221
- Axtell MJ. (2013) Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol*. **64**: 137-159
- Axtell MJ, Meyers BC. (2018). Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. *The Plant cell*, **30**: 272–284.
- Baby J, Vrundha N, Chandy S. (2012) AmiRzyn: PERL Centered Artificial MicroRNA Designing Aid. *International Research Journal of Biological Sciences*, **1**: 18–23
- Bahramian MB, Zarbl H (1999) Transcriptional and posttranscriptional silencing of rodent  $\alpha 1$  (I) collagen by a homologous transcriptionally self-silenced transgene. *Mol and cell biol*.
- Bai S, Harada T, Kasai A. (2018) Application of Systemic Transcriptional Gene Silencing for Plant Breeding. *Applied RNA Bioscience*. 247-263
- Bajczyk M, Bhat SS, Szewc L, Szweykowska-Kulinska Z, Jarmolowski A, Dolata J. (2019) Novel Nuclear Functions of Arabidopsis ARGONAUTE1: Beyond RNA Interference. *Plant Physiol*. 179:1030-1039

- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, **98**: 10037–10041
- Bartel D (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**: 281–297
- Bartel DP. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**: 215–233
- Beclin C, Boutet S, Waterhouse P, Vaucheret H (2002) A branched pathway for transgene-induced RNA silencing in plants. *Curr Biol*. **12**: 684–688
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363-366
- Baumberger N, Baulcombe DC.(2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. USA* **102**: 11928–11933.
- Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC, Palatnik JF. (2013) Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome research*, **23**: 1675–1689
- Bologna NG, Iselin R, Abriata LA, Sarazin A, Pumplun N, Jay F, Grentzinger T, Dal Peraro M, Voinnet O. (2018) Nucleo-cytosolic Shuttling of ARGONAUTE1 Prompts a Revised Model of the Plant MicroRNA Pathway. *Mol Cell*. **69**: 709-719.e5.
- Boratyn GM, Schaffer A, Agarwala R, Altschul SF, Lipman DJ, Madden TL. (2012) Domain enhanced lookup time accelerated BLAST. *Biol Direct*, **7**: 12-25
- Borges F, Martienssen RA. (2015). The expanding world of small RNAs in plants. *Nature reviews. Molecular cell biology*, **16**: 727–741
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK. (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulates salt tolerance in Arabidopsis. *Cell*, **123**: 1279–1291
- Bouche N, Laressergues D, Gascioli V, Vaucheret H (2006). An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *The EMBO journal*, **25**, 3347-3356
- Bozorov TA, Baldwin IT, Kim SG. (2012) Identification and profiling of miRNAs during herbivory reveals jasmonate-dependent and -independent patterns of accumulation in *Nicotiana attenuata*. *BMC Plant Biol*. **12**: 209-222
- Brodersen P, Voinnet O. (2009) Revisiting the principles of microRNA target recognition and mode of action. *Nature Reviews Molecular Cell Biology*, **10**: 141–148
- Brooks BR, Brucoleri RE, Olafson BD, Swaminathan S, Karplus M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, **4**: 187–217
- Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. (2009) CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, **30**: 1545–1614

- Buck KW. (1996) Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv. Virus Res.* **47**, 159–251
- Burdisso P, Suarez IP, Bologna NG, Palatnik JF, Bersch B, Rasia RM. (2012) Second Double-Stranded RNA Binding Domain of Dicer-like Ribonuclease 1: Structural and Biochemical Characterization. *Biochemistry*, **51**: 10159–10166
- Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, i inni. (2012) Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell*, **24**, 3613–29
- Carbonell A, Daròs JA. (2019) Design, Synthesis, and Functional Analysis of Highly Specific Artificial Small RNAs with Antiviral Activity in Plants. *Methods Mol. Biol.* **2028**: 231–246.
- Carmell MA, Hannon GJ. (2004) RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol.* **11**, 214-218
- de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, **34**: W362–W365
- Cenik ES, Fukunaga R, Lu G, Dutcher R, Wang Y, Tanaka Hall TM, Zamore PD. (2011) Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Molecular cell*, **42**: 172–184
- Chan SWL, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. (2004) RNA silencing genes control de novo DNA methylation. *Science*, **303**: 1336
- Chau BL, Lee KA. (2007) Function and anatomy of plant siRNA pools derived from hairpin transgenes. *Plant methods*, **3**: 13-27
- Chekulaeva M, Filipowicz W. (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Current opinion in cell biology*, **21**: 452–460
- Chen J, Im W, Brooks CL. (2005). Application of torsion angle molecular dynamics for efficient sampling of protein conformations. *Journal of computational chemistry*, **26**, 1565-1578
- Chen HM, Li YH, Wu SH. (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. *Proc Natl Acad Sci U S A.* **104**: 3318-3323
- Chen X. (2009) Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol*, **25**: 21-44
- Cho SH, Addo-Quaye C, Coruh C, Arif MA, Ma Z, Frank W, Axtell MJ. (2008). *Physcomitrella patens* DCL3 is required for 22-24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. *PLoS Genet.* **4**: e1000314
- Chomczyński P, Sacchi N. (2006) *The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on.* *Nature Protocols*, **1**: 581–585

- Cogoni C, Irelan JT, Schumacher M, Schmidhauser TJ, Selker EU, Macino. (1996) Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *EMBO J*, **15**: 3153-3163
- Cole C, Barber JD, Barton GJ. (2008) The Jpred 3 secondary structure prediction server. *Nucleic acids research*, **36**: W197–W201
- Conrad C, Rauhut R. (2002) Ribonuclease III: new sense from nuisance. *Int J Biochem Cell Biol.*, **34**, 116-129
- Covey S, Al-Kaff N, Lángara A, Turner, D. (1997) Plants Combat Infection by Gene Silencing. *Nature*, **385**: 781–782
- Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, Takeda A, Sullivan CM, Gilbert SD, Montgomery TA, Carrington JC. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **17**: 997-1003
- Dai X, Zhao PX. (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic acids research*, **39**: W155–W159
- Dalmay T, Hamilton AJ, Rudd S, Angell S, Baulcombe DC (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, **101**: 543–553
- DeLano WL. (2002) The PyMOL molecular graphics system, Version 1.5.0.4 Schrödinger, LLC.
- Deng P, Muhammad S, Cao M, Wu L. (2018) Biogenesis and regulatory hierarchy of phased small interfering RNAs in plants. *Plant Biotechnology Journal*, **16**: 965-975
- Denning EJ, Priyakumar U, Nilsson L, Mackerell AD. (2011) Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *Journal of computational chemistry*, **32**: 1929–1943
- Derrien B, Clavel M, Baumberger N, Iki T, Sarazin A, Hacquard T, Ponce MR, Ziegler-Graff V, Vaucheret H, Micol JL, Voinnet O, Genschik P A (2018) Suppressor Screen for AGO1 Degradation by the Viral F-Box P0 Protein Uncovers a Role for AGO DUF1785 in sRNA Duplex Unwinding. *Plant Cell*. **30**: 1353-1374
- Ding SW, Voinnet O. (2007) Antiviral immunity directed by small RNAs. *Cell*, **130**: 413–26.
- Dolata J, Taube M, Bajczyk M, Jarmolowski A, Szweykowska-Kulinska Z, Bielewicz D (2018) Regulation of plant microprocessor function in shaping microRNA landscape. *Front Plant Sci* **9**:753
- Dolgosheina EV, Morin RD, Aksay G, Sahinalp SC, Magrini V, Mardis ER, Mattsson J, Unrau PJ. (2008). Conifers have a unique small RNA silencing signature. *RNA*, **14**: 1508–1515.
- Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*, **35**: W522–W525

- Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic acids research*, **32**: W665–W667
- Dong Q, Schlueter SD, Brendel V. (2004) PlantGDB, plant genome database and analysis tools. *Nucleic acids research*, **32**: D354–D359
- Du Z, Lee JK, Tjhen R, Stroud RM, James TL. (2008) Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proceedings of the National Academy of Sciences*, **105**: 2391–2396
- Dunoyer P., Himber C, Voinnet O. (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nature genetics*, **37**: 1356–1360
- Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, Alioua A, Himber C, Voinnet O. (2010) An endogenous, systemic RNAi pathway in plants. *The EMBO journal*, **29**: 1699–1712
- Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V. (2008) PlantGDB: a resource for comparative plant genomics. *Nucleic acids research*, **36**: 959–D965
- Eamens AL, Smith NA, Curtin SJ, Wang MB, Waterhouse PM. (2009) The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA*, **15**: 2219–2235
- Elbashir SM, Lendeckel W, Tuschl T. (2001) RNA interference is mediated by 21 and 22 nt RNAs. *Genes Dev.* **15**: 188–200
- Elkayam E, Faehnle CR, Morales M, Sun J, Li H, Joshua-Tor L. (2017) Multivalent Recruitment of Human Argonaute by GW182. *Molecular Cell*, **67**: 646-658
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. (2007) Comparative protein structure modeling using MODELLER. *Current Protocols in Protein Science*, **2**: Unit 2.9
- Fahlgren, N, Hill ST, Carrington JC, Carbonell A. (2016) P-SAMS: a web site for plant artificial microRNA and synthetic trans-acting small interfering RNA design. *Bioinformatics*, **32**: 157-158
- de Felippes FF. (2019). Gene Regulation Mediated by microRNA-Triggered Secondary Small RNAs in Plants. *Plants (Basel, Switzerland)*, **8**: 112
- Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. **391**: 806-11
- Foloppe N, MacKerell Jr AD. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry*, **21**: 86–104
- Fukudome A, Kanaya A, Egami M, Nakazawa Y, Hiraguri A, Moriyama H, Fukuhara T. (2011) Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of Arabidopsis Dicer-like 4. *RNA*, **17**: 750–760



- Fukudome A, Fukuhara T. (2017) Plant dicer-like proteins: double-stranded RNA-cleaving enzymes for small RNA biogenesis. *J Plant Res*, **130**: 33–44
- Fukunaga R, Colpan C, Han BW, Zamore PD. (2014) Inorganic phosphate blocks binding of pre-miRNA to Dicer-2 via its PAZ domain. *The EMBO journal*, **33**: 371–384
- Gan J, Shaw G, Tropea JE, Waugh DS, Ji X, i inni. (2008) A stepwise model for double-stranded RNA processing by ribonuclease III. *Molecular microbiology*, **67**: 143–154
- Gan J, Tropea JE, Austin BP, Waugh DS, Ji X, i inni. (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, **124**: 355–366
- Gascioli V, Mallory AC, Bartel DP, Vaucheret H. (2005) Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. *Current Biology*, **15**: 1494–1500
- Gebert LFR, MacRae IJ. (2019) Regulation of microRNA function in animals. *Nature Reviews Molecular Cell Biology*, **20**: 21–37
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. (2008) miRBase: tools for microRNA genomics. *Nucleic acids research*, **36**: D154–D158
- Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, **18**: 2714–2723
- Guo Z, Wang XB, Wang Y, Li WX, Gal-On A, Ding SW. (2018) Identification of a new host factor required for antiviral RNAi and amplification of viral siRNAs. *Plant Physiol*. **176**: 1587–1597.
- Gurtan AM, Lu V, Bhutkar A, Sharp PA. (2012) In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. *RNA*, **18**: 1116–1122
- Gustafson AM, Allen E, Givan S, Smith D, Carrington JC, Kasschau KD. (2005) ASRP: the Arabidopsis small RNA project database. *Nucleic acids research*, **33**: D637–D640
- Hall TA. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*, **41**: 95–98
- Hamilton AJ, Baulcombe DC. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, **286**: 950–952
- Han MH, Goud S, Song L, Fedoroff N. (2004) The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci USA*, **101**: 1093–1098
- Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**: 887–901
- Hannon GJ. (2002). RNA interference. *Nature*, **418**: 244–251
- Hauptmann J, Dueck A, Harlander S, Pfaff J, Merkl R, Meister G. (2013) Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nat. Struct. Mol. Biol.* **20**: 814–817

- Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, i inni. (2010) The Arabidopsis RNA-directed DNA methylation Argonautes functionally diverge based on their expression and interaction with target loci. *Plant Cell*, **22**, 321–34
- Hedges SB, Blair JE, Venturi ML, Shoe JL. (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* **4**: 1471–2148
- Henz SR, Cumbie JS, Kasschau KD, Lohmann JU, Carrington JC, Weigel D, Schmid M. (2007) Distinct expression patterns of natural antisensetranscripts in Arabidopsis. *Plant Physiol*, **144**: 1247–1255
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680
- Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC, Kneller GR. (2000) Harmonicity in slow protein dynamics, *Chem. Phys.* **261**: 25-37
- Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, Koiwa H, Seki M, Shinozaki K, Fukuhara T. (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. *Plant molecular biology*, **57**: 173–188
- Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC. (2007) Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell*. **19**: 926-942
- Höck J, Meister G. (2008) The Argonaute protein family. *Genome Biol* **9**, 210-222
- Huang Y, Ji L, Huang Q, Vassilyev DG, Chen X, Ma JB. (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature*. **461**: 823-7
- Humphrey W, Dalke A, Schulten K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics*, **14**: 33–38
- Iki T. (2017) Messages on small RNA duplexes in plants. *Journal of Plant Research*, **130**: 7–16
- Ilesh Shah M, Joseph J, Rajagopalan A, Sivaprakasam M. (2017) ImageQuant: An image-based quantitative Immunoassay Analyzer. *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Rochester, MN, USA, **2017**, pp. 420-425
- Im W, Lee MS, Brooks III CL. (2003) Generalized Born Model with a Simple Smoothing Function. *Journal of computational chemistry* **24**:1691-1702
- Jagannadham PTK, Muthusamy SK, Chidambaranathan P. (2019) Micromics: A Novel Approach to Understand the Molecular Mechanisms in Plant Stress Tolerance. *Recent Approaches in Omics for Plant Resilience to Climate Change*. 93-108
- Jensen S, Gassama MP, Heidmann T. (1999) Cosuppression of I transposon activity in Drosophila by I-containing sense and antisense transgenes. *Genetics*. **153**: 1767-74.
- Ji L, Liu X, Yan J, Wang W, Yumul RE, i inni. (2011) ARGONAUTE10 and ARGONAUTE1 Regulate the Termination of Floral Stem Cells through Two MicroRNAs in Arabidopsis. *PLoS Genet* **7**: e1001358

