

Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
Zakład Biologii Molekularnej i Systemowej

Otrzymywanie i charakterystyka białek typu Dicer
pochodzących z modelowej rośliny bobowatej
Medicago truncatula

Production and characterization of Dicer-like proteins
from *Medicago truncatula*, a model legume plant

mgr Aleksander Tworak

Promotor
prof. dr hab. Marek Figlerowicz

Promotor pomocniczy
dr Anna Urbanowicz

Poznań 2014

*Serdecznie dziękuję Promotorowi mojej pracy doktorskiej,
prof. dr hab. Markowi Figlerowiczowi
oraz Promotorowi pomocniczemu, dr Annie Urbanowicz
za opiekę naukową, życzliwość i poświęcony mi czas.*

*Serdecznie dziękuję wszystkim Pracownikom
Zakładu Biologii Molekularnej i Systemowej,
bez Waszej życzliwości i pomocy praca ta nie mogłaby powstać.*

Serdecznie dziękuję Magdzie i Rodzinie, stanowicie inspirację do moich działań.

SPIS TREŚCI

I.	Wprowadzenie	6
I.1.	Niekodujące RNA	7
I.2.	Małe regulatorowe RNA	12
I.2.1.	Prekursory małych regulatorowych RNA u roślin	12
I.2.2.	Biogeneza roślinnych miRNA i siRNA	16
I.2.3.	Roślinne kompleksy wyciszające indukowane przez RNA	20
I.3.	Białka typu Dicer	25
I.3.1.	Budowa białek typu Dicer	27
I.3.2.	Funkcjonowanie białek typu Dicer	31
II.	Cel pracy	36
III.	Materiały i metody	37
III.1.	Materiały	37
III.1.1.	Oligonukleotydy DNA i RNA	37
III.1.2.	Plazmidy bakteryjne	39
III.1.3.	Szczepy bakterii	40
III.1.4.	Pożywki mikrobiologiczne	40
III.1.5.	Roztwory buforowe	41
III.1.6.	Zestawy odczynników	41
III.2.	Metody	42
III.2.1.	Uniwersalne procedury laboratoryjne	42
III.2.1.1.	Rozdział elektroforetyczny kwasów nukleinowych w warunkach natywnych	42
III.2.1.2.	Rozdział elektroforetyczny kwasów nukleinowych w warunkach denaturujących	42
III.2.1.3.	Rozdział elektroforetyczny białek w warunkach denaturujących (SDS-PAGE)	43
III.2.1.4.	Immunodetekcja białek	44
III.2.2.	Analiza genomu i badanie poziomu ekspresji genów	45
III.2.2.1.	Identyfikacja genów <i>in silico</i>	45
III.2.2.2.	Analiza bioinformatyczna sekwencji kodujących	45
III.2.2.3.	Przygotowanie zawiesiny bakterii symbiotycznych <i>Sinorhizobium meliloti</i>	45
III.2.2.4.	Przygotowanie materiału roślinnego	46
III.2.2.5.	Izolacja RNA i odwrotna transkrypcja	46
III.2.2.6.	Badanie poziomu akumulacji mRNA	47
III.2.3.	Klonowanie cDNA	48
III.2.3.1.	Amplifikacja cDNA	48
III.2.3.2.	Klonowanie cDNA w plazmidzie wysokokopijnym	49
III.2.3.3.	Analiza konstruktów plazmidowych	50
III.2.3.4.	Szybka amplifikacja końca 5' cDNA (5' RACE)	51
III.2.4.	Konstrukcja wektorów ekspresyjnych	51
III.2.4.1.	Synteza cDNA do konstrukcji wektorów ekspresyjnych	51
III.2.4.2.	Trawienie dsDNA enzymami restrykcyjnymi do celów preparatywnych	52
III.2.4.3.	Klonowanie sekwencji kodujących w wektorach ekspresyjnych	52
III.2.5.	Produkcja białek	53

III.2.5.1.	Przygotowanie i przechowywanie szczepów bakteryjnych.....	53
III.2.5.2.	Synteza białka w bakteriach	54
III.2.5.3.	Analiza rozpuszczalności białka.....	54
III.2.5.4.	Izolacja i oczyszczanie białka na złożu z immobilizowanym glutationem	55
III.2.6.	Badanie aktywności rybonukleazowej białek	56
III.2.6.1.	Otrzymywanie substratów RNA.....	56
III.2.6.2.	Znakowanie radioizotopowe substratów RNA.....	57
III.2.6.3.	Analiza aktywności rybonukleazowej białek	58
IV.	Wyniki	59
IV.1.	Identyfikacja genów białek DCL w genomie <i>Medicago truncatula</i>	59
IV.2.	Analiza profilu ekspresji genów <i>MtDCL</i>	65
IV.3.	Klonowanie i charakterystyka cDNA wybranych białek MtDCL	70
IV.4.	Produkcja białek MtDCL1 i MtDCL2a	75
IV.5.	Badanie aktywności katalitycznej białek MtDCL1 i MtDCL2a	79
V.	Dyskusja	88
V.1.	Białka DCL kodowane przez genom <i>Medicago truncatula</i>	88
V.2.	Porównanie aktywności katalitycznej białek MtDCL1 i MtDCL2a	94
VI.	Wnioski.....	98
VII.	Streszczenie / Abstract	99
VIII.	Finansowanie	101
IX.	Literatura	102
X.	Załączniki.....	119

WYKAZ STOSOWANYCH SKRÓTÓW

- 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*)
- DRA** – ATPaza aktywowana dupleksem RNA (ang. *duplex RNA activated ATPase*)
- DRB** – białko wiążące dsRNA
(ang. *dsRNA binding protein*)
- DRH** – helikazy spokrewnione z białkami Dicer (ang. *Dicer-related helicase*)
- ddPCR** – cyfrowy PCR w kroplach
(ang. *droplet digital PCR*)
- 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*)
- GST** – S-transferaza glutationu
(ang. *glutathione S-transferase*)
- hc-siRNA** – heterochromatynowy mały interferujący RNA
(ang. *heterochromatic siRNA*)
- lncRNA** – długi niekodujący RNA
(ang. *long non-coding RNA*)
- mRNA** – RNA informacyjny
(ang. *messenger RNA*)
- miRNA** – mikroRNA (ang. *microRNA*)
- NAT-siRNA** – mały interferujący RNA pochodzący z naturalnych transkryptów antysensownych (ang. *natural antisense transcript siRNA*)
- ncRNA** – niekodujący RNA
(ang. *non-coding RNA*)
- PAZ** – domena Piwi/Argonaute/Zwill
(ang. *Piwi/Argonaute/Zwill*)
- 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 wyciszanie genów (ang. *post-transcriptional gene silencing*)
- RACE** – szybka amplifikacja końca cDNA
(ang. *rapid amplification of cDNA end*)
- 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*)
- RNAi** – interferencja RNA
(ang. *RNA interference*)
- SDS-PAGE** – elektroforeza w żelu poliakryloamidowym w warunkach denaturujących (ang. *SDS-polyacrylamide gel electrophoresis*)
- siRNA** – mały interferujący RNA
(ang. *small interfering RNA*)
- sncRNA** – mały niekodujący RNA
(ang. *small non-coding RNA*)
- ssRNA** – jednoniciowy RNA
(ang. *single-stranded RNA*)
- ta-siRNA** – mały interferujący RNA działający *in trans* (ang. *trans-acting siRNA*)
- TGS** – transkrypcyjne wyciszanie genów
(ang. *transcriptional gene silencing*)
- vsRNA** – mały interferujący RNA pochodzenia wirusowego
(ang. *virus-derived siRNA*)

I. WPROWADZENIE

Zgodnie z ogłoszonym na początku lat pięćdziesiątych ubiegłego wieku tzw. centralnym dogmatem biologii molekularnej proces ekspresji informacji genetycznej przebiega od DNA, poprzez RNA, do finalnego produktu jakim są białka. Przyjęty wówczas schemat akcentował szczególnie istotną rolę DNA, jako nośnika informacji, oraz białek. Uważano, że te niezwykle zróżnicowane produkty końcowe procesu ekspresji informacji genetycznej, pełnią większość strukturalnych, katalitycznych i regulatorowych ról warunkujących prawidłowe funkcjonowanie komórki. Równocześnie kwasom rybonukleinowym przypisywano jedynie drugoplanowe role pośrednika w procesie wyrażania informacji genetycznej czy szkieletu spajającego multienzymatyczne kompleksy. Szereg istotnych odkryć dokonanych w ostatnich dwóch dekadach dowodzi jednak, iż zaproponowany ponad 50 lat temu opis jest daleki od rzeczywistości, szczególnie w odniesieniu do komórek eukariotycznych. Wskazówek w tym zakresie dostarczyły liczne projekty sekwencjonowania genomów, a następnie transkryptomów organizmów eukariotycznych w tym i genomu człowieka. Analiza sekwencji genomowych wykazała, iż jedynie ich niewielka część odpowiada za kodowanie białek. Z drugiej strony zaobserwowano, iż większość genomu jest aktywna transkrypcyjnie. Okazało się, że białka nie są jedynym produktem ekspresji informacji genetycznej, a pula cząsteczek RNA obecnych w komórkach jest znacznie większa i bardziej różnorodna niż pierwotnie sądzono.