- Johanson TM, Lew AM, Chong MM. (2013) MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. *Open Biol.* **3**: 130144.
- Kapoor M, Arora R, Lama T, Nijhawan A, Khurana J, Tyagi A, Kapoor S. (2008) Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC genomics*, **9**: 451–457
- Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A Jr, Zhu JK, Staskawicz BJ, Jin H. (2006) A pathogen-inducible endo-genous siRNA in plant immunity. *Proc Natl Acad Sci USA* **103**: 18002–18007
- Khvorova A, Reynolds A, Jayasena SD. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**: 209–216
- Kim JY, Kwak KJ, Jung HJ, Lee HJ, Kang H. (2010) MicroRNA402 affects seed germination of *Arabidopsis thaliana* under stress conditions via targeting DEMETER-LIKE Protein3 mRNA. *Plant and cell physiology*, **51**: 1079–1083
- Kim VN. (2005) Small RNAs: classification, biogenesis, and function. *Mol Cells*, **19**: 1–15
- Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ. (2004) Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *Journal of Biological Chemistry*, **279**: 42230
- Kurowski MA, Bujnicki JM. (2003) GeneSilico protein structure prediction meta-server. *Nucleic Acids Research*, **31**: 3305–3307
- Kurihara Y, Watanabe Y. (2004) *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 12753
- Kwak PB, Tomari Y. (2012) The N domain of Argonaute drives du-plex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* **19**: 145-151
- Kwon SC, Nguyen TA, Choi YG, Jo MH, Hohng S, Kim VN, Woo JS. (2016) Structure of Human DROSHA. *Cell*. **164**: 81-90
- Laganà A, Acunzo M, Romano G, Pulvirenti A, Veneziano D, Cascione L, Giugno R, Gasparini P, Shasha D, Ferro A, i inni. (2014) miR-Synth: a computational resource for the design of multi-site multi-target synthetic miRNAs. *Nucleic acids research*, gku202
- Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, i inni. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**: 2947–2948
- Lau PW, Guiley KZ, De N, Potter CS, Carragher B, MacRae IJ. (2012) The molecular architecture of human Dicer. *Nature structural & molecular biology*, **19**: 436–440
- Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. (2004) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*, **23**: 4051–4060

- Li H, Robertson AD, Jensen JH. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics*, **61**: 704–721
- Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B. (2012). MicroRNA regulation of plant innate immune receptors. *Proc. Natl. Acad. Sci. USA* **109**: 1790–1795
- Li JF, Chung HS, Niu Y, Bush J, McCormack M, Sheen J. (2013) Comprehensive protein-based artificial microRNA screens for effective gene silencing in plants. *The Plant Cell Online*, **25**: 1507–1522
- Liang YH, Lavoie M, Comeau MA, Abou Elela S, Ji X. (2014) Structure of a Eukaryotic RNase III Postcleavage Complex Reveals a Double-Ruler Mechanism for Substrate Selection. *Molecular cell*, **54**: 431–444
- Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes & development*, **17**: 991–1008
- Lindbo JA, Dougherty WG. (1992) Pathogen-derived resistance to a potyvirus: immune and resistant phenotypes in transgenic tobacco expressing altered forms of a potyvirus coat protein nucleotide sequence. *Mol Plant Microbe Interact.* **5**: 144–153
- Lingel A, Simon B, Izaurralde E, Sattler M. (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature structural & molecular biology*, **11**: 576–577
- Liu Q., Yao X., Pi L., Wang H., Cui X., Huang H. (2009) The ARGONAUTE10 gene modulates shoot apical meristem maintenance and leaf polarity establishment by repressing miR165/166 in Arabidopsis. *Plant J.* **58**: 27-40
- Liu C, Axtell MJ, Fedoroff NV. (2012a) The helicase and RNaseIIIa domains of Arabidopsis Dicer-Like1 modulate catalytic parameters during microRNA biogenesis. *Plant physiology*, **159**: 748–758
- Liu F, Bakht S, Dean C. (2012b) Cotranscriptional role for Arabidopsis DICER-LIKE 4 in transcription termination. *Science*, **335**: 1621–1623
- Liu H, Yu H, Tang G, Huang T. (2018a) Small but powerful: function of microRNAs in plant development. *Plant Cell Rep.* **37**: 515-528
- Liu Z, Jia Wang J, Cheng H, Ke X, Sun L, Zhang QC, Wang HW. (2018b) Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate. *Cell*, **173**: 1191-1203.
- Llave C. (2010) Virus-derived small interfering RNAs at the core of plant-virus interactions. *Trends Plant Sci*, **15**: 701–707
- Lobbes D, Rallapalli G, Schmidt DD, Martin C, Clarke J. (2006) SERRATE: A new player on the plant microRNA scene. *EMBO Rep*, **7**, 1052–1058.
- Lu S, Sun YH, Amerson H and Chiang VL. (2007), MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development, *Plant J*, **51**: 1077–1098
- Luo QJ, Mittal A, Jia F, Rock CD. (2012) An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis. *Plant Mol Biol.* **80**: 117-129

- Ma JB, Ye K, Patel DJ. (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, **429**: 318–322
- Ma E, Zhou K, Kidwell MA, Doudna JA. (2012) Coordinated activities of human dicer domains in regulatory RNA processing. *Journal of molecular biology*, **422**: 466–476
- Machida S, Chen HY, Adam Yuan Y. (2011) Molecular insights into miRNA processing by *Arabidopsis thaliana* SERRATE, *Nucleic Acids Research*, **39**: 7828–7836
- MacKerell AD, Bashford D, Bellott M, Dunbrack R, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, i inni. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, **102**: 3586–3616
- MacKerell AD, Feig M, Brooks CL. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry*, **25**: 1400–1415
- MacRae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. (2006) Structural basis for double-stranded RNA processing by Dicer - Suppl. *Science*, **311**: 195–198
- MacRae IJ, Zhou K, Doudna JA. (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nature structural & molecular biology*, **14**: 934–940
- Mallory AC, Vaucheret H. (2004) MicroRNAs: something important between the genes. *Current opinion in plant biology*, **7**: 120–125
- Mallory AC, Vaucheret H. (2006) Functions of microRNAs and related small RNAs in plants. *Nat Genet*. **38**: 31-36
- Mallory AC, Hinze A, Tucker MR, Bouche N, Gascioli V, Elmayan T, Laressergues D, Jauvion V, Vaucheret H, Laux T. (2009) Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. *PLoS Genet*. **5**:E1000646-E1000646
- Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM. (2006) The evolution and diversification of Dicers in plants. *FEBS Lett*. **580**: 2442–2450.
- Markham NR, Zuker M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Bioinformatics*, **453**: 3–31
- Martínez-Jiménez del Carmen V, Méndez-Mancilla A, Portales-Pérez DP.(2018) miRNAs in nutrition, obesity, and cancer: The biology of miRNAs in metabolic disorders and its relationship with cancer development. *Mol. Nutr. Food Res*. **62**: 1600994
- Mateos JL, Bologna NG, Chorostecki U, Palatnik J.F (2010) Identification of MicroRNA Processing Determinants by Random Mutagenesis of Arabidopsis MIR172a Precursor. *Current Biology*, **20**: 49-54
- Matsui T, Segall J, Weil A, Roeder RG. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J. Biol. Chem*. **255**: 11992- 11996

- Meiler J, Baker D. (2003) Coupled prediction of protein secondary and tertiary structure. *Proceedings of the National Academy of Sciences*, **100**: 12105–12110
- Meola N, Gennarino VA, Banfi S. (2009) microRNAs and genetic diseases. *Pathogenetics*, **2**: 7
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, inni (2008) Criteria for annotation of plant microRNAs. *Plant Cell*, **20**: 3186–3190
- Mickiewicz A, Rybarczyk A, Sarzynska J, Figlerowicz M, Blazewicz J. (2016) AmiRNA Designer - new method of artificial miRNA design. *Acta Biochim. Pol.* **63**: 71–77
- Mickiewicz A, Sarzyńska J, Miłostan M, Kurzyńska-Kokorniak A, Rybarczyk A, Łukasiak P, Kuliński T, Figlerowicz M, Błażewicz J. (2017). Modeling of the catalytic core of Arabidopsis thaliana Dicer-like 4 protein and its complex with double-stranded RNA. *Comput. Biol. Chem.* **66**: 44-56.
- Miśkiewicz J, Tomczyk K, Mickiewicz A, Sarzynska J, Szachniuk M. (2017) Bioinformatics study of structural patterns in plant microRNA precursors. *Biomed Res Int.*
- Miśkiewicz J, Szachniuk M. (2018) Discovering Structural Motifs in miRNA Precursors from the Viridiplantae Kingdom *Molecules*, **23**: 1367.
- Mlotshwa S, Pruss GJ, Peragine A, Endres MW, Li J, Chen X, Poethig RS, Bowman LH, Vance V. (2008) DICER-LIKE2 plays a primary role in transitive silencing of transgenes in Arabidopsis. *PLoS One*, **3**: e1755
- Moazed D. (2009) Small RNAs in transcriptional gene silencing and genome defence. *Nature*, **457**: 413-420
- Molnar A, Bassett A, Thuenemann E, Schwach F, Karkare S, Ossowski S, Weigel D, Baulcombe D. (2009) Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. *Plant J.* **58**(1), 165-174
- Morel JB, Godon C, Mourrain P, Béclin C, Boutet S, Feuerbach F, Proux F, Vaucheret H. (2002). Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. *Plant Cell*, **14**: 629–639
- Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel J-B, Jouette D, Lacombe AM, Nikic S, Picault N, et al (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell*, **101**: 533–542
- Mukherjee K, Campos H, Kolaczowski B. (2013) Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Molecular biology and evolution*, **30**: 627–641
- Nakanishi K, Ascano M, Gogakos T, Ishibe-Murakami S, Serganov AA, Briskin D, Morozov P, Tuschl T, Patel DJ. (2013) Eukaryote-Specific Insertion Elements Control Human ARGONAUTE Slicer Activity. *Cell reports*, **3**: 1893–1900
- Napoli C, Lemieux C, Jorgensen R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell*. **2**: 279-289

- Nicholas KB, Nicholas H, and Deerfield D. (1997) GeneDoc: analysis and visualization of genetic variation. *EMBnet.news* 1997, **4**: 1-4
- Needleman SB, Wunsch CD. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* **48**: 443-53
- Nobuta, K, Lu C, Shrivastava R, Pillay M, De Paoli E, Accerbi M, i inni (2008). Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. *Proc. Natl. Acad. Sci. USA* **105**: 14958–14963
- Notredame C, Higgins DG, Heringa J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, **302**: 205–217
- Ossowski S, Schwab R, Weigel D. (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. *The Plant Journal*, **53**: 674–690
- Pagliarani C; Gambino G.(2019) Small RNA Mobility: Spread of RNA Silencing Effectors and its Effect on Developmental Processes and Stress Adaptation in Plants. *Int. J. Mol. Sci.* **20**: 4306.
- Park W, Li J, Song R, Messing J, Chen X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**: 1484-1495
- Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS. (2005) Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc Natl Acad Sci USA*, **102**: 3691–3696
- Park J-E, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, **475**: 201–205
- Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM. (2008) MetaMQAP: a meta-server for the quality assessment of protein models. *BMC bioinformatics*, **9**: 403
- Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**: 2368–2379
- Peters L, Meister G. (2007). Argonaute proteins: mediators of RNA silencing. *Mol. Cell*, **26**: 611–623
- Petri S, Meister G. (2013) siRNA design principles and off-target effects. *Target Identification and Validation in Drug Discovery*. 59–71
- Pirritano M. Götz U; Karunanithi S. Nordström, K. Schulz M.H. Simon M. (2018) Environmental Temperature Controls Accumulation of Transacting siRNAs Involved in Heterochromatin Formation. *Genes.* **9**: 117.
- Polydore S,. Axtell M.J (2018) Analysis of RDR1/RDR2/RDR6-independent small RNAs in *Arabidopsis thaliana* improves MIRNA annotations and reveals novel siRNA loci. *Plant J* **94**: 1051–1063
- Pooggin M.M. (2017) RNAi-mediated resistance to viruses: a critical assessment of methodologies. *Current Opinion in Virology*, **26**: 28-35
- Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Łukasiak P, Bartol N, Błazewicz J, Adamiak RW. (2012) Automated 3D structure composition for large RNAs, *Nucleic Acids Research*, **40**: 1–12