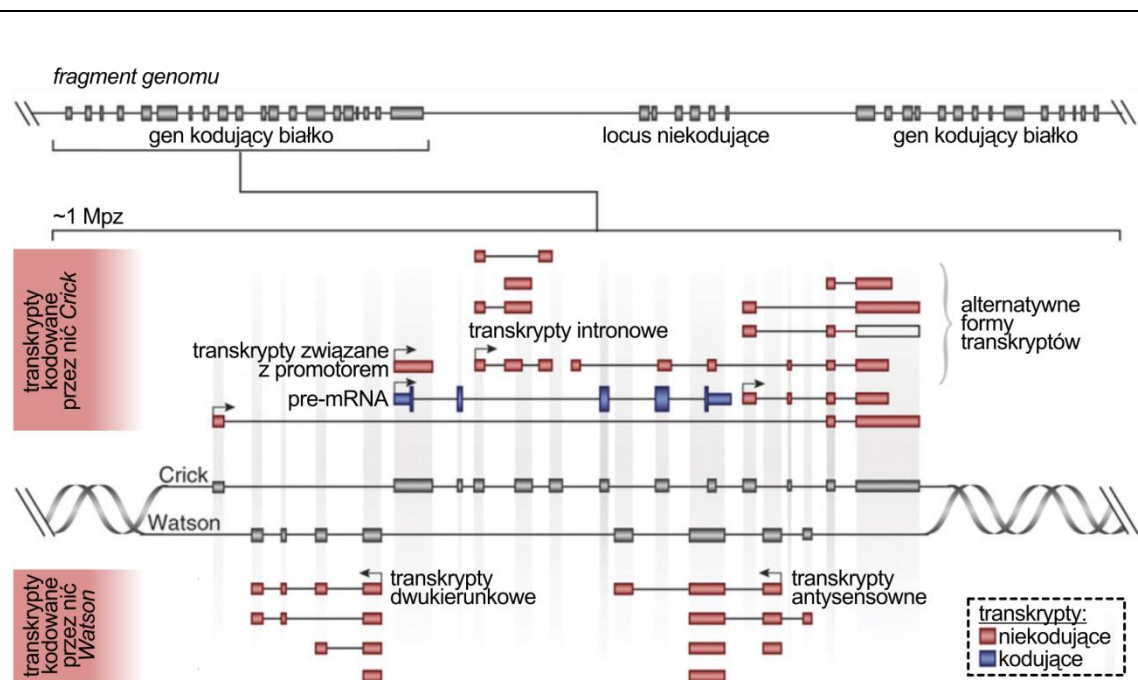
Zidentyfikowano szereg istotnych cech RNA, które predysponują te cząsteczki do pełnienia rozmaitych funkcji biologicznych. RNA może przyjmować zróżnicowane, wysoce zorganizowane struktury wyższego rzędu. Niektóre z tych cząsteczek, określane mianem rybozymów, wykazują aktywność katalityczną. Ponadto cząsteczki RNA posiadają zdolność do specyficznej hybrydyzacji z innymi RNA lub DNA oraz do tworzenia kompleksów z białkami. Dzięki temu cząsteczki RNA mogą nadawać relatywnie uniwersalnym białkom efektorowemu wysoką specyficzności wobec określonej sekwencji nukleotydowej. Ta strukturalna i funkcjonalna wszechstronność RNA leży także u podstaw hipotezy „Świata RNA”, według której połączenie w cząsteczkach RNA funkcji katalitycznej oraz informacyjnej umożliwiło powstanie i ewolucję pierwszym formom życia na Ziemi. Obecnie lepiej wyspecjalizowane biomolekuły, białka oraz DNA, pełnią rolę odpowiednio katalizatorów reakcji oraz nośnika informacji. Jednak w komórkach wszystkich organizmów żywych nadal obecna jest ogromna pula różnorodnych cząsteczek RNA. Ich rola w wielu przypadkach została już określona, w większości jednak nadal pozostaje nieznana.

I.1. NIEKODUJĄCE RNA

Wszystkie transkrypty obecne w komórce (tzw. transkryptom) podzielić można na kodujące i niekodujące, w zależności od tego czy stanowią matrycę w procesie biosyntezy białek. Do niekodujących RNA (ncRNA, ang. *non-coding RNA*) zalicza się kilka znanych od dawna i powszechnych klas transkryptów, takich jak RNA rybosomalne (rRNA, ang. *ribosomal RNA*) i RNA transportujące (tRNA, ang. *transfer RNA*), czy nieco później odkryte małe jądrowe RNA (snRNA, ang. *small nuclear RNA*) i małe jąderkowe RNA (snoRNA, ang. *small nucleolar RNA*). Jednak dopiero postępujący od początku XXI wieku rozwój niezwykle czułych, a równocześnie wysokoprzepustowych technologii mikromacierzowych oraz głębokiego sekwencjonowania pozwolił na identyfikację całego spektrum niekodujących transkryptów występujących w komórkach eukariotycznych. Zaobserwowano, że transkrypcji ulegać może nawet trzy czwarte genomu człowieka czy innych ssaków, mimo iż obszar kodujący białka obejmuje mniej niż 2% każdego z tych genomów [1, 2, 3]. Niekodujące transkrypty stanowią pulę niezwykle zróżnicowanych pod względem wielkości cząsteczek, mogących pochodzić z obu nici DNA i z wszystkich rejonów genomu, wliczając w to także geny kodujące białka (rejony promotorowe, introny, fragmenty egzonów) [4]. Na różnorodność ncRNA wpływ ma między innymi zjawisko alternatywnego składania, występowanie alternatywnych miejsc inicjacji i terminacji transkrypcji oraz możliwość kombinacji tych zjawisk [5]. W konsekwencji poszczególne transkrypty często są wzajemnie komplementarne lub zawierają fragmenty o identycznej sekwencji (Rys. 1.1). Szacuje się, że u człowieka z każdego genu kodującego białko powstaje średnio 10 różnych typów kodujących i niekodujących transkryptów [6]. Ponadto w przypadku połowy genów obok zwykłych transkryptów, powstają także transkrypty o sekwencji antysensownej. W efekcie różnorodność ncRNA może znacznie przewyższać różnorodność transkryptów kodujących białka [7, 8]. Najobszerniejszy katalog zidentyfikowanych ludzkich genów oraz transkryptów powstaje w ramach projektu GENCODE [9]. Baza ta (wersja 19) obejmuje obecnie ponad 57 tysięcy ludzkich genów (z czego jedynie niecałe 21 tysięcy koduje białka) oraz prawie 200 tysięcy różnorodnych transkryptów.

Obecność w komórkach tak dużej i zróżnicowanej puli ncRNA sugeruje, że mają one istotne znaczenie biologiczne. Wskazuje na to kilka przesłanek. Część ncRNA pochodzi z wysoce konserwatywnych sekwencji zlokalizowanych poza regionami

kodującymi białka [10]. Ponadto transkrypcja niektórych ncRNA jest regulowana na różnych etapach rozwoju organizmu, jest zależna od stanów chorobowych, niekiedy wykazuje też specyficzność tkankową [11]. Ważną wskazówką dotyczącą potencjału funkcjonalnego ncRNA jest zaobserwowana pozytywna korelacja pomiędzy stopniem skomplikowania organizmów, a udziałem sekwencji niekodujących w ich genomach [12]. Zależność taka, wbrew wcześniejszym przypuszczeniom, nie zachodzi w odniesieniu do żadnego innego prostego parametru opisującego genom, jak choćby wielkość czy liczba genów kodujących białka. Głównym wyzwaniem ewolucyjnym związanym ze wzrostem złożoności organizmów wydaje się być stworzenie mechanizmów umożliwiających precyzyjną kontrolę zachodzących w nich procesów. W przypadku organizmów prokariotycznych, których genomy złożone są w około 90% z sekwencji kodujących białka, stwierdzono, że wraz z liniowym wzrostem wielkości genomu, a więc liczby kodowanych białek, następuje potęgowy wzrost liczby białek regulatorowych. W konsekwencji wzrost złożoności komórek, których procesy kontrolne oparte są w głównej mierze na działaniu białek regulatorowych limitowany



Rysunek 1.1. Różnorodność niekodujących transkryptów w komórkach eukariotycznych. Źródłem niekodujących RNA (kolor czerwony) są zarówno geny jak i loci niekodujące. Transkrypcja odbywać się może z promotora genu (może być dwukierunkowa) oraz z intronów. Dla wielu genów powstają transkrypty antysensowne. Zróżnicowane formy transkryptów powstają w wyniku ich alternatywnego składania, występowania alternatywnych miejsc inicjacji i terminacji transkrypcji oraz kombinacji tych zjawisk (reprodukcja za zgodą © Macmillan Publishers Ltd) [13].

jest przez możliwość zakodowania w genomie niezbędnej ich liczby [14]. Limit ten tożsamy jest z wielkością największych znanych genomów prokariotycznych, liczących około 10 milionów par zasad. Stąd przypuszczenie, że ewolucja komórek eukariotycznych wymagała wykorzystania bardziej ekonomicznego systemu regulatorowego, opartego o ncRNA, a skala tego zjawiska nie jest jeszcze dobrze poznana [15].

Jak istotne znaczenie funkcjonalne mogą mieć ncRNA obecne w komórkach zademonstrowano jeszcze zanim w pełni zdano sobie sprawę ze złożoności eukariotycznego transkryptomu. W latach dziewięćdziesiątych XX wieku nastąpiło odkrycie mechanizmu regulacji ekspresji genów zwanego interferencją RNA (RNAi, ang. *RNA interference*). Zaobserwowano wtedy, że wprowadzenie do komórek dwuniciowego RNA (dsRNA, ang. *double-stranded RNA*) indukuje proces wyciszania ekspresji genów zawierających sekwencje homologiczne do wprowadzonego RNA [16]. Intensywne badania tego zjawiska ujawniły, iż kluczową rolę pełnią w nim krótkie jednoniciowe cząsteczki RNA. Zgromadzone do dziś dane wskazują, że te ncRNA, znane jako małe regulatorowe RNA, pełnią podstawową rolę w procesie regulacji ekspresji większości genomów eukariotycznych. U człowieka wpływają na poziom ekspresji co najmniej połowy genów kodujących białka [17].

Obecnie podziału niekodujących transkryptów dokonuje się ze względu na ich wielkość. Długie niekodujące RNA (lncRNA, ang. *long non-coding RNA*) obejmują pulę cząsteczek o długości powyżej 200 nukleotydów, natomiast małe niekodujące RNA (sncRNA, ang. *small non-coding RNA*) to cząsteczki krótsze niż 200 nukleotydowe. Podział ten w dużej mierze wynika z możliwości łatwej separacji tych dwóch frakcji RNA w procesie ich izolacji z komórek [4]. W ramach każdej grupy wyróżniono szereg podklas, jednak z dużą dozą pewności można stwierdzić, że wciąż nie znamy ich pełnej różnorodności.

Jak już wspomniano, do grupy lncRNA zaliczane są transkrypty złożone z co najmniej 200 nukleotydów. Niekiedy jednak ich długość przekracza nawet i 100 tysięcy nukleotydów [18]. Biogeneza lncRNA przypomina proces powstawania cząsteczek mRNA. Wykazano, że regiony promotorowe fragmentów genomu kodujących lncRNA wiążą różnorodne czynniki transkrypcyjne, posiadają także charakterystyczne modyfikacje białek histonowych [19, 20]. Większość lncRNA jest transkrybowana przez polimerazę RNA II (Pol II) i ulega typowym dla mRNA modyfikacjom potranskrypcyjnym: przyłączeniu czapeczki 7-metyloguanozynowej na końcu 5`,

poliadenylacji końca 3' oraz procesowi składania. W odróżnieniu od genów kodujących białka, sekwencje kodujące lncRNA są mniej konserwatywne, ich ekspresja zachodzi na dużo niższym poziomie i często jest specyficzna dla określonego rodzaju komórki lub tkanki [21]. Zidentyfikowano także stabilne niekodujące transkrypty będące produktami polimerazy RNA III (Pol III) [22], oraz specyficznej dla roślin polimerazy RNA V (Pol V) [23]. LncRNA powstają także w trakcie składania mRNA oraz biogenezy snoRNA [24].

W ramach projektu GENCODE skatalogowano dotąd ponad 23 tysiące długich niekodujących transkryptów, pochodzących z ponad 13 tysięcy loci rozmieszczonych w genomie człowieka. Liczby te odnoszą się do wersji 19 bazy danych i rosną z każdą kolejną jej aktualizacją [9]. Jak dotąd tylko niewielkiej części lncRNA przypisano jakąś konkretną funkcję biologiczną, niemniej grupa ta wciąż się powiększa. Na podstawie poczynionych dotąd obserwacji postuluje się, iż lncRNA mogą pełnić cztery główne, niewykluczające się wzajemnie funkcje: (i) cząsteczek sygnałowych biorących udział w procesie regulacji ekspresji genów, (ii) sond zapewniających białkom specyficzność wobec określonych sekwencji w DNA lub RNA, (iii) szkieletów spajające kompleksy rybonukleoproteinowe oraz (iv) tzw. pułapek lub przynęt (ang. *decoys*) stanowiących alternatywne miejsca wiązania czynników regulatorowych (białek oraz RNA) [25, 26].

Ciekawym przykładem lncRNA są tzw. gąbki mikroRNA (ang. *microRNA sponges*), długie niekodujące RNA, w obrębie których występuje wiele miejsc wiązania małych regulatorowych RNA typu mikroRNA (miRNA, ang. *microRNA*; zob. Rozdz. I.2) [27]. Jednym z pierwszych poznanych niekodujących RNA tego typu był roślinny transkrypt IPS1 (ang. *Induced by phosphate starvation 1*). U *Arabidopsis thaliana* wpływa on na stabilność mRNA genu *PHO1* poprzez wiązanie cząsteczek miRNA miR399 uczestniczących w wyciszaniu tego mRNA. Z uwagi na niepełną komplementarność w miejscu wiązania nie dochodzi do trawienia IPS1, następuje natomiast wydajne unieczynnienie cząsteczek miR399 [28]. Kolejne doniesienia wskazują także na rolę słabo dotąd poznanych kolistych RNA (circRNA, ang. *circular RNA*) w regulacji ekspresji genów poprzez kompetycyjne wiązanie miRNA. Przykładowo koliste RNA ciRS-7 (ang. *circular RNA sponge for miR-7*) zawiera ponad 70 miejsc wiązania ludzkiego mikroRNA miR-7, dzięki czemu efektywnie ogranicza jego aktywność [29].

Drugą grupę ncRNA stanowią krótkie transkrypty, o długości nieprzekraczającej 200 nukleotydów. W bazie GENCODE (wersja 19) skatalogowano

dotąd ponad 9 tysięcy ludzkich genów dla sncRNA [9]. Wśród sncRNA wyróżnia się kilka dobrze zdefiniowanych klas o zróżnicowanych funkcjach. Najwcześniej poznane tRNA występują we wszystkich organizmach, zaangażowane są w proces translacji, a za syntezę ich prekursorów odpowiada Pol III. SnRNA występują w jądrach organizmów eukariotycznych i biorą udział między innymi w procesie składania transkryptów oraz regulacji transkrypcji. Z kolei snoRNA zlokalizowane są głównie w jąderku oraz ciałkach Cajala w jądrach komórek eukariotycznych. Częsteczki te odgrywają istotną rolę w procesie chemicznej modyfikacji innych niekodujących RNA. Zarówno snRNA jak i snoRNA transkrybowane są w większości przez Pol II [30]. Największą grupę poznanych dotąd sncRNA stanowią jednak tzw. małe regulatorowe RNA. Są to cząsteczki o zróżnicowanym pochodzeniu, które łączy sposób funkcjonowania. Te około 20-30-nukleotydowe cząsteczki nadają dużej grupie białek regulatorowych z rodziny Argonaute (AGO, ang. *Argonaute*) specyficzność wobec określonych sekwencji RNA/DNA. Białka AGO w kompleksach z małymi regulatorowymi RNA pełnią w komórce różnorodne funkcje, zazwyczaj związane z regulacją ekspresji genów.

I.2. MAŁE REGULATOROWE RNA

Wśród małych regulatorowych RNA najlepiej poznaną grupę stanowią około 21-nukleotydowe miRNA. Powstają one z kodowanych w genomie jednoniciowych prekursorów. MiRNA występują zarówno w organizmach roślinnych jak i zwierzęcych, u człowieka skatalogowano dotąd ponad 3 tysiące kodujących je genów (GENCODE, wersja 19; [9]). Drugą powszechnie występującą grupę małych regulatorowych RNA stanowią małe interferujące RNA (siRNA, ang. *small interfering RNA*). Mają one długość około 20-25 nukleotydów, a powstają z prekursorów o egzogennym lub endogennym pochodzeniu. Prekursorami siRNA mogą być zarówno jednoniciowe jak i dwuniciowe RNA. Choć siRNA zidentyfikowano u większości organizmów eukariotycznych, ich największą różnorodność obserwuje się u roślin.

Osobną grupę małych regulatorowych RNA tworzą RNA oddziałujące z białkami Piwi (piRNA, ang. *Piwi-interacting RNA*). Białka Piwi należą do specyficznej głównie dla komórek linii zarodkowych podklasy białek typu AGO. W odróżnieniu od miRNA i siRNA cząsteczki piRNA są nieco dłuższe, około 20-30 nukleotydowe. Szczegółowy mechanizm biogenezy i funkcjonowania piRNA nie jest jeszcze dobrze poznany. Obecnie wydaje się, że ich główną funkcją jest wyciszenie różnorodnych ruchomych elementów genetycznych w komórkach rozrodczych zwierząt, co warunkuje także prawidłowy rozwój gonad. U roślin jak dotąd nie wykazano obecności małych regulatorowych RNA typu piRNA [31].

I.2.1. PREKURSORY MAŁYCH REGULATOROWYCH RNA U ROŚLIN

MiRNA oraz siRNA powstają z prekursorów o zróżnicowanym pochodzeniu. Cechą wspólną prekursorów jest obecność w ich strukturze fragmentu dwuniciowego, z którego wycinana jest cząsteczka regulatorowa. Większość roślinnych prekursorów miRNA kodowana jest przez geny stanowiące pojedyncze, niezależne jednostki transkrypcyjne [32]. Znane są także roślinne geny miRNA zlokalizowane w intronach genów kodujących białka, bądź kodujące kilka prekursorów, niemniej sytuacje takie są znacznie bardziej powszechne u zwierząt [33]. Około 20% znanych miRNA u *Arabidopsis thaliana*, ryżu (*Oryza sativa*) oraz topoli (*Populus trichocarpa*) pochodzi z tzw. transkryptów policistronowych. W większości przypadków transkrypty takie obejmują tandemowe powtórzenia prekursorów tego samego miRNA. Sugeruje to, że powstały one w wyniku lokalnych duplikacji, prowadzących do podniesienia poziomu

ekspresji miRNA w ramach danej jednostki transkrypcyjnej [34]. U ryżu zidentyfikowano także specyficzną grupę miRNA, których geny zlokalizowane są na anstysensownej nici DNA w stosunku do genu podlegającego regulacji. Znane są one jako nat-miRNA (ang. *natural antisense miRNA*) [35].

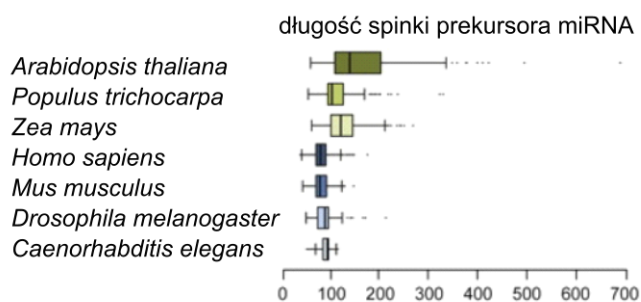
Transkrypcja genów kodujących miRNA prowadzi do powstania tzw. pierwotnych prekursorów (pri-miRNA, ang. *primary miRNA*) i zachodzi przy udziale Pol II. Promotory roślinnych genów miRNA wykazują nadreprezentację wielu znanych motywów regulatorowych, w tym motywu TATA, co sugeruje możliwość wiązania różnorodnych czynników transkrypcyjnych i złożony proces regulacji ich ekspresji [36]. Jednym z czynników regulujących transkrypcję genów miRNA jest Mediator, wielobiałkowy kompleks oddziałujący z Pol II i odpowiedzialny za wiązanie licznych aktywatorów i represorów transkrypcji [37]. Wykazano także wpływ biotycznych i abiotycznych czynników stresowych na poziom ekspresji poszczególnych genów miRNA [38]. Potranskrypcyjna modyfikacja pri-miRNA obejmuje przyłączenie czapeczki 7-metyloguanozynowej na końcu 5', poliadenylację końca 3', często także proces składania [39]. Niektóre pri-miRNA podlegać mogą alternatywnemu składaniu [40].

Pri-miRNA mogą mieć długość wielu tysięcy nukleotydów. Sekwencja dojrzałego miRNA zlokalizowana jest zawsze w obrębie struktury typu spinki do włosów (ang. *hairpin structure*) – w jej dwuniciowym fragmencie (tzw. trzonie). Trzony spinek, z których generowane są miRNA charakteryzuje obecność wielu niesparowań. Ponadto u roślin struktury te mają bardzo zróżnicowaną długość, od około 70 do ponad 900 nukleotydów. Odróżnia je to od ich zwierzęcych odpowiedników, charakteryzujących się dość jednorodną długością, z reguły nieprzekraczającą 100 nukleotydów (Rys. 1.2) [41].

W odróżnieniu od prekursorów miRNA dwuniciowe fragmenty prekursorów siRNA są w pełni komplementarne lub zawierają niewielką liczbę niesparowań. Prekursory te mają także dużo bardziej zróżnicowane pochodzenie, częściowo odzwierciedlone w stosowanej klasyfikacji siRNA. Mogą one być utworzone poprzez syntezę drugiej nici na matrycy jednoniciowego transkryptu, bądź z dwóch wzajemnie komplementarnych transkryptów. W pierwszym przypadku w proces zaangażowana jest grupa białek o aktywności polimerazy RNA zależnej od RNA (RDR, ang. *RNA-dependent RNA polymerase*). Białka te kodowane są w genomach między innymi wirusów RNA, roślin i bezkręgowców, nie wykazano natomiast ich obecności u takich

organizmów jak *Drosophila melanogaster*, mysz czy człowiek [42]. *Arabidopsis thaliana* posiada 6 genów kodujących polimerazy RNA zależne od RNA (RDR1-6). Przynajmniej 3 białka RDR (RDR1, RDR2, RDR6) biorą udział w procesie biogenezy siRNA.