- Prasad A, Sharma N, Muthamilarasan M, Rana S, Prasad S. (2019) Recent advances in small RNA mediated plant-virus interactions. *Critical Reviews in Biotechnology*, **39**:587-601
- Pumplin N, Voinnet O. RNA silencing suppression by plant pathogens: defence, counter- defence and counter-counter-defence. *Nat Rev Microbiol*. 2013;11(11):745–60.
- Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA. (2010) Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA*, **16**: 474–481
- Qu F., Ye X., Morris T.J (2008) Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 14732-14737
- Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature*, **443**: 1008-1012
- Rajeswaran R, Pooggin MM. (2012) RDR6-mediated synthesis of complementary RNA is terminated by miRNA stably bound to template RNA. *Nucleic acids research*, **40**: 594–599
- Ratcliff F, Harrison B, Baulcombe DC. (1997) A Similarity between Viral Defense and Gene Silencing in Plants. *Science*, **276**: 1558–1560
- Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. (2002) MicroRNAs in plants. *Genes Dev*. **16**: 1616-1626
- Rodrigues JP, Levitt M, Chopra G. (2012) KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Research*, **40**: W323–W328
- Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S. (2010) Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. *Genes Dev*. **24**: 1010–1021
- Roy A, Kucukural A, Zhang Y. (2010) I-TASSER: a unified Platform for automated protein structure and function prediction. *Nature protocols*, **5**: 725–738
- Ruiz MT, Voinnet O, Baulcombe DC. (1998) *Initiation and maintenance of virus-induced gene silencing. Plant Cell*, **10**: 937-946.
- Šali A, Blundell TL. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815
- Sarzyńska J, Mickiewicz A, Miłostan M, Łukasiak P, Błażewicz J, Figlerowicz M, Kuliński T. (2010) Flexibility of Dicer Studied by Implicit Solvent Molecular Dynamics Simulations. *Computational Methods in Science and Technology*, **16**: 97–104.
- Sasaki T, Shimizu N. (2007) Evolutionary conservation of a unique amino acid sequence in human DICER protein essential for binding to Argonaute family proteins. *Gene*, **396**: 312–320
- Schiebel W, Haas B, Marinkovic S, Klanner A, Sanger HL (1993b) RNA-directed RNA polymerase from tomato leaves. II. Catalytic in vitro properties. *J Biol Chem*. **268**: 11858–11867



- Schirle NT, Sheu-Gruttadauria J, MacRae IJ. (2014) Gene regulation. Structural basis for microRNA targeting. *Science*, **346**: 608-13
- Schultz J, Milpetz F, Bork P, Ponting CP. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences*, **95**: 5857–5864
- Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D. (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. *The Plant Cell Online*, **18**: 1121–1133.
- Shabalina SA, Koonin EV. (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol*. **23**: 578-87.
- Shen W, Smith NA, Zhou C, and Wang M. (2014) High-Level Accumulation of Exogenous Small RNAs Not Affecting Endogenous Small RNA Biogenesis and Function in Plants. *Journal of Integrative Agriculture*, **13**: 1017–1023
- Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC. (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell*. **24**: 859-874
- Silva D, Jemc C. (2015) Sorting Out Identities: An Educational Primer for Use with “Novel Tools for Genetic Manipulation of Follicle Stem Cells in the Drosophila Ovary Reveal an Integrin-Dependent Transition from Quiescence to Proliferation. *Genetics*, **201**: 13-22
- Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, Sehr P, Pillai RS, Carlomagno T. (2011) Recognition of 2'-O-Methylated 3'-End of piRNA by the PAZ Domain of a Piwi Protein. *Structure*, **19**: 172–180
- Song L, Axtel MJ, Fedoroff NV. (2010) RNA Secondary Structural Determinants of miRNA Precursor Processing in Arabidopsis. *Current Biology*, **20**: 37-41
- Song MS, Rossi JJ. (2017). Molecular mechanisms of Dicer: endonuclease and enzymatic activity. *The Biochemical journal*, **474**: 1603–1618
- Stein P, Svoboda P, Anger M, Schultz RM. (2003) RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA*, **9**: 187-92
- Suhre K, Sanejouand YH. (2004) Elnemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research*, **32**: W610–W614
- Sunkar R, Zhu JK. (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *The Plant Cell Online*, **16**: 2001–2019
- Sunkar R, Li YF, Jagadeeswaran G. (2012) Functions of microRNAs in plant stress responses. *Trends in plant science*, **17**: 196–203
- Szarzynska B, Sobkowiak L, Pant BD, Balazadeh S, Scheible WR, Mueller-Roeber B, Jarmolowski A, Szweykowska-Kulinska Z (2009) Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. *Nucleic Acids Res*. **37**:3083–3093

- Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**: 123–132
- Takeda A, Iwasaki S, Watanabe T, Utsumi M, Watanabe Y. (2008) The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol.* **49**: 493–500
- Takeshita D, Zenno S, Lee WC, Nagata K, Saigo K, Tanokura M. (2007) Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human dicer. *Journal of molecular biology*, **374**: 106–120
- Taylor DW, Ma E, Shigematsu H, Cianfrocco MA, Noland CL, Nagayama K, Nogales E, Doudna JA, Wang H-W. (2013) Substrate-specific structural rearrangements of human Dicer. *Nature structural & molecular biology*, **20**: 662–670
- Tian Y, Simanshu DK, Ma JB, Park JE, Heo I, Kim VN, Patel DJ. (2014) A Phosphate-Binding Pocket within the Platform-PAZ-Connector Helix Cassette of Human Dicer. *Molecular cell*, **53(4)**:606–616
- Vaucheret H. (2006). Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes Dev.* **20**: 759–771
- Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gascioli V, Mallory AC, Hilbert JL, Bartel DP, Cr  t   P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol. Cell*, **16**: 69–79
- Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review*, **159**: 98
- Voinnet O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**: 669–687
- Vu TV, Choudhury NR, Mukherjee SK. (2013) Transgenic tomato plants expressing artificial microRNAs for silencing the pre-coat and coat proteins of a begomovirus, Tomato leaf curl New Delhi virus, show tolerance to virus infection. *Virus Res.* **172**: 35–45
- Wallner B, Elofsson A. (2003) Can correct protein models be identified? *Protein Science*, **12**: 1073–1086
- Wallner B, Elofsson A (2005) Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics*, **21**: 4248–4254
- Walter AE, Turner DH, Kim J, Lyttle MH, M  ller P, Mathews DH, Zuker M. (1994) Coaxial stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proceedings of the National Academy of Sciences*, **91**: 9218–9222
- Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, **461**: 754–761
- Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, Ecker JR, Tang H, Pikaard CS. (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev*, **26**: 1825–1836

- Wu L, Zhou H, Zhang Q, Zhang J, Ni F, Liu C, Qi Y. (2010). DNA methylation mediated by a microRNA pathway. *Mol. Cell*, **38**: 465–475
- Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants. *PLoS Biol.* **2**: 0642–0652
- Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC. (2005) Expression of Arabidopsis MIRNA genes. *Plant Physiology*, **138**: 2145–2154
- Xie, Z., Cheng, H. (2017) Interplay and transition between small RNA-directed posttranscriptional and transcriptional gene silencing in plants. *Ind J Plant Physiol.* **22**: 371.
- Yang L, Liu Z, Lu F, Dong A, Huang H (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J*, **47**, 841–850
- Yang X, Ren W, Zhao Q, Zhang P, Wu F, He Y. (2014) Homodimerization of HYL1 ensures the correct selection of cleavage sites in primary miRNA. *Nucleic Acids Res.* **42**: 12224–12236
- Yoshikawa M, Peragine A, Park MY, Poethig RSA. (2005) Pathway for the biogenesis of *trans*-acting siRNAs in Arabidopsis. *Genes Dev.* **19**: 2164–2175
- Yu D, Fan B, MacFarlane SA, Chen Z. (2003) Analysis of the involvement of an inducible Arabidopsis RNA-dependent RNA polymerase in antiviral defense. *Mol Plant Microbe Interact*, **16**: 206–216
- Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935.
- Yuan C, Wang J, Harrison AP, Meng X, Chen D, Chen M. (2015). Genome-wide view of natural antisense transcripts in Arabidopsis thaliana. *DNA research : an international journal for rapid publication of reports on genes and genomes*, **22**: 233–243
- Zamore PD, Haley B. (2005). Ribo-gnome: The big world of small RNAs. *Science*, **309**: 1519–1524
- Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, i inni (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* **25**: 2540–2553
- Zhang H, Kolb F, Jaskiewicz L, Westhof E, Filipowicz W. (2004), Single processing center models for human Dicer and bacterial RNase III. *Cell*, **118** 57–68
- Zhang X, Zhao H, Gao S, Wang WC, Katiyar-Agarwal S, Huang HD, Raikhel N, Jin H. (2011) Arabidopsis Argonaute 2 regulates innate immunity via miRNA393(\*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12. *Mol Cell*, **42**, 356–366.
- Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, Zhou X, Gao S, Lu L, Niu D, Chen Z, Leung C, i inni (2012) Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biol*, **13**: R20
- Zhang C, Wu Z, Li Y, Wu J. (2015) Biogenesis, function, and applications of virus-derived small RNAs in plants. *Front. Microbiol.* **6**: 1237

- Zhang Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, **9**: 40–48.
- Zheng X, Zhu J, Kapoor A, Zhu JK. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* **26**: 1691–1701.
- Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, Tu A, Jacobsen SE.(2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons *Nat. Struct. Mol. Biol.*, **19**: 870-875
- Żok T, Popenda M, Szachniuk M. (2014) MCQ4Structures to compute similarity of molecule structures. *Central European Journal of Operations Research*, **22**:457-473

## 9. Załączniki

**Załącznik nr 1.** Sekwencje miRNA wykorzystane do wyznaczenia zależności sekwencyjnych oraz profili termodynamicznych

Nazwa z miRBase	Sekwencja	start	koniec	ramię	długość
MI0000184	UUGACAGAAGAUAGAGAGCAC	4	24	5prim	21
MI0000185	UUGACAGAAGAUAGAGAGCAC	19	39	5prim	21
MI0000186	UUGACAGAAGAUAGAGAGCAC	24	44	5prim	21
MI0000189	UUUGGAUUGAAGGGAGCUCUA	163	183	3prim	21
MI0000190	UGCCUGGCUCCCUGUAUGCCA	4	24	5prim	21
MI0000191	UGCCUGGCUCCCUGUAUGCCA	5	25	5prim	21
MI0000192	UGCCUGGCUCCCUGUAUGCCA	5	25	5prim	21
MI0000193	UCA AUGCAUUGAAAGUGACUA	38	58	5prim	21
MI0000193_47_67	UGAAAGUGACUACAUCGGGGU	47	67	5prim	21
MI0000194	UCGAUAAACCUCUGCAUCCAG	89	109	3prim	21
MI0000195	UCGAUAAACCUCUGCAUCCAG	79	99	3prim	21
MI0000197	UGGAGAAGCAGGGCACGUGCA	19	39	5prim	21
MI0000198	UGGAGAAGCAGGGCACGUGCA	3	23	5prim	21
MI0000199	UCGGACCAGGCUUCAUCCCC	86	106	3prim	21
MI0000200	UCGGACCAGGCUUCAUCCCC	137	157	3prim	21
MI0000201	UCGGACCAGGCUUCAUCCCC	134	154	3prim	21
MI0000202	UCGGACCAGGCUUCAUCCCC	96	116	3prim	21
MI0000203	UCGGACCAGGCUUCAUCCCC	103	123	3prim	21
MI0000204	UCGGACCAGGCUUCAUCCCC	87	107	3prim	21
MI0000205	UCGGACCAGGCUUCAUCCCC	119	139	3prim	21
MI0000206	UCGGACCAGGCUUCAUCCCC	78	98	3prim	21

MI0000207	UCGGACCAGGCUUCAUUCCCC	85	105	3prim	21
MI0000208	UGAAGCUGCCAGCAUGAUCUA	19	39	5prim	21
MI0000209	UGAAGCUGCCAGCAUGAUCUA	10	30	5prim	21
MI0000210	UCGCUUGGUGCAGGUCGGGAA	18	38	5prim	21
MI0000211	UCGCUUGGUGCAGGUCGGGAA	19	39	5prim	21
MI0000212	CAGCCAAGGAUGACUUGCCGA	18	38	5prim	21
MI0000213	UGAUUGAGCCGUGCAAUAUC	59	79	3prim	21
MI0000214	UGAUUGAGCCGCGCCAAUAUC	88	108	3prim	21
MI0000215	AGAAUCUUGAUGAUGCUGCAU	78	98	3prim	21
MI0000216	AGAAUCUUGAUGAUGCUGCAU	71	91	3prim	21
MI0000218	UUUGGAUUGAAGGGAGCUCUU	168	188	3prim	21
MI0000544	UUGGACUGAAGGGAGCUCCCU	154	174	3prim	21
MI0000976	CAGCCAAGGAUGACUUGCCGG	43	63	5prim	21
MI0000977	CAGCCAAGGAUGACUUGCCGG	113	133	5prim	21
MI0000978	UGAGCCAAGGAUGACUUGCCG	41	61	5prim	21
MI0000979	UGAGCCAAGGAUGACUUGCCG	69	89	5prim	21
MI0000980	UGAGCCAAGGAUGACUUGCCG	32	52	5prim	21
MI0000981	UGAGCCAAGGAUGACUUGCCG	144	164	5prim	21
MI0000981_204_224	UCCGGCAAGUUGACCUUGGCU	204	224	3prim	21
MI0000982	UAGCCAAGGAUGACUUGCCUG	46	66	5prim	21
MI0000983	UAGCCAAGGAUGACUUGCCUG	40	60	5prim	21
MI0000984	UAGCCAAGGAUGACUUGCCUG	51	71	5prim	21
MI0000985	UAGCCAAGGAUGACUUGCCUG	43	63	5prim	21
MI0000986	UAGCCAAGGAUGACUUGCCUG	44	64	5prim	21
MI0000987	UAGCCAAGGAUGACUUGCCUG	42	62	5prim	21
MI0000988	UAGCCAAGGAUGACUUGCCUG	45	65	5prim	21
MI0000989	UUGAGCCGUGCCAAUAUCACG	78	98	3prim	21

MI0000990	UUGAGCCGUGCCAAUAUCACG	90	110	3prim	21
MI0000991	AGAAUCUUGAUGAUGCUGCAG	98	118	3prim	21
MI0000992	AGAAUCUUGAUGAUGCUGCAG	88	108	3prim	21
MI0001000	AAGCUCAGGAGGGAUAGCGCC	18	38	5prim	21
MI0001001	AAGCUCAGGAGGGAUAGCGCC	15	35	5prim	21
MI0001002	UUCGCAGGAGAGAUAGCGCCA	68	88	5prim	21
MI0001007	CUGAAGUGUUUGGGGGAACUC	63	83	3prim	21
MI0001008	CUGAAGUGUUUGGGGGGACUC	70	90	3prim	21
MI0001009	CUGAAGUGUUUGGGGGGACUC	70	90	3prim	21
MI0001010	CUGAAGUGUUUGGGGGAACUC	70	90	3prim	21
MI0001011	CUGAAGUGUUUGGGGGAACUC	65	85	3prim	21
MI0001012	CUGAAGUGUUUGGGGGGACUC	82	102	3prim	21
MI0001013	UCCACAGCUUUCUUGAACUG	11	31	5prim	21
MI0001014	UCCACAGCUUUCUUGAACUU	11	31	5prim	21
MI0001015	UCAUUGAGUGCAGCGUUGAUG	11	31	5prim	21
MI0001016	UCAUUGAGUGCAUCGUUGAUG	11	31	5prim	21
MI0001017	UGUGUUCUCAGGUCACCCUU	75	95	3prim	21
MI0001018	UGUGUUCUCAGGUCACCCUG	86	106	3prim	21
MI0001019	UGUGUUCUCAGGUCACCCUG	85	105	3prim	21
MI0001020	UGCCAAAGGAGAUUUGCCCUG	93	113	3prim	21
MI0001021	UGCCAAAGGAGAGUUGCCCUG	105	125	3prim	21
MI0001022	UGCCAAAGGAGAGUUGCCCUG	84	104	3prim	21
MI0001023	UGCCAAAGGAGAUUUGCCCCG	70	90	3prim	21
MI0001024	UGCCAAAGGAGAUUUGCCUCG	79	99	3prim	21
MI0001025	UGCCAAAGGAGAUUUGCCCCG	88	108	3prim	21
MI0001069	UAUGAGAGUAUUAUAAGUCAC	12	32	5prim	21
MI0001070	CGAAACUGGUGUCGACCGACA	108	128	5prim	21

MI0001072	UUAGAUUCACGCACAAACUCG	103	123	3prim	21
MI0001079	UUUAAAUCAUUAUCUUUUGGU	72	92	5prim	21
MI0001080	AUGCACUGCCUCUCCUGGC	120	140	3prim	21
MI0001085	UUUGGAUUGAAGGGAGCUCCU	195	215	3prim	21
MI0001086	UUGGACUGAAGGGAGCUCCU	170	190	3prim	21
MI0001087	UGGAGAAGCAGGGCACGUGCG	11	31	5prim	21
MI0001088	UAAGCUGCCAGCAUGAUCUUG	12	32	5prim	21
MI0001089	GGAAUCUUGAUGAUGCUGCAU	95	115	3prim	21
MI0001424	AUAGUUUCUCUUGUUCUGCAC	6	26	5prim	21
MI0001425	UCAUCUUCAUCAUCUGUCA	25	45	5prim	21
MI0001426	AACAGAGCAGAAACAGAACAU	3	23	5prim	21
MI0001427	GGUUCGUACGUACACUGUUCA	57	77	3prim	21
MI0001428	GAAGGUAGUGAAUUUGUUCGA	78	98	3prim	21
MI0001429	UAAUGUGAUGAUGAACUGACC	48	68	3prim	21
MI0001430	UUAUGAAUGCUGAGGAUGUUG	4	24	5prim	21
MI0001431	UAAACUAAUCACGGAAAUGCA	4	24	5prim	21
MI0001441	UUUUGGAAAUUGUCCUACG	14	34	5prim	21
MI0005103	UUUGCUUCCAGCUUUUGUCUC	63	83	3prim	21
MI0005104	UUGGUUACCAUAUGGCAUC	70	90	3prim	21
MI0005106	UCUAAGUCUUCUAUUGAUGUU	82	102	3prim	21
MI0005108	UGGCUUGGUUAUGUACACCG	7	27	5prim	21
MI0005109	UGAUUGGAAAUUCGUUGACU	106	126	3prim	21
MI0005109_142_162	UUCUGCUAUGUUGCUGCUCAU	142	162	3prim	21
MI0005110_150_170	UUCUUCGUGAAUAUCUGGCAU	150	170	3prim	21
MI0005111	UUAGAGUUUUCUGGAUACUUA	6	26	5prim	21
MI0005112	ACAAACACCUUGGAUGUUCUU	12	32	5prim	21
MI0005113	AAGCUUUGCUCGUUCAUGUUC	269	289	3prim	21



MI0005379	UGC GGGAAGCAUUUGCACAUG	50	70	5prim	21
MI0005380	UGGGUGGUGAUCAUAUAAGAU	120	140	3prim	21
MI0005381	UUCUCAAGAAGGUGCAUGAAC	115	135	3prim	21
MI0005383	UUAGAUGACCAUCAACAAACU	106	126	3prim	21
MI0005385	CAAAUUAAGCUUCAAGGUAG	134	154	3prim	21
MI0005386	UAACUAUUUUUGAGAAGAAGUG	124	144	3prim	21
MI0005388	UGCUGGGAUCGGGAUUCGAAA	50	70	5prim	21
MI0005388_266_286	UUGAUUCCCAUCCAAGCAAG	266	286	3prim	21
MI0005390	UGGUAGCAGUAGCGGUGGUAA	178	198	3prim	21
MI0005391	UGGAGAAGAUACGCAAGAAAG	363	383	3prim	21
MI0005391_50_70	UUCUUGCAUAUGUUCUUAUC	50	70	5prim	21
MI0005393	AAACGAACAAAAACUGAUGG	113	133	3prim	21
MI0005393_77_75	AUCAGUUUCUUGUUCGUUUCA	55	75	5prim	21
MI0005394	UUUUCUUCUACUUCUUGCACA	136	156	3prim	21
MI0005395	UACCAACCUUCAUCGUUCCC	50	70	5prim	21
MI0005397	UACGAGCCACUUGAAACUGAA	50	70	5prim	21
MI0005398	UCAUGGUCAGAUCCGUCAUCC	128	148	3prim	21
MI0005399	UUUAGGUCGAGCUUCAUUGGA	50	70	5prim	21
MI0005400	UGGUAAGAUUGC UUAUAAGCU	55	75	5prim	21
MI0005400_114_134	UUUAUAAGCCAUCUACUAGUU	114	134	3prim	21
MI0005401	CGGCUCUGAUACCAAUUGAUG	92	112	3prim	21
MI0005402	UUGAAUUGAAGUGCUUGAAUU	288	308	3prim	21
MI0005404	UAACUAAACAUUGGUGUAGUA	118	138	3prim	21
MI0005406_119_139	UGGGUGGCAAACAAAGACGAC	119	139	3prim	21
MI0005407	AAGAUAAAGCGCCUAGUUCUG	137	157	3prim	21
MI0005409	UAGACCAUUUGUGAGAAGGGA	35	55	5prim	21
MI0005410	UCACUCCUCUUCUUCUUGAUG	159	179	3prim	21

MI0005412	GAUGAGGAUAGGGAGGAGGAG	3	23	5prim	21
MI0005413	GAUGAGGAUAGGGAGGAGGAG	3	23	5prim	21
MI0005414	GAUGAGGAUAGGGAGGAGGAG	3	23	5prim	21
MI0005415	GAUGAGGAUAGGGAGGAGGAG	3	23	5prim	21
MI0005434	UUUUGUAUGUUGAAGGUGUAU	344	364	3prim	21
MI0005435	UUUCGUUGUCUGUUCGACCUU	11	31	5prim	21
MI0005436	UCUCUCUGUUGUGAAGUCAAA	18	38	5prim	21
MI0005437	UCAAUAGAUUGGACUAUGUAU	86	106	3prim	21
MI0005438	GAUGGAUAUGUCUUCAAGGAC	93	113	3prim	21
MI0005438_24_44	CCUUGGAGAAAUAUGCGUCAAA	24	44	5prim	21
MI0005439	AUAUGCUGGAUCUACUUGAAG	75	95	3prim	21
MI0005439_11_31	UCCAAUAGGUCGAGCAUGUGC	11	31	5prim	21
MI0005440	UUGAGAGCAACAAGACAUAAU	239	259	3prim	21
MI0005440_48_68	UUAUGUCUUGUUGAUCUCAAU	48	68	5prim	21
MI0005441	UCAGGUAUGAUUGACUUCAAA	4	24	5prim	21
MI0005442	AUGAAUUUGGAUCUAAUUGAG	9	29	5prim	21
MI0005442_108_128	UUUUUCCUCAAAUUUAUCCAA	108	128	3prim	21
MI0005443	UCAAGGAACGGAUUUUGUUA	18	38	5prim	21
MI0005445	UUGAACAUGGUUUUUAGGAA	24	44	5prim	21
MI0005446	CUUCUUAAGUGCUGAUAAUGC	166	186	3prim	21
MI0005447	AUUGGUUCAAUUCUGGUGUUG	229	249	3prim	21
MI0005447_240_260	UCUGGUGUUGAGAUAGUUGAC	240	260	3prim	21
MI0005448	UAAUUUGGUGUUUCUUCGAUC	65	85	3prim	21
MI0008303	UGAGAGAAGUGAGAUGAAAUC	4	24	5prim	21
MI0008303_13_33	UGAGAUGAAAUCUUGAUUGG	13	33	5prim	21
MI0008303_84_104	AAUUAAGAUUUCAUCUUCU	84	104	3prim	21
MI0008304	UACUAAGUAGAGUCUAAGAGA	39	59	5prim	21

MI0008305	UAAGUUAAGAUUUGUGAAGAA	10	30	5prim	21
-----------	-----------------------	----	----	-------	----

## Załącznik nr 2. Sekwencje genów ulegających ekspresji w *A. thaliana* - fragment

>gi|2121272|gb|AF000146.1|AF000146 PLN Arabidopsis thaliana dihydroorotase (pyr4) mRNA, complete cds.

G TTCGTTTTGTTGTGTTGTAAGCATCCCTGTTGTGCATTTGTTTTATCCAGGTGAATTCCTGCAAAGCCTTGAGCAGCTTTCAGATGATAAAGACTTTGGTCTCTCCTTATAGTGGATTTGGATCTCAGAAGTTGAAGTTTGACAGATCTTCTGAGAAAGTGAAGACGAGAGCAGTAAGAATGGAGCTCACAACTCACTCAACCTGATGATTGGCATCTTCATCTCCGTGACGGTGATCTTCTTCATGCTGTTGTTCCACAGTGCAGTAATTTAAGAGAGCGATTGTGATGCCGAATCTGAAGCCTCTGTACCTCCACTGCAGCTGCCATTATTACCGGAAATTTATCATGAAAGCTTGGCATCTGAGAGTAGCTTGTATCCACTTATGACACTTATTTGACTGACAAAACCTACCTGAAGAGATCAGGCTTGCAAGGAAAGTGGTGTGTTTATGCGGTGAAGCTGTACCCTGCCGAGCAACAACCAACTCTCAAGATGGTGTACGGATCTCTTTGGAAAATGCTTACCGGTACTAGAAGAGATGGTCAAACAAAACATGCCTTGTGTTTGTGGGAGGTCACAGATCCGAGTATCGATGTCTTTGACCGAGAAAATCTTCATTGAGACAGTCTGCAGCCTTAATCCAACGCCTTCCGACGTGAAAGTAGTGATGGAACACATCACACCATGGATGCTGTGAATTTTGTGAATCTTCAAAGAAGGGTCTGTGGGTGCAACAGTCAACCACAACATCTCCTTCTCAACAGGAACGCCCTTTTCAAGGTGGATTACAACCTCACTACTGTCTTCCCGTCTCAAAGAGAAATACATAGAGAAGCCATTGTCAAAGCTGTAACACTAGTGAAGCAAGAAGTTCTTCTCGGCACAGATAGTCTCCACATGAACGGAGTAGAAAAGAATCATCCTGCGGATGTGCTGGTATTTACAGCGCTCCATTGCCTTGTCCCTATACGCCAAGGTCTTTGATGAGGCGGGTGCCTGGACAAGTTGGAAGCTTTCACCAAGTTTCAATGGACCTGATTTCTATGGCCTCCGAGAACTCGTCAAAGATCACACTGAAGAAATCTCCTTGAAGGTTCCAGACGTTTTCACTTCCATTCCGGAGAGATCGTCCCTATGTTTGTGGAGAAACCTTCAATGGCAACCGTTGAAATAATGTGGTGAATTTGAAGTTGTCAAATTTTTTCCCAAATCATGAACTCGTACCTTTTGGATTGAATTTTACGGCTGAGTTATGCCCTGCATTTTTTGAGGAATAAAGCAAGATACTGACCAATAAAAAAAAAAAAA

>gi|2121274|gb|AF000147.1|AF000147 PLN Arabidopsis thaliana UMP/CMP kinase (pyr6) mRNA, complete cds.