W puli endogennych małych regulatorowych RNA identyfikowanych u *Arabidopsis* hodowanych w standardowych warunkach 10% stanowią miRNA, a pozostałe 90% – siRNA [43]. Największą grupę małych regulatorowych RNA u roślin stanowią tzw. heterochromatynowe małe interferujące RNA (hc-siRNA, ang. *heterochromatic siRNA*). Powstają one z obecnych w genomie, często w formie heterochromatyny, ruchomych elementów genetycznych i innych sekwencji powtórzonych. Specyficzna dla roślin Polimeraza RNA IV (Pol IV) odpowiada za transkrypcję tych sekwencji, warunkuje także dalsze etapy biogenezy hc-siRNA [44]. Wykazano, że biogeneza ponad 90% roślinnych siRNA zależna jest od aktywności Pol IV [45]. Transkrypty Pol IV stanowią matryce, na których RDR2 syntetyzuje nić komplementarną [43]. Białka Pol IV i RDR2 oddziałują *in vivo*, natomiast w testach *in vitro* RDR2 jest aktywne jedynie w kompleksie z Pol IV [46, 47]. Ponadto działanie obu białek wspomaga przynajmniej jeden kofaktor: białko modyfikujące strukturę chromatyny CLASSY1 (CLSY1) [48]. Sugeruje to, że dwuniciowe prekursorzy hc-siRNA tworzone są niezwykle szybko, prawdopodobnie dobudowanie drugiej nici RNA przez RDR2 następuje w trakcie lub tuż po etapie transkrypcji.



Rysunek 1.2. Długość struktur typu spinki do włosów obecnych w roślinnych i zwierzęcych prekursorach miRNA. Spinki roślinnych prekursorów są zwykle dłuższe, dodatkowo charakteryzują się bardziej zróżnicowaną długością niż ich zwierzęce odpowiedniki. Prostokątami zaznaczono zakres długości spinek zidentyfikowanych w danym organizmie – od 25 do 75 percentyla (reprodukcja za zgodą © ASPB) [41].

Osobną grupę stanowią tzw. drugorzędowe siRNA, w których powstawanie zaangażowane są inne małe regulatorowe RNA, z reguły miRNA. Do grupy tej należą roślinne małe interferujące RNA działające *in trans* (ta-siRNA, ang. *trans-acting siRNA*). Źródłem ta-siRNA są zarówno niektóre komórkowe mRNA jak i specyficzne niekodujące transkrypty pochodzące z genów *TAS* [49]. U *Arabidopsis* zidentyfikowano dotąd 4 rodziny genów *TAS*, transkrypcja wszystkich zachodzi z udziałem Pol II. Pierwotne transkrypty *TAS* mają długość około tysiąca nukleotydów i podobnie jak mRNA posiadają czapkę 7-metyloguanozynową na końcu 5` i poliadenylowany koniec 3` [50]. Transkrypty stanowiące źródło ta-siRNA rozpoznawane są przez jedno lub kilka specyficznych miRNA, co prowadzi do ich rozcięcia przez białko z rodziny AGO [51]. Do niektórych z powstałych fragmentów RNA komplementarną nicią dobudowuje enzym RDR6, tworząc prekursor drugorzędowych ta-siRNA. Nie są jasne mechanizmy decydujące o wykorzystaniu danego fragmentu pierwotnego transkrypty jako matrycy dla RDR6, zwłaszcza że z reguły po przecięciu przez białko AGO następuje szybka degradacja powstałych produktów (zob. Rozdz. I.2.3). Wykazano, że długość cząsteczki miRNA biorącej udział w przecięciu pierwotnego transkrypty (22 nukleotydy zamiast typowych 21) oraz obecność białka SGS3 (ang. *Suppressor of gene silencing 3*) mają wpływ na stabilizację RNA stanowiącego matrycę dla RDR6 [52, 53].

Dwuniciowe prekursor roślinnych siRNA mogą także zostać utworzone przez hybrydyzację dwóch transkryptów. Komplementarne fragmenty RNA pochodzący mogą zarówno z transkrypcji zachodzącej na obu niciach DNA w danym locus, jak i z dwóch osobnych loci. Powstają w ten sposób tzw. siRNA pochodzące z naturalnych transkryptów antysensownych (NAT-siRNA, ang. *natural antisense transcript siRNA*), oznaczane odpowiednio *cis*-NAT-siRNA lub *trans*-NAT-siRNA. U *Arabidopsis* ponad 30% genomu ulega jednoczesnej transkrypcji z obu nici DNA. Jednak tylko dla około 6% par antysensownych transkryptów potwierdzono akumulację odpowiadających im cząsteczek *cis*-NAT-siRNA [54, 55]. Prekursor *cis*-NAT-siRNA nie mają wspólnej drogi biogenezy. Akumulacja *cis*-NAT-siRNA u *Arabidopsis* zwykle zależna jest od obecności Pol IV, a w niektórych przypadkach także od białek typu RDR (RDR6 lub RDR2) [55]. W większości scharakteryzowanych przypadków jeden z transkryptów tworzących prekursor *cis*-NAT-siRNA ulega ekspresji konstytutywnej, podczas gdy ekspresja drugiego indukowana jest w określonej sytuacji, na przykład przez czynnik stresowy [56]. Jak dotąd nie są znane przypadki funkcjonowania w roślinach *trans*-

NAT-siRNA, pochodzących jak na przykład u myszy z transkryptów genu i homologicznego pseudogenu [57, 58, 59].

Małe niekodujące RNA są także wykorzystywane przez rośliny do obrony przed wirusami. Najczęściej prekursorzy siRNA pochodzenia wirusowego (vsiRNA, ang. *virus-derived siRNA*) stanowią transkrypty, w których obecna jest struktura typu spinki, podobnie jak ma to miejsce w przypadku prekursorów miRNA. Dotyczy to zarówno wirusów DNA, RNA jak i wiroidów. Ponadto genomy wielu wirusów RNA przekształcane są w procesie replikacji w dwuniciową formę pośrednią, przy udziale wirusowych białek typu RDR. W obu przypadkach dwuniciowe fragmenty RNA pochodzenia wirusowego stanowią źródło tzw. pierwszorzędowych vsiRNA. Z kolei przy udziale roślinnych białek RDR (RDR6 i RDR1) powstają tzw. drugorzędowe vsiRNA [60].

I.2.2. BIOGENEZA ROŚLINNYCH miRNA I siRNA

Kluczowy etap biogenezy, wspólny dla większości małych regulatorowych RNA, obejmuje wycięcie z dwuniciowego fragmentu prekursora, krótkiego dupleksu RNA zawierającego cząsteczkę regulatorową (Rys. 1.3). Za trawienie prekursorów małych regulatorowych RNA odpowiadają wyspecjalizowane białka o aktywności rybonukleazy typu III (RNazy III). U zwierząt funkcjonują dwie niezależne grupy rybonukleaz zaangażowanych w biogenezę małych regulatorowych RNA: Drosha oraz Dicer [61]. Białko Drosha zlokalizowane jest w jądrze i bierze udział w biogenezie większości zwierzęcych miRNA. Działa ono w dużym kompleksie białkowym znanym jako Mikroprocesor, obejmującym między innymi białko DGCR8, które w testach *in vitro* warunkuje wydajne trawienie prekursorów [62]. Drosha dokonuje pierwszego trawienia w obrębie spinki pri-miRNA (przecięcia obu nici w obrębie jej trzonu), co prowadzi do powstania tzw. pre-miRNA (ang. *precursor miRNA*). Pre-miRNA podlega eksportowi do cytoplazmy z udziałem Eksportyny 5, gdzie rybonukleaza Dicer wyzwala dupleks RNA. Dupleks ten tworzy cząsteczka miRNA oraz komplementarna do niej tzw. nić pasażerska oznaczana jako miRNA*. Rybonukleaza Dicer odpowiada także za biogenezę różnego typu zwierzęcych siRNA [33, 63].

U roślin za trawienie prekursorów wszystkich typów małych regulatorowych RNA odpowiada jedna rodzina rybonukleaz: białka typu Dicer (DCL, ang. *Dicer-like proteins*) stanowiące homologi zwierzęcych białek Dicer. Białka DCL pełnią tym

samym rolę zwierzęcych białek Drosha oraz Dicer. W szlaku biogenezy miRNA odpowiadają za trawienie zarówno pri-miRNA jak i pre-miRNA, co prowadzi do utworzenia dupleksów miRNA/miRNA*. Uważa się, że biogeneza małych regulatorowych RNA u roślin zachodzi w całości w jądrze, jako że zarówno DCL, jak i pozostałe zaangażowane w ten proces białka wykazują taką lokalizację. Dopiero właściwe cząsteczki regulatorowe podlegają eksportowi do cytoplazmy.

Obecnie wyróżnia się cztery rodzaje roślinnych białek typu Dicer, każde wyspecjalizowane w generowaniu określonej grupy małych regulatorowych RNA. DCL1 bierze udział głównie w biogenezie miRNA i generuje produkty o długości około 21 nukleotydów, z kolei białka DCL2, DCL3 i DCL4 zaangażowane są głównie w biogenezę siRNA i generują produkty, z których większość ma długość odpowiednio 22, 24 i 21 nukleotydów [64]. Cechą charakterystyczną produktów generowanych przez wszystkie białka typu Dicer jest obecność 2-nukleotydowych wystających końców 3' [65].

W przypadku DCL1 wpływ na wydajność oraz precyzję cięcia mają przynajmniej dwa kofaktory tworzące kompleks z enzymem: białko wiążące dsRNA HYL1 (znane także jako DRB1, ang. *dsRNA binding protein 1*) oraz białko posiadające charakterystyczny motyw strukturalny palca cynkowego SERRATE (SE) [66, 67]. Większość z komórkowej puli DCL1 i HYL1 skupiona jest w charakterystycznych strukturach jądrowych, tzw. *Dicing bodies*, z reguły zlokalizowanych w pobliżu jąderka. Obecne są tam także białka SE oraz cząsteczki pri-miRNA, co sugeruje, że cięcie prekursorów miRNA odbywa się głównie w obrębie tych struktur [68].