CCAAATCTCTCCGTGTTTCTTCTGCTTCAGGTGATATTTAAACAAAGTCTTAGTACAATGGGATCTGTTGATGCTGCTAATGGAAAGTGGGAAGAAACCTACAGTTATATTTGTTCTTGGTGGTCCAGGAAGTGGAAAAGGTACCCAGTGTGCTTATATTGTTGAACATTATGGTTACACACATCTGAGTGTGAGATCTTCTTAGAGCTGAGATTAATCAGGTTCTGAAAATGGAACATGATCCAGAATATGATTAAGAAGGGGAAGATTGACCTTCTGAGGTTACTATCAAGCTTCTACAGAAAGCTATCAGGAAAACGGGAATGACAAGTTCCTCATTGATGGTTCCCTCGAATGAGGAAAACCGAGCAGCATTGAAAAGTTACTGAGATTGAACCAAAGTTTGTCTTATCTCGATTGCTCAGAGAAGATGGAGAAAGCGCTGTTGGGCGAAACAGGGGAGAGAGGATGACAATATTGAGACTATAAGGAAGCGCTTAAAGGTGTTTCTTGAACTAGCTTACCAGTGATTCATTACTACGAAGCTAAGGGGAAAGTTAGGAAGATTAATGCTGCAAAGCCATTGAAGCTGTCTTCAGGAGGTGAAGGCAATTTTTTCTCCTGAAGCTGAGAAGGTTGAAGCCTAGGCTGCAGCATCGTAAAAAGATACCGAAAACCGGATAAATTGATCAGGTAGAGTTGGGTGTGTTGCATCTTCTTACAGGAAGTCTCTGCTGGCCGTATTTGTTTTTGTCTTATTAACCATATGATTGATGATTACTGTTATTATACACAATTTACCTATTATATAATTTTTGTTTTGTTTCAATTTTTGTATTTGATGGGATTAATAAAAAAAAAAAAAAAAAAAAA

>gi|2145038|gb|AF000228.1|AF000228 PLN Arabidopsis thaliana p-hydroxyphenylpyruvate dioxygenase mRNA, complete cds.

CCACGCTCCGAGTTTTAGCAGAGTTGGTGAATCAATGGGCCACAAAACGCCCGTTTCAGAGAATCAAACCATGATGACGGCGTGCCTGCTCGCCGGATTCAAGCTCGTCGGATTTTCAAAGTTCGTAAGAAAGAAATCAAAGTCTGATAAATTCAAGGTTAAGCGCTTCCATCACATCGAGTTCTGGTGCGGCAGCAACCAACGTCGCTCGCTTCTCCTGGGGTCTGGGGATGAGATTCTCCGCCAATCCGATCTTCCACCGGAAACATGGTTACGCCTTACTACTCACCTCCGGTACCTCCGATTCTTTTACTGCTCCTTACTCTCCGTCTCTCCGCCGGAGAGATTAACCGACAACCACAGCTTATCCCAAGTTTGCATCACGGCTTGTGCTTCTTCTTCTTCCATGGTCTCGGTGTTAGAGCCGTTGCGATTGAAGTAGAAGACGCAGAGTCAGCTTCTCCATCAGTGTAGCTAATGGCGCTATTCCTTCGTCGCTCCTATCGTCTCAATGAAGCAGTTACGATCGCTGAGGTTAACTATACGGCGATGTTGTTCTCCGATATGTTAGTTACAAAGCAGAAGATACCGAAAATCCGAATTTGCCAGGGTTCGAGCGTGTAGAGGATGCGTCTGTTCCATTGGATTATGGTATCCGGCGGCTTGACCACGCCGTGGGAAACGTTCTGAGCTTGGTCCGGCTTAACTTATGTAGCGGGTCTACTGGTTTTCAACATTCGAGAGTTCACAGCAGACGACGTGGAACCGCCGAGAGCGGTTAAATTCAGCGGTCTGGCTAGCAATGATGAAATGGTTCTTACCATTAAACGAGCCAGTGCACGGAAACAAGAGGAAGAGTCAGATTGAGACGATTTGGAACATAACGAAGGCGCAGGGCTACAACATCTGGCTCTGATGAGTGAAGACATTTAGGACCCTGAGAGATGAGGAAGAGGAGCAGTATTGGAGGATTCGACTTTCATGCCTTCTCCTCCGCTACTACTACCAGAACTCAAAGAACGGTCCGGCAGCTGCTCAGCGATGATCAGATCAAGGAGTGTGAGGAATAGGGATTCTGTAGACAGAGATGATCAAGGGACGTTGCTTCAAATCTTCAAAAACCACTAGGTGACAGGCCGACGATATTTATAGAGATAATCCAGAGAGTAGGATGCATGATGAAAGATGAGGAAGGGAAAGGCTTACCAGAGTGGAGGATGTGGTGGTTTTGGCAAAGGCAATTTCTGAGCTCTTCAAGTCCATTGAAGAATACGAAAGACTCTTGAAGCCAAACAGTTAGTGGGATGAACAAGAAGAACAACCAACTAAAGGATTGTGTAATTAATGAAAACGTTTTATCTTCAAAAACATGTTATACAACATCTCATTTAAAAACGAGATCAATCAAAAATACAATCTTAAATTCAAAACCAAAAAAAAAAAAA

>gi|2196703|gb|AF000977.1|AF000977 PLN Arabidopsis thaliana MEK1 mRNA, complete cds.

CGACGGTTCGTCATTCTTCTCTAAGCTTTATCTTTCTGATCTCTGACCTTGTGTCTCTGTGCATGTGTGTTTATCTGCTACTGTGTT  
 TTGTCGGTTTGATCTCCAAAAACCAGACATGAACAGAGGAAGCTTATGCCCTAATCCCATCTGTCTCCCTCTTGGCAATCCATCTCCA  
 AATTCTTAACACAGAGTGAACGTTTAAAGATGGAGATCTTCGAGTGAACAAAGATGGAATCCAGACCGTGTCTCTGTCCGAACCAGGA  
 GCTCCACCTCTATTGAGCCATTGGACAACCAATTGAGTTTGGCAGATTTAGAAGTGATCAAAGTCATTGGCAAAGGAAGTAGTGGTAAT  
 GTCCAGTTGGTCAAACACAACTCACTCAACAGTTTTTCGCTCTTAAAGTTCATTCAATTGAACACAGAAGAATCAACATGTCGGGCGATT  
 CTCAGGAGCTGAGAATAAACTTGAGCTCGCAATGTCCATATCTTGTCTCATGTTATCAATCTTCTACCACAACGGTCTTGTTCATCATA  
 TTGGAATTCATGGATGGTGGATCCCTTGCAGACTTGTAAAGAAAGTCGGAAAAGTTCCTGAAAACATGCTATCTGCCATCTGCAAGCGA  
 GTTCTTCGAGGTCTTTGTTATATTATCATCATGAGAGGCGAATCATTTCATCGGGACTTAAAGCCTTCAAACCTGCTAATCAATCATAGAGGTG  
 AAGTCAAGATCACAGACTTTGGTGTGAGCAAGATCTTGACAAGCACAAGTAGTCTTCTAATCTTTCGTGGGCACATACCCTTATATGTC  
 TCCAGAGAGAATCAGCGGGAGTTTGTACAGTAACAAGAGCGATATTTGGAGCTTGGGACTGGTTTTGCTCGAATGTGCAACGGGTAAT  
 TCCCGTATACTCTCCAGAACAAGAAAGGATGGAGTAGCGTGTACGAGCTTGTGGACGCCATTGTTGAAAACCCGCTCCTTGTGCAC  
 CTCCAATCTCTTTCAGAGTTTTGCTCCTCATCTCGCAATGTGTACAAAAGATCCAAGGGACAGAAAATCAGCAAAGGAGCTTCT  
 GGAACACAAGTTCGTAAGATGTTTGAAGATTCGGATACAAATCTCTCGGCTTACTTCACCGACGCAGGATCTTGTATCCCCACTTGCT  
 AACTAGAACCAGTTTGAACAATCTTTAACCTCCAAGTTATTATTCATATCCACTAGAAGAGATTTTATGAAAGATAAATTTACGAGAT  
 GTTATGACTTAAGAGACTTGTCTACAAAATTTGATTTAATGTGTGACTTTCCACATCACTATTATAAACACAACATATCATC

>gi|4100568|gb|AF001394.1|AF001394 PLN Arabidopsis thaliana fatty acid desaturase/cytochrome b5 fusion protein mRNA, complete cds.

CTTCAAAAATCCGATTCCAGATCCATTCTGGGCAAACAAGGTTGGTGTCTCTAATCTCAAAGCTTTTTTCAAATTCGGAAAA  
 AGCAAATCGTGGGAGAGATTCATCTTCTCTGTGCGTTATCGGATCTCGGAGCTTTTGGTTGCTCGTCAATGGCGGAAGAGACGGA  
 GAAAAAGTACATTACGGACGAAGATCTTAAAAACACAACAATCTGGAGATCTATGGATCGCGATTCAAGGCAAGGTCTACAACGCTCT  
 CCGATTGGATTAACACTATCCCGGAGGCGACACGGTATTCTCAATCTCGTTGGTCAAGACGTCACCGATGCTTTCATCGCATTTATCC  
 CGGAACCGCTTGGCACCATCTCGACCATCTTCCACCGTTACCACATCAGAGATTTCCAAGTCTCCGAAGTCTCACGCGATTACCGTCGT  
 ATGGCTGCCGAGTTTCGTAACCTCGGTCTCTCGAAAACAAGGTACGTTACTCTCTACACTTAGCCTTCGTGCGCCCATGTTCTCG  
 GAGTTCTACGGTGTGTTGGCTGTACCT

### Załącznik nr 3. Sekwencje kodujące genów AGB1 i IMM1

>NM\_202954.3 Arabidopsis thaliana guanine nucleotide-binding protein subunit beta (AGB1), mRNA, complete cds

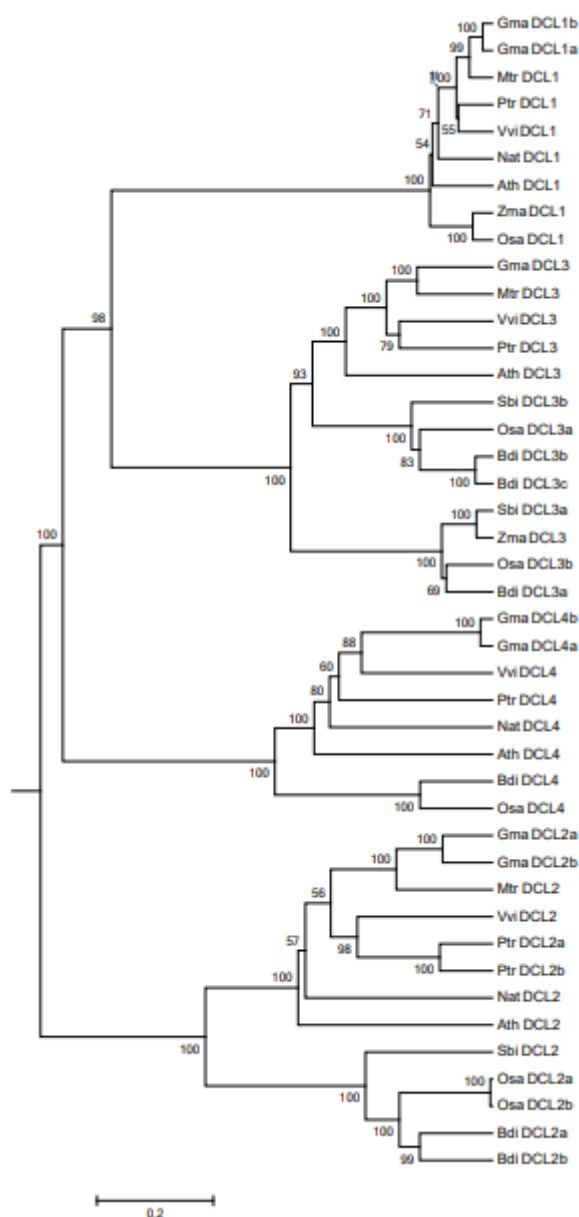
TAGTCTCAGGTCAGAGAAGCCATCATCAACATTCACAAGAGAGCCGTGTTTGTGTCTTGACTGATTCTT  
 CTCTCAAGCTTTTTTAATCTCTCTCTCTTTTCCACGTAATTCCCCAAATCCATTCTTCTAGGGTTCC  
 ATCTCCCTCTCTCAATCATGAACCTTCTCTCTCTTAGACCCACAAAGTTTCCCCCTTTTATTTGATCG  
 GCGACGGAGAAGCCTAAGTCTGATCCCGAATGTCTGTCTCCGAGCTCAAAGAACGCCACGCCGTCGCTA  
 CGGAGACCGTTAATAACCTCCGTGACCAGCTTAGACAGAGACGCCTCCAGCTCCTCGATACCGATGTGGC  
 GAGGTATTACGCGGCGCAAGGACGTACTCGGGTGTGCTTCGGAGCAACGGATCTGGTTTGTGTGCTACT  
 CTTACGGGACACACCGGAAAGGTTTATTCATTAGATTGGACACCGGAGGGAACCGGATTGTCAGTGCAT  
 CTCAAGATGGGAGATTAATCGTGTGGAATGCTCTAACGAGTCAGAAAACATGCTATTAACCTCCCTTG  
 TGCATGGGTTATGACATGTGCTTTCTCTCCAAATGGTCAAGTTCGGTTGCGTGTGGTGGATTAGACAGTGTA  
 TGTCTATCTTTAGCCTTAGCTCAACGGCGGACAAGGATGGAAGTGTACCGGTTTCAAGAATGCTCACTG  
 GTCACAGGGGATATGTTTCGTGCTGTGAGTATGTCCAAATGAGGATGCCACCTTATCACCAGTTCAGG  
 TGATCAAACCTGTATCTTATGGGATGTAACACTACTGGTCTCAAACCTTCTGTTTTTGGCGGTGAATTTACG  
 TCTGGACATACTGCTGATGTACTAAGCGTCTCAATCAGTGGATCAAACCCAAACTGGTTTATATCTGGTT  
 CATGCGATTCCACAGCACGGTTGTGGGACACTCGTGTGCAAGCCGAGCAGTGCGTACCTTTCATGGTCA  
 CGAGGGAGATGTTAATACGGTCAAGTTCTTTCCGGATGGGTATAGATTTGGGACTGGATCAGACGATGGA

ACATGCAGGCTGTATGACATAAGGACTGGTCACCAACTCCAGGTCTATCAGCCACATGGTGATGGTGAGA  
ACGGACCTGTCACCTCCATTGCATTCTCTGTGTCAGGGAGACTTCTTTTCGCTGGCTATGCGAGCAACAA  
CACTTGCTACGTTTGGGATACCCCTCTTGGGAGAGTTTTCTTGACAGTTGATTGGATTTGGGATTACAGC  
AGGATTCACACAGGAATAGAATAAGCTGTTTGGGGTTGTCAGCAGATGGAAGTGCCTTGTGTACAGGAAG  
TTGGGATTCAAATCTAAAGATATGGGCGTTTGGAGGACACAGGAGAGTGATTTGAAGAAGATTTAACGAA  
AGTAGGAGTCACGTCTCCAGTTGTTGGTTAATATATTTCTGTAGTCGGGAAGTAAGGTTTCGGTTTGTGGAA  
GGTGTGTTGGTTTGAATAGTGGAGTGGTTAGAAGAATTAAACTTCCCTTTTGTAGTGTGCTTTGATTTA  
TTTATTTCTTCATTGGGAACTAAACTCCTTCAACACGCTACTCAATGTGAATTCTGTAATCAATTGTGTA  
CCCACCAGTCTTTACTTTACTATCATCTCTTC

>NM\_118352.4 Arabidopsis thaliana Alternative oxidase family protein (IM), mRNA,  
complete cds

GCGUCGACGUCGAUGGGGAUGAGAAGUAGUGGAGAAUCCGCUCACAUUGGGAUUCGUCAUUCUUCUUCUA  
AAACCCGCAAAAUUCUCCAUUUCUACCAAAAAUAUCCAACUUUUACUUUUCUUUCCUGUGAAAUAUCU  
GCUCAAAUCUUUGGUUCCUGACGGAGAUGGCGGCGAUUUCAGGCAUCUCCUCUGGUACGUUGACGAUUUC  
ACGGCCUUUGGUUACUCUUCGACGCUCUAGAGCCGCCGUUUCGUACAGCUCUCCUCACCGAUUGCUUCAU  
CAUCUUCUUCUCUUCUUCGUCGUCUGCUAUUAAGGAACAUAUCGAGUCCAAGCAACGAUUUUGCAAG  
ACGAUGAAGAGAAAGUGGUGGAGGAAUCGUUUAAAGCCGAGACUUCUACUGGUACAGAACCACUUGA  
GGAGCCAAAUAUGAGUUCUUCUUAACUAGUGCUUUUGAGACAUGGAUCAUAAGCUUGAGCAAGGAGUG  
AAUGUUUUCCUUACAGACUCGGUUUAUAAGAUACUUGACACUUUGUAUCGUGACCGAACAUUGCAAGGU  
UCUUUGUUUCUUGAGACAAUUGCUAGAGUGCCUUUUUUGCGUUUAUGUCUGUGCUACAUAUGUAUGAGAC  
CUUUGGUUGGUGGAGGAGAGCAGAUUAUUUGAAAGUACACUUUGCUGAGAGCUGGAAUGAAAUGCAUCAC  
UUGCUCAUAAUGGAAGAAUUGGGUGGAAAUCUUGGUGGUUGAUCGUUUUCUGGCUCAGCACAUAGCAA  
CCUUCUACUACUUCUAGACAGUGUUCUUGUAUAUCUUAAGCCUAGAAUGGCAUAUCACUUUUCGGAUG  
UGUGGAGAGUCAUGCAUAUGAGACUUAUGAUAAAUUUCUCAAGGCCAGUGGAGAGGAGUUGAAGAAUAUG  
CCUGCACCAGUAUCGAGUAAAUAUCUUAUCGGGAGGUGACUUGUACUUAUUUGAUGAGUUCCAAACAU  
CAAGAACUCCAAUACUCGAAGACCAGUAAUAGAAAUCUUAUCGAUGUUGUUGAACAUAAGAGAUGA  
UGAAGCAGAACACUGCAAGACAAUGAGAGCUUGUCAGACUCUAGGCAGUCUGCGUUCUCCACACUCCA  
UUAGAAGAUGAUGAUACUGAAGAAGAAUCAGGGUGUGUUGUUCUGAGGAGGCUCAUUGCGAAGGUAUUG  
UAGACUGCCUCAAGAAAUCCAUAACAAGUUAUAAAUAUAGAAAGUAAACUAAAAAGAUUAUUUGUAUCA  
GCUCAUGAACAAUAGAUUAUUAUCCAUUAUCUUGGGAUUAAGGAUUAUUGGAAAUUCCAUCGUUGUG  
CUAGUGUGAGAGAAUCAAAUACCCUAAUGAUGUAAAUGUACUUGAUGAGCUUAAGUCGUUGUAGACC  
AUUUUAUCCAUAUCUACUAGCCUUUGGUUGCAUU

**Załącznik nr 4.** Drzewo filogenetyczne 43 białek DCL w zakresie modelowanego fragmentu.



The evolutionary history was inferred using the Minimum Evolution method [1]. The optimal tree with the sum of branch length = 7.06428085 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [2] and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm [3] at a search level of 1. The Neighbor-joining algorithm [4] was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 499 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [5].

1. Rzhetsky A & Nei M (1992) A simple method for estimating and testing minimum evolution trees. *Molecular Biology and Evolution* 9:945-967.

2. Zuckerkandl E & Pauling L (1965) Evolutionary divergence and convergence in proteins, pp. 97-166 in *Evolving Genes and Proteins*, edited by V. Bryson and H.J. Vogel. Academic Press, New York.

3. Nei M & Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

4. Saitou N & Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

5. Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.