Dla większości znanych rodzin miRNA u *Arabidopsis* wykazano, że produktom generowanym przez DCL1 towarzyszy niewielka pula cząsteczek wytwarzanych przez DCL3. Białko DCL3 może zatem konkurować o prekursory miRNA z DCL1, co prowadzi do powstawania nieco dłuższych, około 24-nukleotydowych wariantów miRNA, tzw. długich miRNA (lmiRNA, ang. *long miRNA*) [69]. Znane są także prekursory miRNA podlegające trawieniu wyłącznie przez DCL4 [70, 71]. Wiele z nich to stosunkowo młode ewolucyjnie prekursory miRNA. Obecna w nich struktura typu spinki zawierająca miRNA ma bardziej stabilną strukturę i zawiera mniej niesparowań, niż ma to miejsce w przypadku substratów DCL1. Często duża część takiego prekursora wykazuje komplementarność do genu podlegającego regulacji przez dany miRNA. W starszych ewolucyjnie pri-miRNA jedynie sekwencja miRNA wykazuje komplementarność do podlegającej regulacji sekwencji docelowej. Wskazuje

to na jeden ze sposobów powstawania genów miRNA. Dwa ułożone tandemowo i odwrócone względem siebie duplikaty danego genu (lub jego fragmentu) umożliwiają mogą powstawanie transkryptów zawierających strukturę typu spinki z niewielką liczbą niesparowań. Tego typu cząsteczki stanowią mogą substraty dla enzymów DCL3 czy DCL4. Stopniowa akumulacja mutacji może powodować zmianę w preferencji cięcia prekursora przez poszczególne enzymy, na korzyść DCL1 [72].

Dużo mniej wiadomo na temat biogenezy siRNA. Roślinne białka DCL wykazują specjalizację w cięciu prekursorów poszczególnych typów siRNA (Tabela 1.1). Za cięcie prekursorów hc-siRNA (zależnych od Pol IV i RDR2) odpowiada białko DCL3. Dwuniciowe prekursorzy ta-siRNA (zależne od Pol II i RDR6) cięte są głównie przez białko DCL4, które funkcjonuje w kompleksie z białkiem DRB4, homologiem HYL1 (DRB1). W testach *in vitro* obecność DRB4 warunkuje aktywność katalityczną DCL4 [73]. Z kolei NAT-siRNA, których szlaki biogenezy są najbardziej zróżnicowane, wykazują w dużej mierze zależność od DCL3 i DCL1, przy czym wykazano, że wiele prekursorów podlega trawieniu przez oba te białka [55]. Także DCL2 uczestniczy w biogenezie niektórych NAT-siRNA [74]. Za cięcie prekursorów pochodzących z wirusów RNA zwykle odpowiadają białka DCL4 i DCL2, natomiast z wirusów DNA – DCL3 [75].

Tabela 1.1. Porównanie funkcji roślinnych białek typu Dicer (DCL) i ich produktów.

Białko	DCL1	DCL2	DCL3	DCL4
Długość generowanych produktów	21 nt	22 nt	24 nt	21 nt
Główny produkt endogeny	miRNA	NAT-siRNA	hc-siRNA (zależne od Pol IV i RDR2)	ta-siRNA (zależne od Pol II i RDR6)
Rola w produkcji siRNA pochodzenia wirusowego	nieznaczna	uzupełnia DCL4 w generowaniu vsiRNA wirusów RNA	główne białko generujące vsiRNA wirusów DNA	główne białko generujące vsiRNA wirusów RNA

Na przykładzie biogenezy vsiRNA wykazano wymiennosc funkcji oraz hierarchie działania poszczególnych roślinnych białek DCL. DCL4 wraz z kofaktorem DRB4 odpowiadają za produkcję siRNA (długości 21 nukleotydów) pochodzących z większości wirusów roślinnych. Stąd aktywność obu białek podlega często inhibicji przez wirusowe supresory. Stwierdzono, że w takiej sytuacji białko DCL2 produkuje z tych samych prekursorów alternatywną pulę 22-nukleotydowych vsiRNA,

wykazujących działanie zbliżone do produktów DCL4 [76]. Podobną zależność zaobserwowano w przypadku prekursorów ta-siRNA, których cięcie przy braku aktywności DCL4 zachodzi z udziałem DCL2 i DCL3 [77].

Białka DCL2, DCL3 i DCL4, podobnie jak DCL1, wykazują lokalizację jądrową, a ich istotna frakcja kolokalizuje wraz z DCL1 w charakterystycznych ziarnistościach w okolicy jąderka. Struktury te, poprzez skupienie szlaków biogenezy miRNA i różnych rodzajów siRNA wpływać mogą na obserwowaną wymiennosc funkcji poszczególnych białek DCL [78]. Jedynie białka DCL1 i DCL3 posiadają kanoniczny sygnał lokalizacji jądrowej [79]. Sugeruje to, że DCL2 i DCL4 mogą być także obecne w cytoplazmie, co w przypadku DCL2 zostało potwierdzone doświadczalnie [64]. Obecność obu białek w cytoplazmie wydaje się uzasadniona ze względu na fakt, iż materiał genetyczny roślinnych wirusów RNA z reguły nie trafia do jądra komórkowego.

Co najmniej kilka białek wpływa na stabilizację prekursorów i dojrzałych cząsteczek miRNA i siRNA oraz prawidłowy przebieg procesu ich biogenezy w komórkach roślinnych. Białko DAWDLE (DDL), które może oddziaływać między innymi z DCL1, stabilizuje pierwotne prekursory miRNA i siRNA przed etapem ich cięcia [80]. Inne roślinne białko, TOUGH (TGH), wpływa na wydajność cięcia prekursorów przez białka typu Dicer, prawdopodobnie wspomagając wiązanie cząsteczek RNA z enzymem [81]. Czapeczka metyloguanozynowa obecna w prekursorach generowanych przez Pol II wiązana jest przez kompleks białek CBP20 i CBP80 (ang. *Cap binding protein*), co warunkuje prawidłowy przebieg składania pri-miRNA [82].

Produkty generowane przez białka DCL podlegają 2'-O-metylacji na końcu 3' przez jądrową metylotransferazę HEN1 [83]. Modyfikacja ta bezpośrednio zapobiega trawieniu miRNA i siRNA przez komórkowe 3'-5' egzonukleazy. Metylacja końca 3' chroni także cząsteczki przed urydylacją, która zapoczątkowuje alternatywny szlak degradacji RNA w komórce [84]. Poza metylacją na stabilizację małych regulatorowych RNA wpływa szereg czynników, działających zarówno *in cis* jak i *in trans* [85]. Stabilność poszczególnych miRNA może być uwarunkowana obecnością określonych motywów sekwencyjnych [86, 87]. Wykazano także, że większa część puli dojrzałych miRNA w komórce związana jest z białkowymi kompleksami efektorowymi (zob. Rozdz. I.2.3), co zapewnia ich stabilizację i wydłuża czas półtrwania [88, 89, 90]. Dojrzałe cząsteczki małych regulatorowych RNA podlegają eksportowi do cytoplazmy,

za co odpowiada prawdopodobnie białko HASTY, homolog zwierzęcej Eksportyny 5 [91].

I.2.3. ROŚLINNE KOMPLEKSY WYCISZAJĄCE INDUKOWANE PRZEZ RNA

Małe regulatorowe RNA funkcjonują w komórkach w kompleksach białkowych znanych jako RISC (kompleks wyciszający indukowany przez RNA, ang. *RNA induced silencing complex*). Główny składnik RISC stanowią białka AGO, które bezpośrednio wiążą dojrzałe cząsteczki miRNA bądź siRNA. Cząsteczka małego regulatorowego RNA w kompleksie z AGO służy jako sonda zapewniająca specyficzne wiązanie RISC do docelowego transkryptu. Różnorodne komórkowe transkrypty podlegają w ten sposób specyficznej regulacji, której efektem jest wyciszenie loci z których pochodzą.

Białka AGO występują zarówno u organizmów prokariotycznych jak i eukariotycznych i cechują się wysoką konserwatywnością [92]. Posiadają one charakterystyczną budowę domenową obejmującą: domenę końca N, domenę PAZ (ang. *Piwi/Argonaute/Zwill*, obecna także w białkach Dicer), domenę MID (ang. *middle*) i domenę PIWI. Domeny PAZ i MID wiążą odpowiednio koniec 3' i 5' dojrzałej cząsteczki małego regulatorowego RNA [93, 94]. PIWI przyjmuje strukturę charakterystyczną dla domen rybonukleazy H i odpowiada za aktywność endonukleazową białek AGO [95, 96].

Z generowanych przez białka DCL dupleksów RNA z reguły tylko jedna nić stanowi cząsteczkę regulatorową i zostaje włączona do RISC. Zaobserwowano, iż w przypadku dupleksów miRNA/miRNA* cząsteczka regulatorowa ulega z reguły dużo wyższej akumulacji w komórce niż nić towarzysząca. Obserwacja ta wskazuje na preferencyjne włączanie miRNA do RISC i związaną z tym stabilizację cząsteczki [97]. U roślin, podobnie jak u zwierząt, wybór danej nici RNA często podyktowany jest niższą stabilnością termodynamiczną dupleksu w obrębie końca 5' cząsteczki regulatorowej, w odniesieniu do przeciwnego końca dupleksu [98]. Ponadto wykazano, że na wybór nici z dupleksu miRNA/miRNA* wpływa obecność białka HYL1 (kofaktora DCL1) i jego regulatora CPL1, a także białka szoku cieplnego HSP90 [99, 100]. W niektórych przypadkach także aktywność endonukleazowa białka AGO warunkuje usunięcie nici towarzyszącej z dupleksu [101].

Genom *Arabidopsis* koduje 10 białek Argonaute (AGO1-10), które wykazują zróżnicowane preferencje w wiązaniu poszczególnych klas małych regulatorowych RNA [102]. Przykładowo AGO1 tworzy kompleksy głównie z miRNA, a AGO4, AGO6 i AGO9 z hc-siRNA [103, 104]. Z drugiej strony wiadomo, że białka AGO konkurują o wiązanie poszczególnych małych regulatorowych RNA, a na preferencje wiązania wpływać może między innymi zróżnicowany profil ekspresji poszczególnych genów *AGO* w różnych tkankach rośliny [105]. Ponadto w przypadku niektórych białek AGO wykazano preferencyjne wiązanie małych RNA posiadających wybrany nukleotyd na końcu 5': urydynę (AGO1), adenozyne (AGO2 i AGO4) czy cytydynę (AGO5) [106, 107]. Urydyna obecna jest na końcu 5' większości roślinnych miRNA. Jej zamiana na adenozyne skutkuje zmianą preferencji wiązania małego RNA do białek AGO (z AGO1 na AGO2) i zniesieniem efektu wyciszenia sekwencji docelowej [107]. MiRNA rzadko tworzą homogenne populacje, częściej stanowią zbiór cząsteczek różniących się długością w zakresie 1-2 nukleotydów, co wynikać może zarówno z alternatywnego cięcia prekursorów przez różne białka DCL, jak i nieprecyzyjnego działania poszczególnych białek [108]. W rezultacie różne warianty danego miRNA działać mogą w kompleksach z różnymi białkami AGO. W specyficznym przypadku włączenie alternatywnej formy miRNA do białka AGO4 (typowo odpowiedzialnego za wyciszenie sekwencji powtórzonych w genomie z udziałem hc-siRNA) prowadzi do wyciszenia genu tego miRNA [72].