## Załącznik nr 5. MSA 43 sekwencji roślinnych białek DCL (zakres modelowanego fragmentu)

Osa DCL1	ILSGRDGQNSQFIKL	MYSVNCVNVGTSKDPFVVTQLSNFA	IFGNE	DAEYL-STWML	FVARTMIKASVFRGRI	ETES	VLLS	HVRLMSI	LDVDDVDPST	107				
Zma DCL1	ILSGRDGQSSQFIKL	LYSVNCVNIKTSKDPFVVTQLSNFA	IFGNE	DAEYL-STWML	FVARTIIPKASVFRGPI	ETES	ILLKS	HVRLMSI	LDVDDVDPST	107				
Gma DCL1a	ILSEKDACNNYKLLHL	MYAVKCNELGHSKDPFLTQVSNFA	LFNGE	DAEYL-SMSML	FVARTVTKASVFSGLN	ETES	ASL	HVRLMSI	LDVDDVDPST	107				
Gma DCL1b	ILSGKDACNNYKLLHL	MYAVKCNELGHSKDPFLTQVSNFA	LFNGE	DAEYL-SMSML	FVARTVTKSSVFRGLN	ETES	ASL	HVRLMSI	LDVDDVDPST	107				
Mtr DCL1	ILSGKDACNNYKLLHL	MYTIKCNVGHGSKDPFLTQVSDFA	LFGTE	DAEYL-SMSML	FVARTVTKTSVFRSD	ETES	TSL	HVRLMSI	LDVDDVDPST	107				
Vvi DCL1	ILLGKDCNSSLVHL	MYAVKCNVFGSKDPFLTQVSDVFL	LFNGE	DAEYL-SMSML	FVARTVTKASVFWGPI	ETES	ASL	HVRLMSI	LDVDDVDPST	107				
Ptr DCL1	ILCGRDGCNNSKVLHL	LYGVRCLNIGTNDPFLTQVSNFA	LFNGE	DAEYL-SMSML	FVARTVTKASVFRGRI	ETES	ASL	HVRLMSI	LDVDDVDPST	107				
Nat DCL1	ILSGKDCSDSSKLVHL	MYAIKCNIGTSKDPFLTQVSEFA	LFNGE	DAEYL-SMSML	FVARTVTKATVFRGPI	ETES	ASL	HVRMSI	LDVDDVDPST	107				
Ach DCL1	VSSGKEVCESSKLFHL	MYNVRVDFGSSKDPFLSEVSEFA	LFNGE	DAEYL-SMSML	VARAMITKASVAFKSD	TEN	SSL	HVRLMSI	LDVDDVDPST	107				
Osa DCL2a	RSFS-----RLGIYYC	KISLEGC-----PKTASPTDII	ALKCD	GSDFI-SSSFKL	PGQDN-ASVTKVYGI	HQNE	VIAAR	OTTILSFL	IGDDHLEYSNG	96				
Osa DCL2b	RSFS-----RLGIYYC	KISLEGC-----PKTASPTDII	ALKCD	GSDFI-SSSFKL	PGQDN-ASVTKVYGI	HQNE	VIAAR	OTTILSFL	IGDDHLEYSNG	96				
Bd1 DCL2a	RSFS-----HLGIYHC	KISLGGC-----LKTIASPTDII	AVKCD	GSDFI-ANSFKL	SGAQDF-LSVSKYVGN	HQNE	VIAAR	OTTILSFL	IGDDHLEYSNG	97				
Bd1 DCL2b	LSFS-----CLGIYIC	MSLEGF-----LQITAPTDDII	AMKCD	GSDFV-SRSFKL	YGTQDY-TVVAKYGI	HQNE	VIAAR	OTSISL	LINNAIK	91				
Sb1 DCL2	LSFC-----HRGLIYC	KISLEGC-----LQITAPTDDII	AVKCD	GSDFI-SSSFKL	MGYQDY-LSFTKYVGN	HQNE	VIAAR	OTTILSFL	IGDDHLEYSNG	96				
Gma DCL2a	VSNV-----SHTIYHC	WELSQN-----FSYDISVQDIF	AIRIE	DPEIG-CTQDF	GFDRGS-WSLKYRGT	N	SPNII	LLC	QVTLRGL	IHSMMKLTAS	97			
Gma DCL2b	VSNV-----SQTWYHC	FMEFNK-----FSYDVSQDIF	LMRIE	DPEIG-CMDF	GFDRGS-LSVNFYRGT	N	SPD	LLC	QVTLRGL	IHSMMKLTAS	97			
Mtr DCL2	MSNN-----SSTIYIC	LIELKQN-----FSYDITVQDIF	ATRYE	DPEIG-CMDF	CFDRGS-LAVNRYGN	N	SPD	LLC	QVSIIGL	MSKMKDKESFS	97			
Vvi DCL2	SLKD-----PVPYHC	LIELNQK-----YAYEYSPQIV	AVRSE	EYDVG-NVNF	QWRGT-MTVSNVGV	H	TAEV	LMCR	QITLRVL	LDRAIDKGV	95			
Ptr DCL2a	GPRN-----LTKYYC	LIELNQK-----FAYDIPVHDV	VVRTE	ESDVIRSMGF	EAERGL-TVMRYGI	D	ERVII	LLC	QITLFAV	LHSHVNLKEV	98			
Ptr DCL2b	GPRN-----LTKYYC	LIELNQK-----FAYDIPVHDV	VVRTE	ESDVIRSMGF	EAERGL-LAVSRYGI	D	QVPL	LLC	QITLFEV	LHREVNKLEEV	98			
Nat DCL2	CAND-----TEAVYIC	VDLQHD-----SYSSQLCGII	AVRTR	KFDDERLTF	DMDKGS-LIVQNVSV	R	TSE	LRCQ	QVSLFRI	LDLDSKLEDA	97			
Ach DCL2	FSAQ-----ETIYHF	LIRMKPN-----SPRNLHNDV	LTRVV	EDDIG-TNFR	EDHRGT-LAVTISVGF	A	TRV	LFRC	QITLFRV	LHSHVENLMEA	97			
Osa DCL3a	IHDS-----TALNT	KDFLCD-----QEGENYAGEV	LMERF	DDVA-PSKWL	FILPNKMYTT	TPRKL	Q	NKK	Q	KKG	Q	EFFNGI	FGLFHGRKSKGQDF	103
Osa DCL3b	VSKK-----TALKQS	KMNFVCD-----QAGIYSEFV	LIDAT	PDEVA-TLEIDL	YLDH-KMVKTS	SSGLE	DAQ	VEQA	Q	GLFNGI	FGLFTRSKVNPAPREF	102		
Bd1 DCL3a	THES-----ENVSLNA	RDFRCD-----QEGENYAGEV	LMESA	DDVA-CSKWL	FILPNKMYTT	TPRKL	Q	NKK	Q	REG	Q	EFFNGI	FGLFHGRKSKGQDF	103
Bd1 DCL3b	ASKR-----TVIKLQG	KINFSCD-----QVGQYSEFV	LIDTI	ANAAA-NIDINL	YLDH-KMVKAS	SPCGLE	DVQ	VEQA	Q	GLFNGI	FGLFTRSKVNPAPREF	102		
Bd1 DCL3c	ASKR-----TVIKLQG	KINFSCD-----QVGQYSDVFL	LIDTI	ANAAA-NIDINL	YLDH-KMVKAS	SPCGLE	DVQ	VEQA	Q	GLFNGI	FGLFTRSKVNPAPREF	102		
Zma DCL3	IHES-----ETVILNA	TFDFRCD-----QEGELYAGEV	LMESV	DDVA-RSEIDL	FILPNKMYTT	TPRKL	Q	NKE	Q	CKG	Q	EFFNGI	FGLFHGRKSKGQDF	103
Sb1 DCL3a	IHES-----ETVILNS	TFDFRCD-----QEGFTFAGEV	LMESV	DDVA-HSEIDL	FILPNKMYTT	TPRKL	Q	NKE	Q	CKG	Q	EFFNGI	FGLFHGRKSKGQDF	103
Sb1 DCL3b	GDKL-----SVTKLQG	KINFVCD-----QVQKCYDFV	LIDTS	IAKAAA-TINIDL	YLDH-KMVKAS	SPCGLE	DAR	VEKT	Q	GLFNGI	FGLFTRSKVNPAPREF	102		
Gma DCL3	GDKL-----TRVKFNA	KFEFTCN-----IVSIYSGFA	LIESE	DDVA-NVNDL	YLVY-KIWKAS	SSCQ	D	DAE	EMRA	Q	GLFNGI	FGLVVKKSAG-EREF	101	
Mtr DCL3	GDKP-----DGAKFHA	KFEFKCN-----IVSIYSGFV	LIVESL	DDVG-NIELDL	YLYS-KMVKTS	SSCQ	D	DAE	EMRA	Q	GLFNGI	FGLVVKKSAG-EREF	101	
Vvi DCL3	GDKI-----DGAIFQA	KDFSCN-----IVNELYSGFV	LIESK	DDVG-NIEMEL	YLYS-KFVPTS	SSCQ	S	DAE	EMKAMF	Q	GLFNGI	FGLVVKKSAG-EREF	102	
Ptr DCL3	GEKL-----DGATFHA	KDFSCS-----IVSGIYSGFI	LIESK	DDVG-NIELDL	YLYA-KIWKSSI	SSCQ	H	DAE	EMTKA	Q	GLFNGI	FGLVVKKSAG-EREF	102	
Ach DCL3	GENI-----DGATFQA	KDFDCCN-----ISGVYSSFI	LIEST	AEVVG-KVMDL	YLYR-KLWKAS	SPCQ	Q	SOE	EMVKA	Q	GLFNGI	FGLVVKKSAG-EREF	99	
Osa DCL4	LKLD-----SLNLHF	VYKFIPI-----PEDRYQMF	GFVNF	PVEAE-TLQVLD	YHARGRI	VKAGI	KHLG	KAFA	EMLAH	Q	GLFNGI	FGLVVKKSAG-EREF	102	
Bd1 DCL4	CKQD-----CLLDLHF	VYKFFPI-----PADRHYQMF	GFVNF	PEBAE-KLEVD	YHARGRI	VKTG	KHLG	TV	FDKE	EMVLA	Q	GLFNGI	FGLVVKKSAG-EREF	97
Gma DCL4a	INED-----NIVRLNS	VYKFCPY-----PEDRVYKEF	GFVNF	PEBAE-KLELD	YHAGRS	YMTF	VFFV	GF	NKD	KMAEN	Q	GLFNGI	FGLVVKKSAG-EREF	97
Gma DCL4b	INED-----NIVRLNS	VYKFCPY-----PEDRVYKEF	GFVNF	PEBAE-KLELD	YHAGRS	YMTK	FVFF	GF	NKD	KMAEN	Q	GLFNGI	FGLVVKKSAG-EREF	97
Vvi DCL4	SNL-----EHICLNS	VYKFTPI-----PEDRIYKQF	GFVNF	PEBAE-KLELD	YHAGRS	YMTK	FVFF	GF	NKD	KMAEN	Q	GLFNGI	FGLVVKKSAG-EREF	96
Ptr DCL4	TELE-----KPIHLNS	VYFCPV-----PEDRIYKQF	GFVNF	PEAD-KMSEI	YHAGRS	YMTK	VFS	GLS	KFTD	THATN	Q	GLFNGI	FGLVVKKSAG-EREF	97
Nat DCL4	TETE-----SPVCLNS	VYINFSPC-----PIDRVYKQF	GFVNF	PEBAE-RMKLD	YHAGRS	YVET	EL	IPS	GATFENN	VOLAEX	Q	GLFNGI	FGLVVKKSAG-EREF	97
Ach DCL4	DPST-----SCVNLHS	VYIMFVPH-----PADRIYKQF	GFVNF	PEBAE-TMDLD	YHAGRS	YSVK	FPS	VTE	FDDND	RLAEX	Q	GLFNGI	FGLVVKKSAG-EREF	97

Osa_DCL1	----	TPWDPAKAY	F	FPVGAERKCTD-PURE	I	MTL	NNIVNTDAMNPLQARPDVYLG	TNERTLGGDRREYGF	GKLRHGTAFGQKAHPTYGIRGAI	AIEFDIVKASGLVPARD	: 215	
Zma_DCL1	----	TPWDPAKAY	F	FPVGAERCMD-LLRE	I	MTL	NSIVNSDAMNPLQARPDVYLG	TNERTLGGDRREYGF	GKLRHGTAFGQKAHPTYGIRGAI	AIEFDIVKASGLVPRG	: 215	
Gma_DCL1a	----	TPWDPAKAY	F	FPVMDGKSDV-PMNQ	I	MTL	ETIIGADAWKNPLQARPDVYLG	TNERTLGGDRREYGF	GKLRHGMAFGQKSHPTYGIRGAV	AQFDVVVKASGLVPRD	: 215	
Mtr_DCL1b	----	TPWDPAKAY	F	FPVMDGKSDV-PMNQ	I	MTL	ETIIGADAWKNPLQARPDVYLG	TNERTLGGDRREYGF	GKLRHGMAFGQKSHPTYGIRGAV	AQFDVVVKASGLVPRD	: 215	
Vvi_DCL1	----	TPWDPAKAY	F	FPVMDGKSSD-PMNQ	I	MTL	ETIIGADAWKNPLQARPDVYLG	TNERTLGGDRREYGF	GKLRHGMAFGQKSHPTYGIRGAV	AQFDVVVKASGLVPRD	: 215	
Ptr_DCL1	----	TPWDPAKAY	F	FPVMDKSED-PIRQ	I	MTL	ERIIRTDGWSNPLQARPDVYLG	TNERTLGGDRREYGF	GKLRHGMAFGQKSHPTYGIRGAV	AQDVVVRASGLVPRD	: 215	
Nat_DCL1	----	TPWDPAKAY	F	FPVMDKSDV-PIKE	I	MTL	ENIIGDAMSNRQARPDVYLG	TNERTLGGDRREYGF	GKLRHGMAFGQKSHPTYGIRGAV	AQFDVVVKASGLVPRG	: 215	
Ath_DCL1	----	TPWDPAKAY	F	FPVTDMSGD-PIKD	I	MTL	KKITKTAWDNPLOQARPDVYLG	TNERTLGGDRREYGF	AKLRHGMAFGKLSHPTYGVRGAI	ANFDVVVKASGLVPHRS	: 215	
Osa_DCL2a	----	IKYFHEMQVPIGVVY	LL	PLVSGR-PIKG	I	MTL	KKITKTAWDNPLOQARPDVYLG	TNERTLGGDRREYGF	GKLRHNI	VFGKSHPTYGIRGAV	ASFDVVVRASGLLVPVRD	: 143
Osa_DCL2b	----	IKYFHEMQVPIGVVY	LL	PLVSGR-PIKG	I	MTL	KKITKTAWDNPLOQARPDVYLG	TNERTLGGDRREYGF	GKLRHNI	VFGKSHPTYGIRGAV	ASFDVVVRASGLLVPVRD	: 143
Bdi_DCL2a	----	IKQFHDMEASVGVVY	LL	PLVSGK-PIKG	I	MTL	KFSSSPIYEANNKHMTH	----	----	----	: 143	
Bdi_DCL2b	----	IKQFHDMEASVGVVY	LL	PLVSGK-PIKG	I	MTL	KFSSSPIYEANNKHMTH	----	----	----	: 143	
Sb1_DCL2	----	IKSL-LEMQSSGAVY	LL	PLVSGK-PIKG	I	MTL	KFSSSPIYEANNKHMTH	----	----	----	: 143	
Gma_DCL2a	----	LD-KCYLEDDVEIDY	LL	PLAIGKEKSH-PIKG	I	MTL	KSVPESSFTCKY	----	----	----	: 142	
Gma_DCL2b	----	LD-KCYLEDDVEIDY	LL	PLAIGKEKSH-PIKG	I	MTL	KSVPESSFTCKY	----	----	----	: 142	
Vvi_DCL2	----	E-KYSLEDDLEIDY	LL	PLSIAIBETPS-PIKG	I	MTL	SSIHPSKIQCRI	----	----	----	: 141	
Ptr_DCL2a	----	FD-RYDLGNDQMDVY	LL	PLSTNSREIPS-PIKG	I	MTL	GSVFFSHENASN	----	----	----	: 141	
Ptr_DCL2b	----	LE-GLDLGSGAEIDY	LL	PLAFRSCSQPS-PIKG	I	MTL	SSVLFYKNEEHFN	----	----	----	: 145	
Nat_DCL2	----	LKG-LELGTGVMDY	LL	PLPAIRSRQPS-PIKG	I	MTL	SSVLFYKNEEHFN	----	----	----	: 145	
Ath_DCL2	----	LAAVQLPVGSAVDY	LL	PLSLGTSQNPQ-PIKG	I	MTL	NSVLFPSQVLGDKKH	----	----	----	: 147	
Osa_DCL3a	----	LN-GLHLRDLGVALDY	LL	PLSTHSHETSLS-PIKG	I	MTL	RSVNLTSHEVLEKHEH	----	----	----	: 146	
Osa_DCL3b	----	LNKREDTFVWNTASVY	LL	PLTNP-SFDSN--VC	I	MTL	ESCAGAVEQLWSSYQGDENVIP	----	----	----	: 163	
Bdi_DCL3a	----	IFRKGHEIQWNTESMY	LL	PLLRD-SSYIQDLS	I	MTL	DAATAVKLMRRIYSENKREL	----	----	----	: 159	
Bdi_DCL3b	----	IFRKGHEIQWNTESMY	LL	PLLRD-SSYIQDLS	I	MTL	DAATAVKLMRRIYSENKREL	----	----	----	: 159	
Bdi_DCL3c	----	ILNKDDNLIWNNANMY	LL	PLMDP-TLESHDSTC	I	MTL	GFCTDALEQLRNNMYTEDGNLH	----	----	----	: 162	
Zma_DCL3	----	IFTKGHEIQWSDNMY	LL	PLLRN-SSQIQHDPN	I	MTL	DEAATAVRLMRKIHSDGKMNJ	----	----	----	: 161	
Sb1_DCL3a	----	IFRKGHEIQWSDNMY	LL	PLLRN-SSQIQHDPN	I	MTL	DEAATAVRLMRKIHSDGKMNJ	----	----	----	: 161	
Sb1_DCL3b	----	ILKDDDAFLWKNENMY	LL	PLVDS-TQDSHKSVC	I	MTL	ESCSGVEQLRNLILEDGNNLH	----	----	----	: 163	
Gma_DCL3	----	LLKQDNTSLWSPKHY	LL	PLEKLNDCI	I	MTL	ESCSGVEQLRNLILEDGNNLH	----	----	----	: 163	
Mtr_DCL3	----	LLQKDELLWSPKNSY	LL	PLEKSNDCI	I	MTL	DIAATTVGLMRSTYS	SDGKHNH	----	----	: 161	
Vvi_DCL3	----	LLNTQKSLWSSNNMY	LL	PLESSNIPDES	I	MTL	NSCASAIFKLRKFS	SLVGTGDCDDNGT	----	----	: 170	
Ptr_DCL3	----	LLQKETTLLWSPNNMY	LL	PLEPWSISSNDWCK	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Ath_DCL3	----	LLQFTDSSLWHPAFM	LL	PLVETNDLAS--SAT	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Osa_DCL4	----	KLGN-DVTLEINSTFY	LL	PLIKQKCYGD--RFM	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Bdi_DCL4	----	ILCN-DDTLEISSTFY	LL	PLIKQKCCGD--IFV	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Gma_DCL4a	----	DLGMSAESHTSGTSTFY	LL	PLVVLQBYGN--AME	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Gma_DCL4b	----	DLGMSAESHTSGTSTFY	LL	PLVVLQBYGN--AME	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Vvi_DCL4	----	HLGKSDFCSSSTFY	LL	PLVILNFCEN--MT	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Ptr_DCL4	----	PLGK-DALSKSCPTFY	LL	PLVIFHYSER--RVT	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Nat_DCL4	----	SLEK-EDYVDSASKY	LL	PLVNLCG-HN--KTS	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	
Ath_DCL4	----	PLELQDSSRTSKSTFY	LL	PLCLCHDGS--VTS	I	MTL	NSCASAIEFRORF	SLVTEYSDNSK	----	----	: 197	