Białka AGO uczestniczą w procesie tzw. potranskrypcyjnego wyciszenia genów (PTGS, ang. *post-transcriptional gene silencing*). Ich działanie polega na rozcięciu transkryptów podlegających regulacji, co prowadzi do degradacji powstałych fragmentów z udziałem komórkowych rybonukleaz [109, 110]. Jak dotąd u *Arabidopsis* potwierdzono aktywność endonukleazową białek AGO1, AGO2, AGO4, AGO7 i AGO10 [106, 107, 111, 112, 113, 114]. Aktywność ta nie jest jednak warunkiem koniecznym dla wyciszenia danego locus w genomie. Alternatywny mechanizm PTGS stanowi inhibicja translacji docelowych mRNA. Sposób, w jaki dochodzi do inhibicji nie jest do końca jasny, ale wydaje się zależeć od regionu, w którym do docelowego transkryptu przyłączony zostaje RISC. Zaobserwowano między innymi zaburzenie inicjacji translacji oraz jej przedwczesną terminację [115].

U *Arabidopsis* większość miRNA i wiele siRNA działa w kompleksach z białkiem AGO1. Analizując mutanty *AGO1* zaobserwowano, że aktywność endonukleazowa białka warunkuje prawidłowe funkcjonowanie roślinnych miRNA i

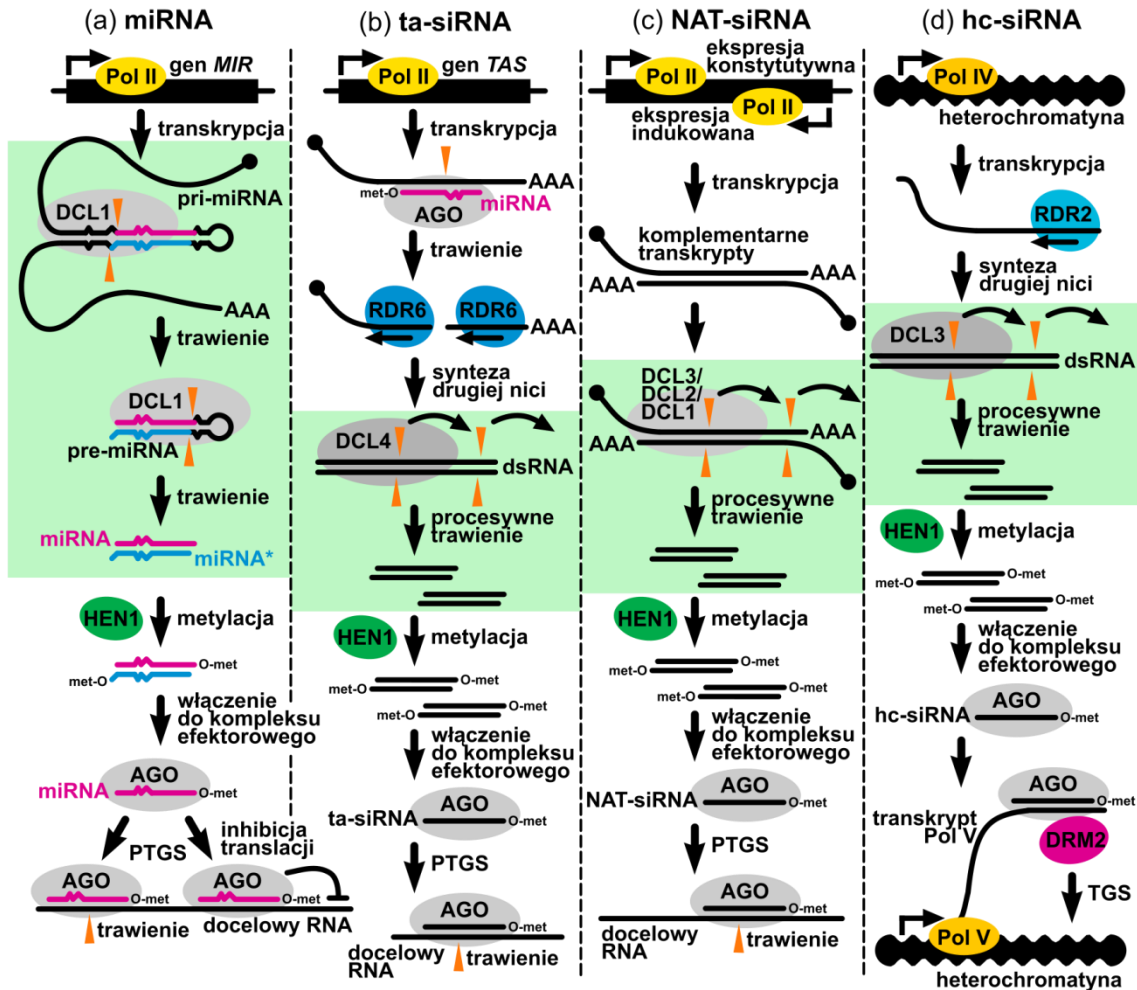
siRNA, a co za tym idzie prawidłowy rozwój roślin [101, 102]. Z drugiej strony wykazano, że dla wielu roślinnych miRNA i siRNA działających w kompleksach z AGO1 inhibicja translacji ma znaczący wpływ na obserwowany efekt wyciszenia danego locus [115, 116]. Wskazuje to, że oba mechanizmy działania RISC: trawienie docelowych transkryptów jak i inhibicja ich translacji, mają istotne znaczenie w procesie PTGS u roślin. Udział obu mechanizmów w wyciszaniu poszczególnych loci jest zróżnicowany. Jak dotąd nie wskazano także żadnej właściwości docelowego transkryptu (na przykład liczba miejsc wiązania miRNA, ich lokalizacja w obrębie cząsteczki, stopień komplementarności miRNA/transkrypt), która faworyzowałaby bądź jego trawienie, bądź inhibicję translacji [116].

Sekwencje roślinnych miRNA i docelowych transkryptów wykazują bardzo wysoki stopień komplementarności [117]. Z kolei u zwierząt komplementarność ta ogranicza się najczęściej do pierwszych 8 nukleotydów od końca 5' miRNA [118]. Wydaje się, że ta obserwacja odzwierciedla istotną różnicę w sposobie funkcjonowania roślinnych i zwierzęcych miRNA, która wpływa na skalę regulacji u obu grup organizmów. Zakładając wiązanie miRNA do transkryptu w obrębie jedynie pierwszych 8 nukleotydów cząsteczki regulatorowej dla każdej konserwatywnej rodziny zwierzęcych miRNA zidentyfikowano średnio około 500 potencjalnych docelowych transkryptów. Szacuje się, że u zwierząt ponad połowa genów kodujących białka podlega regulacji przez miRNA [119]. Ponadto zwierzęce transkrypty często podlegają regulacji przez więcej niż jeden miRNA, co wskazuje, że działanie zwierzęcych miRNA ma charakter kombinatoryczny [120]. U roślin potencjalne miejsca wiązania konserwatywnych miRNA zidentyfikowano w mniej niż 1% komórkowych transkryptów [121]. Ponadto u roślin jeden transkrypt podlega z reguły regulacji przez jeden określony miRNA [108, 122].

AGO4 oraz jego najbliższe homologi AGO6 i AGO9 odpowiadają za funkcjonowanie hc-siRNA u *Arabidopsis*. Wiadomo, że kompleksy hc-siRNA i specyficznych dla nich białek AGO biorą udział w procesie metylacji DNA kierowanej przez RNA (RdDM, ang. *RNA-directed DNA methylation*). Proces ten prowadzi do modyfikacji struktury chromatyny i stanowi formę transkrypcyjnego wyciszenia genów (TGS, ang. *transcriptional gene silencing*). Wiele transkryptów, których poziom regulowany jest poprzez TGS pochodzi z transpozonów i innych sekwencji powtórzonych w genomie. Szczegóły funkcjonowania hc-siRNA pozostają jednak w wielu aspektach niejasne [123]. Za biogenezę hc-siRNA oraz ich docelowych

transkryptów odpowiadają dwie specyficzne dla roślin polimerazy RNA: odpowiednio Pol IV i Pol V [124]. Obie polimerazy generują niekodujące RNA, z których wiele pochodzi z obszarów genomu podlegających metylacji, do których są rekrutowane za pośrednictwem specyficznych kofaktorów [125, 126, 127, 128].

Jak zaobserwowano białka AGO w kompleksie z hc-siRNA wiążą się z transkryptami Pol V w miejscu ich powstawania. W ten sposób białka AGO kierowane są do sekwencji podlegających RdDM. Zakłada się, że proces ten możliwy jest zarówno dzięki komplementarności pomiędzy cząsteczkami hc-siRNA, a docelowymi transkryptami, jak i dzięki oddziaływaniom pomiędzy białkami zaangażowanymi w TGS: w tym AGO4 z Pol V. AGO4 może także oddziaływać z komórkową metylotransferazą DRM2, co prawdopodobnie prowadzi do metylacji loci, z których pochodzą transkrypty generowane przez Pol V [129, 130]. Nie wiadomo, jak duże znaczenie dla tego procesu ma aktywność endonukleazowa AGO4. Wykazano, że jest ona niezbędna dla włączenia do AGO4 właściwej nici hc-siRNA z dupletu generowanego przez DCL3. Nie jest natomiast jasne czy następuje także trawienie docelowych transkryptów [131].



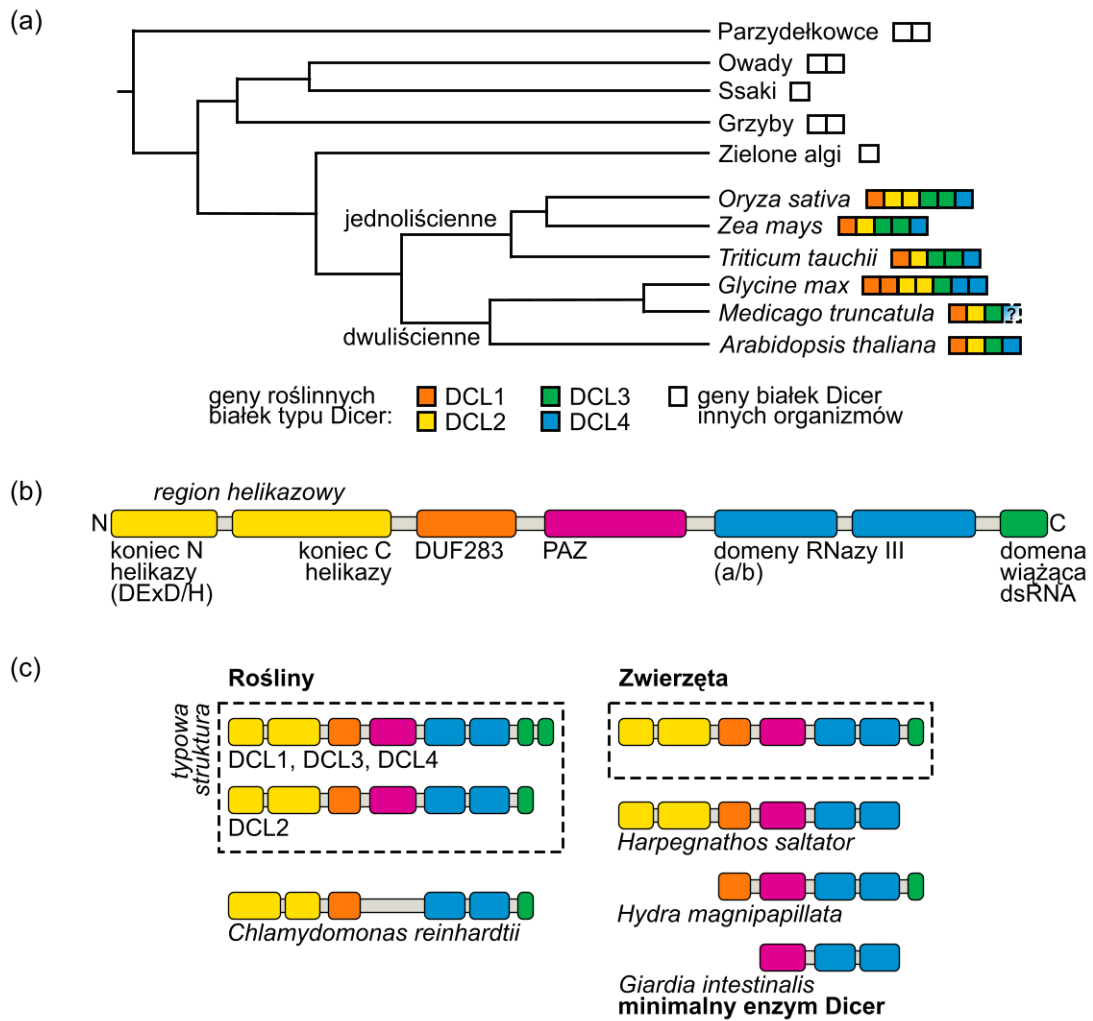
Rysunek 1.3. Szlaki biogenezy małych regulatorowych RNA u roślin. (a) Pierwotne prekursorsy miRNA (pri-miRNA) syntezowane są przez Pol II. Po obróbce potranskrypcyjnej pri-miRNA ulegają trawieniu przez DCL1, co prowadzi do wyzwolenia dupletu miRNA/miRNA*. Dojrzałe cząsteczki miRNA funkcjonują w kompleksach z białkami AGO, jako sondy zapewniające specyficzne wiązanie do transkryptów. Białka AGO odpowiadają za degradację docelowych transkryptów bądź inhibicję ich translacji. (b) Cząsteczki ta-siRNA powstają z transkryptów genów *TAS*, które uległy trawieniu przez białka AGO. Trawieniem tym kierują specyficzne miRNA. Niektóre fragmenty transkryptów *TAS* stanowią matrycę dla RDR6, a powstały dsRNA są substratem dla DCL4. Działanie białek AGO w kompleksach z ta-siRNA opiera się głównie na trawieniu docelowych transkryptów. (c) Prekursory NAT-siRNA powstają z pary wzajemnie komplementarnych transkryptów. Z reguły jeden z nich jest stale obecny w komórce, a produkcję drugiego indukują określone czynniki np. stresowe. Utworzony dsRNA stanowi substrat dla białek DCL1, DCL2 bądź DCL3. Działanie białek AGO w kompleksach z NAT-siRNA także opiera się głównie na trawieniu docelowych transkryptów. (d) Prekursory hc-siRNA stanowią transkrypty Pol IV, syntezowane z reguły na matrycy różnorodnych sekwencji powtórzonych w genomie. DsRNA powstają przez syntezę drugiej nici z udziałem RDR2 i stanowią substrat dla DCL3. Sekwencje docelowe dla hc-siRNA w kompleksach z białkami AGO stanowią długie niekodujące transkrypty Pol V, także powstające z genomowych sekwencji powtórzonych. Hc-siRNA kierują białka modyfikujące strukturę chromaty do powtarzających się w genomie sekwencji. Białko HEN1 metyluje końce 3' wszystkich typów małych regulatorowych RNA, co zapobiega ich degradacji. Etapy biogenezy z udziałem białek DCL zaznaczono na zielono.

I.3. BIAŁKA TYPU DICER

Białka Dicer wykazują aktywność rybonukleazową, a ich działanie polega na wycięciu z prekursora krótkiego dupletu RNA zawierającego dojrzałą cząsteczkę regulatorową (miRNA bądź siRNA). U kręgowców, nicieni a także niektórych drożdży i alg występuje jedno uniwersalne białko Dicer odpowiedzialne za generowanie całej puli małych regulatorowych RNA. Rośliny, owady, pierwotniaki oraz niektóre grzyby posiadają z reguły kilka homologicznych białek Dicer (Rys. 1.4a). Białka występujące u roślin znane są jako DCL. Występowanie kilku białek Dicer w organizmie wiąże się z reguły z ich funkcjonalną specjalizacją. Przykładowo u *Drosophila melanogaster* występują dwa homologiczne białka: Dicer-1 i Dicer-2, wyspecjalizowane w generowaniu odpowiednio miRNA i siRNA [132].

Genom *Arabidopsis thaliana* koduje 4 białka typu Dicer (DCL1-4), reprezentujące 4 wyspecjalizowane funkcjonalnie grupy białek DCL występujących u roślin. Dodatkowo w wielu gatunkach roślin zidentyfikowano duplikacje genów kodujących poszczególne białka DCL. Spośród jednoliściennych, ryż (*Oryza sativa*) posiada 6 białek DCL (w tym po 2 izoformy DCL2 i DCL3), natomiast kukurydza (*Zea mays*) 5 białek DCL (2 izoformy DCL3) [133, 134]. U dwuliściennej soi (*Glycine max*) występuje 7 białek DCL, w tym po 2 izoformy DCL1, DCL2 i DCL4 [135, 136]. W genomie lucerny (*Medicago truncatula*) udało się dotąd zidentyfikować jedynie 3 geny kodujące białka DCL (DCL1, DCL2 i DCL3) [137, 138]. Projekt sekwencjonowania genomu *Medicago* nie został jednak jeszcze ukończony.

Jak dotąd jedynie u roślin zaobserwowano występowanie tak licznej grupy homologicznych białek DCL. Analizy filogenetyczne wskazują, że pojawienie się czterech białek DCL miało miejsce na wczesnych etapach ewolucji roślin i zbiegło się w czasie z wykształceniem wielokomórkowości. Sugeruje to, że zróżnicowanie szlaków biosyntezy małych regulatorowych RNA związane było ze wzrostem złożoności organizmów. Ponadto rośliny wykorzystują małe regulatorowe RNA jako jeden z głównych systemów obrony przed wirusami i innymi patogenami. U kręgowców, które posiadają tylko jedno białko Dicer, rolę tę pełni w dużej mierze specjalnie wykształcony układ immunologiczny. Stąd prawdopodobnie liczba obecnych w roślinach białek DCL odzwierciedla także ich istotną rolę w tym procesie [133, 139].



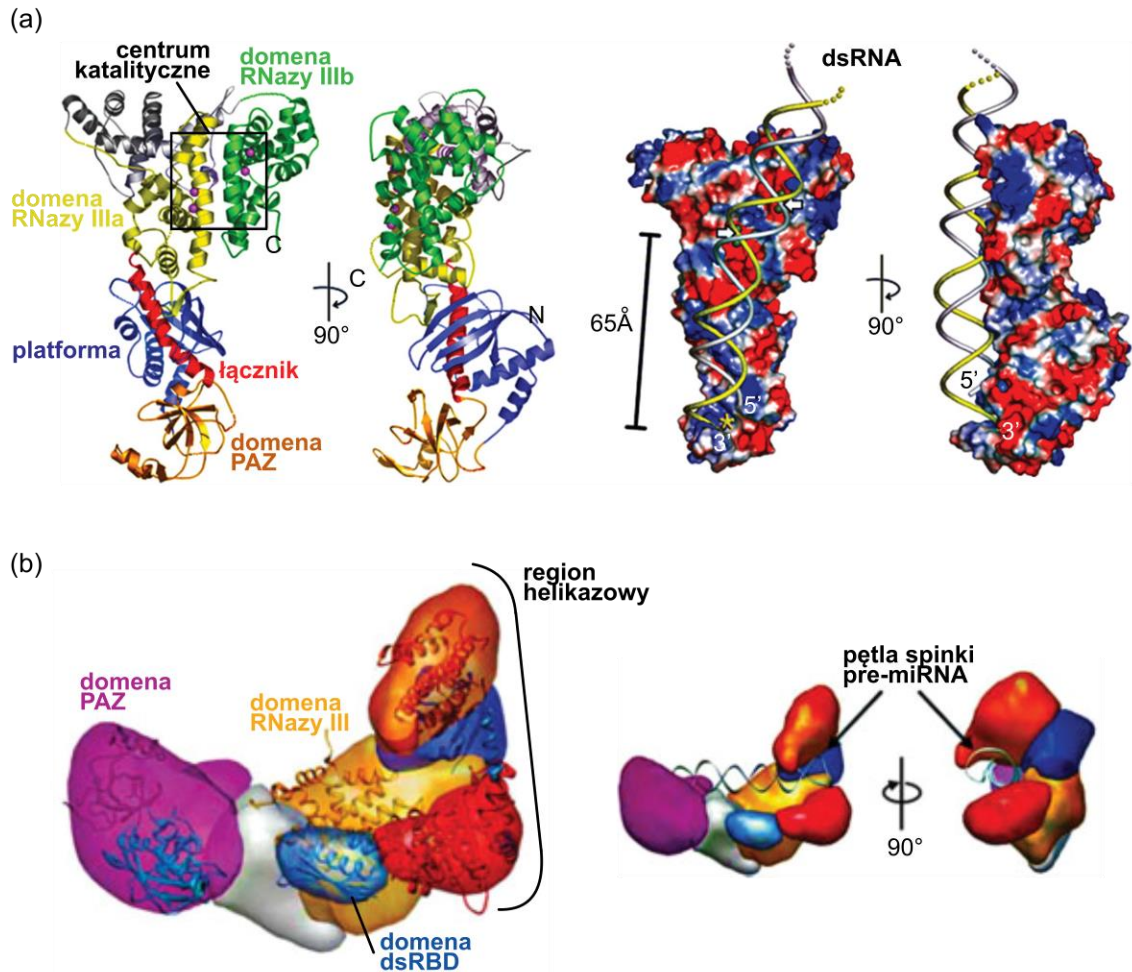
Rysunek 1.4. Białka Dicer i DCL. (a) Schemat ewolucji genów kodujących białka Dicer i DCL. Liczba genów *DCL* zidentyfikowana u wybranych gatunków roślin jedno- i dwuliściennych w porównaniu z liczbą genów białek Dicer charakterystyczną dla innych taksonów. W przypadku *Medicago truncatula* obecność genu kodującego przynajmniej jedną izoformę *DCL4* nie została dotąd potwierdzona (na podstawie [133]). (b) Kompozycja domenowa białek typu Dicer i DCL (bez zachowania skali). Od końca aminowego są to kolejno: domena DEXD/H i domena końca C helikazy (tworzą razem region helikazowy, kolor żółty), DUF283 (pomarańczowy), PAZ (różowy), dwie domeny RNazy III: a i b (niebieski) oraz domena wiążąca dsRNA (zielony). (c) Różnice pomiędzy typową kompozycją domenową zwierzęcych białek Dicer i roślinnych DCL (w ramkach). Charakterystyczną cechą roślinnych białek *DCL1*, *DCL3* i *DCL4* jest obecność dwóch domen wiążących dsRNA na końcu C. Poniżej przedstawiono przykłady znanych wariantów białek Dicer i DCL. Białko z pierwotniaka *Giardia intestinalis* zbudowane jest jedynie z dwóch typów domen (PAZ i RNazy III), przez co uznawane jest za minimalny enzym Dicer.