Osa_DCL1	YCKQLSPREHSDAREGDF	SENLDKTYVY	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	DIID	YPPANK	ILEALAAS	QTL	CYERAB	413
Zma_DCL1	SPREHSEASNGEF	SENLDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAIQL	KDIG	YPPANK	ILEALAAS	QTL	CYERAB	413
Gma_DCL1a	SPREHSEAHGEGES	EETHDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	NMN	YPLASK	ILGALAAS	QTL	CYERAB	416
Gma_DCL1b	SPREHSEAHGEGES	EETHDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	NMN	YPLASK	ILGALAAS	QTL	CYERAB	416
Mtr_DCL1	SPREHSEGHEDET	EETHDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	NMN	YPLASK	ILGALAAS	QTL	CYERAB	416
Vv1_DCL1	SPREHSDSNEGDA	EENLDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	DMGN	YPPAAK	ILEALAAS	QTL	CYERAB	412
Ptr_DCL1	SPREHSEEHGEL	EATVDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	DMGN	YPPAAK	ILEALAAS	QTL	CYERAB	416
Nat_DCL1	SPREHSDSNEGDA	EENLDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	DMGN	YPPAAK	ILEALAAS	QTL	CYERAB	414
Ath_DCL1	SPREHSDSNEGDA	EENLDKTYV	YLPBELC	LVHPLPGSVRGAQR	UFSIMR	RVESMLLAVQLK	DMGN	YPPAAK	ILEALAAS	QTL	CYERAB	414
Osa_DCL2a	HNCYSKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Osa_DCL2b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Bd1_DCL2a	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Bd1_DCL2b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Sb1_DCL2	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Gma_DCL2a	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Gma_DCL2b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Mtr_DCL2	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Vv1_DCL2	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Ptr_DCL2	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Nat_DCL2	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Ath_DCL2	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Osa_DCL3a	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Osa_DCL3b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Bd1_DCL3a	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Bd1_DCL3b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Bd1_DCL3c	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Sb1_DCL3a	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Sb1_DCL3b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Gma_DCL3	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Mtr_DCL3	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Vv1_DCL3	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Ptr_DCL3	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Ath_DCL3	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Osa_DCL4	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Bd1_DCL4	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Gma_DCL4a	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Gma_DCL4b	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Vv1_DCL4	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Ptr_DCL4	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Nat_DCL4	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332
Ath_DCL4	HNYKKEK	EPGDRYS	VELPEL	CRIMSP	SANNLHIFS	YVQV	QPTV	YPPQFD	PVLK	ILEALTKK	QDSQES	332

	480	500	520	540	560	580
Osa_DCL1	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YQYALNKTLSI QADRFA SR A A G VLPVFEESREYEP SIFDEESTG--CELQKESYDDY-ADNMQEDGEEIGDS :	500	520	540	560	580
Zma_DCL1	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MI YQYALNKNLSI QADRFA SR A A G VLPVFEETRDSPE SIFDEEFTPS-SELQKNLYDDYDGDMSQEDGEEIGDS :	500	520	540	560	580
Gma_DCL1a	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YQYALSGLQSI QADRFA SR A A G VLPVFEEDTKDGESSLFDQERSIS-KIERMDCHTNGYED-EMEDGELESDS :	500	520	540	560	580
Mtr_DCL1b	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YQYALSGLQSI QADRFA SR A A G VLPVFEEDTKDGESSLFDQERSIS-KIERMDCHTNGYED-EMEDGELESDS :	500	520	540	560	580
Vv1_DCL1	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YRYALNKGQSI QADRFA SR A A G VLPVFEEDTKDESSLFDQERSIF-KAERMDN-TDEFED-EMEDGELESDS :	500	520	540	560	580
Ptr_DCL1	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YQYALCKGQSI QADRFA SR A A G VLPVFEEDTKETESSLFDHERPFPSETAPGNDRGDGYDDMEMEDGELESDS :	500	520	540	560	580
Nat_DCL1	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YQYALNKGQSI QADRFA SR A A G VLPVFEEDTKDGSYIFDQEKSLAEDRTGMNHLDDYEN-EIEDGELESDA :	500	520	540	560	580
Ath_DCL1a	GDV LK VVSRFLFK POKHEGQ LTRM QOM S MW YQYALNKGQSI QADRFA SR A A G VLPVYDEDTNEESSMFGNEIENGTVAAKTLADEFEDDEAEAGELEDTDS :	500	520	540	560	580
Osa_DCL2a	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Osa_DCL2b	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Bd1_DCL2a	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Bd1_DCL2b	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Sb1_DCL2	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Gma_DCL2a	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Gma_DCL2b	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Mtr_DCL2	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Vv1_DCL2	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Ptr_DCL2a	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Ptr_DCL2b	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Nat_DCL2	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Ath_DCL2	GDSFLK VTRHFLSE RLQHEGLTKM KNLISMAA CQLACSSNVGY HAEEFNRDIIICL-----DYDERDNKKSIFLAPNGMYS :	500	520	540	560	580
Osa_DCL3a	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Osa_DCL3b	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Bd1_DCL3a	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Bd1_DCL3b	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Bd1_DCL3c	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Zma_DCL3	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Sb1_DCL3a	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Sb1_DCL3b	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Gma_DCL3	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Mtr_DCL3	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Vv1_DCL3	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Ptr_DCL3	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Ath_DCL3	GDSV LK VVCDLFLR PMKHEGLSDM SKAVCAAT HKHGWRS LQGVYRDAE DERRVAPGQISLR-PFPCNGIET-----AFVPHRRYIRDPSFFVVKPCD :	500	520	540	560	580
Osa_DCL4	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Bd1_DCL4	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Gma_DCL4a	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Gma_DCL4b	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Vv1_DCL4	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Ptr_DCL4	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Nat_DCL4	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580
Ath_DCL4	GDAFLK VVGRHFLR PDKHEGLTRR SDVNSH YELSTRKKLQWRDQQLQVIRTAQFALRPRCKVVCNTDVEVR-----LHQMIDHPDNR--ENCLNRC :	500	520	540	560	580



	700	720	740	760	780	800	*
Osa_DCL1	VLDL	TRHL	FTYD	TRHL	ETQ	RE	RE
DCL1	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Zma_DCL1	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Gma_DCL1a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL1b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Mtr_DCL1	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Vv1_DCL1	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Ptr_DCL1	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Nat_DCL1	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Osa_DCL2a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL2b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Bd1_DCL2a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL2b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Sb1_DCL2	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Gma_DCL2a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL2b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Mtr_DCL2	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Vv1_DCL2	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Ptr_DCL2a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL2b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Nat_DCL2	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Osa_DCL3a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL3b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Bd1_DCL3a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL3b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Bd1_DCL3c	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Zma_DCL3	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Sb1_DCL3a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL3b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Gma_DCL3	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Mtr_DCL3	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Vv1_DCL3	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Ptr_DCL3	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Ath_DCL3	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Osa_DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Bd1_DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Gma_DCL4a	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
DCL4b	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Vv1_DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Ptr_DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Nat_DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL
Ath_DCL4	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL	TRHL

	820	*	840
Osa_DCL1	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	774
Zma_DCL1	PIIDRMVTPD-TLPHVRLQE	PIIDRMVTPD-TLPHVRLQE	776
Gma_DCL1a	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	779
Gma_DCL1b	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	779
Mtr_DCL1	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	778
Vv1_DCL1	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	777
Ptr_DCL1	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	780
Nat_DCL1	PIIHMVTPRE-TLPHVRLQE	PIIHMVTPRE-TLPHVRLQE	779
Ath_DCL1	PIIQMVTPE-TLPHVRLQE	PIIQMVTPE-TLPHVRLQE	771
Osa_DCL2a	RIIEHTATPE-TIEPDVRLQE	RIIEHTATPE-TIEPDVRLQE	662
Osa_DCL2b	RIIEHTATPE-TIEPDVRLQE	RIIEHTATPE-TIEPDVRLQE	659
Bd1_DCL2a	RIIEHTATPE-TVACDVKLQE	RIIEHTATPE-TVACDVKLQE	664
Bd1_DCL2b	RIIEHTATPE-TVERDVKLQE	RIIEHTATPE-TVERDVKLQE	652
Sb1_DCL2	RIIEHTATPD-TLEDVVKLHE	RIIEHTATPD-TLEDVVKLHE	624
Gma_DCL2a	PIIEPLVTPE-TLKHPIRLNE	PIIEPLVTPE-TLKHPIRLNE	656
Gma_DCL2b	PIIEPLVTPE-TLKHPIRLNE	PIIEPLVTPE-TLKHPIRLNE	656
Mtr_DCL2	PIIEPLVTPE-TLKHPIRLNE	PIIEPLVTPE-TLKHPIRLNE	651
Vv1_DCL2	PIIEPLVTPE-TLKHPIRLNE	PIIEPLVTPE-TLKHPIRLNE	658
Ptr_DCL2a	PIIEPLVTPE-TVKEHARLGE	PIIEPLVTPE-TVKEHARLGE	676
Ptr_DCL2b	PIIEPLVTPE-TLKHPIRLNE	PIIEPLVTPE-TLKHPIRLNE	665
Nat_DCL2	PIIEPLVTPE-TVKEHARLGE	PIIEPLVTPE-TVKEHARLGE	668
Ath_DCL2	PIIEPLVTPE-TVKEHARLGE	PIIEPLVTPE-TVKEHARLGE	660
Osa_DCL3a	PIISPLITPD-KLAPYRLLE	PIISPLITPD-KLAPYRLLE	717
Osa_DCL3b	PIISPLITPD-KLAPYRLLE	PIISPLITPD-KLAPYRLLE	714
Bd1_DCL3a	PIISPLITPE-NLEPFRLLLE	PIISPLITPE-NLEPFRLLLE	673
Bd1_DCL3b	PIISPLITPE-NLEPFRLLLE	PIISPLITPE-NLEPFRLLLE	716
Bd1_DCL3c	PIISPLITPE-NLEPFRLLLE	PIISPLITPE-NLEPFRLLLE	715
Zma_DCL3	PIISPLITPD-KLAPYRLLE	PIISPLITPD-KLAPYRLLE	719
Sb1_DCL3a	PIISPLITPD-KLAPYRLLE	PIISPLITPD-KLAPYRLLE	719
Sb1_DCL3b	PIISPLITPE-NLEPFRLLLE	PIISPLITPE-NLEPFRLLLE	715
Gma_DCL3	PIISPLITPD-KLAPYRLLE	PIISPLITPD-KLAPYRLLE	723
Mtr_DCL3	PIISPLITPE-KLEPFRLLM	PIISPLITPE-KLEPFRLLM	754
Vv1_DCL3	PIISPLITPD-KLEPFRLLNE	PIISPLITPD-KLEPFRLLNE	726
Ptr_DCL3	PIISPLITPE-KLEPFRLLVE	PIISPLITPE-KLEPFRLLVE	716
Ath_DCL3	PIISPLITPD-KLQEPFRLLNE	PIISPLITPD-KLQEPFRLLNE	720
Osa_DCL4	MLIKVLTFA-NMHTNMPRLRE	MLIKVLTFA-NMHTNMPRLRE	682
Bd1_DCL4	MLIKVLSFS-GMHNNMPRLRE	MLIKVLSFS-GMHNNMPRLRE	680
Gma_DCL4a	SFLDLSIKFSSSLQESVRLRE	SFLDLSIKFSSSLQESVRLRE	688
Gma_DCL4b	SFLDLSIKFSSSLQESVRLRE	SFLDLSIKFSSSLQESVRLRE	688
Vv1_DCL4	SFLDRLSFS-DLQNNPIRLQE	SFLDRLSFS-DLQNNPIRLQE	676
Ptr_DCL4	SFLDRLSFS-NLQNNVRLKE	SFLDRLSFS-NLQNNVRLKE	687
Nat_DCL4	SFLKRWLSFT-RLQNNKRLHE	SFLKRWLSFT-RLQNNKRLHE	678
Ath_DCL4	SFLDLSVKNLS-NLQSPKRLLE	SFLDLSVKNLS-NLQSPKRLLE	686

L p 6 P 4e6 2