I.3.1. BUDOWA BIAŁEK TYPU DICER

Większość białek z rodziny Dicer występujących u kręgowców, owadów i roślin posiada w swej strukturze 6 typów domen (Rys. 1.4b). Od końca aminowego białka są to kolejno domena obejmująca motyw DExD/H (ang. *DExD/H box*) i domena końca C helikazy (ang. *helicase C*), tworzące razem region helikazowy, a następnie domeny DUF283 (ang. *domain of unknown function*), PAZ, RNazy III oraz domena wiążąca dsRNA (dsRBD, ang. *dsRNA binding domain*). Pełnej długości ludzkie białko Dicer ma masę cząsteczkową około 220 kDa, masa cząsteczkowa białek DCL z *Arabidopsis* (AtDCL) wynosi od 155 do 215 kDa. U niższych organizmów eukariotycznych białka Dicer często pozbawione są jednej bądź kilku z wymienionych domen (Rys. 1.4c). Przykładowo, Dicer z pierwotniaka *Giardia intestinalis*, jak dotąd jedyne białko z rodziny Dicer, dla którego określono strukturę krystaliczną, posiada w swej strukturze jedynie domenę PAZ oraz dwie domeny RNazy III. Wskazuje to na ich kluczową rolę w procesie generowania małych regulatorowych RNA przez białka Dicer [140]. Dane strukturalne dotyczące ludzkiego białka Dicer, uzyskane za pomocą mikroskopii elektronowej, pozwoliły na określenie jego ogólnej topologii i rozmieszczenia poszczególnych domen (Rys. 1.5) [141].

Trzon katalityczny wszystkich białek typu Dicer stanowi wewnątrzcząsteczkowy dimer dwóch domen RNazy III (oznaczanych jako a i b). Każda z domen odpowiada za przecięcie jednej nici dwuniciowego fragmentu prekursora. Aktywność katalityczna RNazy III zależna jest od obecności jonów Mg^{2+} i prowadzi do rozerwania pojedynczego wiązania fosfodiesterowego w obrębie nici RNA. Powstające w ten sposób produkty posiadają grupę fosforanową na końcu 5' oraz grupę hydroksylową na końcu 3' [142]. Rozmieszczenie domen RNazy III w strukturze białka determinuje obserwowany wzór cięcia dwuniciowych prekursorów; jak już wcześniej wspomniano generowane przez enzym pre-miRNA oraz dupлексы RNA (w tym dupлексы miRNA/miRNA*) posiadają charakterystyczne 2-nukleotydowe wystające końce 3'. Wykazano, że tego typu końce mogą być specyficznie wiązane przez domenę PAZ, oraz że wiązanie końca pre-miRNA do domeny PAZ warunkuje powstawanie cząsteczek miRNA o prawidłowej długości [143, 144].

Dane strukturalne dotyczące białek typu Dicer z człowieka oraz *Giardia intestinalis* wskazują, że domena PAZ oraz dimer RNazy III są odseparowane przestrzennie, a dystans pomiędzy nimi odpowiada długości około 25-nukleotydowej



Rysunek 1.5. Struktura białek typu Dicer. (a) Struktura białka Dicer z *Giardia intestinalis*, minimalnego białka typu Dicer (kod PDB: 2FFL). W strukturze krystalicznej (po lewej) domena PAZ (pomarańczowa) połączona jest z domeną RNazy IIIa (żółta) przez α -helikalny łącznik (czerwony). Łącznik stabilizowany jest przez strukturę platformy (niebieska). Domeny RNazy IIIa i RNazy IIIb (zielona) tworzą wewnątrzcząsteczkowy dimer, a ich miejsca katalityczne (w czarnym kwadracie) zlokalizowano za pomocą jonów Er^{3+} (fioletowe). Model kompleksu białka Dicer z *Giardia intestinalis* i fragmentu dsRNA (po prawej) wskazuje, że odległość pomiędzy domeną PAZ (miejsce wiązania końca RNA, żółta gwiazdka), a centrum katalitycznym (miejsca cięcia obu nici RNA, zaznaczone białymi strzałkami) determinuje długość generowanych produktów. Kolorem niebieskim i czerwonym zaznaczono regiony białka o charakterze odpowiednio zasadowym i kwasowym (© 2006 AAAS) [140]. (b) Topologia ludzkiego białka Dicer określona za pomocą mikroskopii elektronowej. W strukturze białka wskazano lokalizację domen PAZ (fioletowa), RNazy III (pomarańczowa), dsRBD (niebieska) oraz regionu helikazowego złożonego z trzech subdomen (pomarańczowa, niebieska, czerwona) tworzących strukturę w kształcie klamry. W obszarze poszczególnych domen dokonano dokowania znanych struktur homologów (po lewej). Model kompleksu ludzkiego białka Dicer i pre-miRNA (po prawej) wskazuje, że prekursor, którego koniec został związany w obrębie domeny PAZ zajmuje centralny kanał widoczny w strukturze enzymu. Jednocześnie pętla spinki prekursora mieści się w strukturze klamry utworzonej przez region helikazowy białka (reprodukcja za zgodą © Macmillan Publishers Ltd) [141].

helisy dsRNA. Sugeruje to, że długość produktów generowanych przez białka typu Dicer determinowana jest przez odległość pomiędzy domeną PAZ, gdzie następuje wiązanie końca prekursora, a centrum katalitycznym enzymu [140, 141]. W przypadku enzymu z *Giardia* powierzchnia łącząca te domeny wykazuje dodatni potencjał elektrostatyczny, który wspomaga wiązanie dsRNA [145]. Badania struktury krystalicznej domeny PAZ ludzkiego białka Dicer wykazały obecność dwóch kieszeni wiążących oba końce nici tworzącej pre-miRNA. Wiązany jest zarówno wystający 2-nukleotydowy koniec 3' RNA jak i grupa fosforanowa końca 5', co ma istotne znaczenie dla wydajności trawienia pre-miRNA przez Dicer [143, 146]. Kieszenie wiążące wystający koniec 3' prekursorów małych regulatorowych RNA w zwierzęcych białkach Dicer oraz roślinnych DCL1, DCL2 i DCL3 posiadają jednolity, ujemny potencjał elektrostatyczny, natomiast kieszenie w białkach DCL4 różnych gatunków roślin wykazują bardzo niejednorodny potencjał. Cecha ta może być związana z przystosowaniem DCL4 do trawienia różnorodnych prekursorów siRNA pochodzących od bardzo zmiennych i szybko ewoluujących wirusów [139].

W obrębie końca aminowego większości białek typu Dicer zlokalizowany jest region helikazowy, klasyfikowany jako helikaza nadrodziny 2. Obejmuje on domenę o charakterystycznym motywie sekwencyjnym DExD/H oraz domenę końca karboksylowego helikazy. Domeny typu DExD/H wiążą i hydrolizują ATP, dzięki czemu możliwe jest rozplatanie dupleksu RNA oraz translokacja substratu [147]. Region helikazowy nie występuje między innymi w białku Dicer z *Giardia*, co wskazuje, że nie jest on niezbędny dla zachowania podstawowej aktywności katalitycznej enzymu [140]. W zwierzęcych białkach Dicer region helikazowy przyjmuje kształt klamry, a jego lokalizacja w pobliżu centrum katalitycznego wskazuje na możliwość oddziaływania z fragmentem prekursora powyżej wycinanego dupleksu (Rys. 1.5) [141]. Wykazano, że region ten może oddziaływać z pętlą spinki pre-miRNA, dzięki czemu prawdopodobnie następuje rozróżnienie krótszych prekursorów miRNA od dłuższych dupleksów [148, 149]. Delecja regionu helikazowego w ludzkim białku Dicer podnosi jego aktywność katalityczną i znosi preferencje cięcia wobec poszczególnych substratów [144, 150].

Rola regionu helikazowego została najlepiej scharakteryzowana w przypadku białek Dicer z *Drosophila melanogaster*. Spośród dwóch białek Dicer u *Drosophila* Dicer-1 nie posiada funkcjonalnego motywu DExD/H w części helikazowej, co współgra z preferencją białka do trawienia stosunkowo krótkich struktur spinki

obecnych w zwierzęcych pre-miRNA. Z kolei region helikazowy Dicer-2 wiąże i hydrolizuje ATP, co umożliwia procesywnie cięcie długich dwuniciowych prekursorów siRNA [151]. Wydajne trawienie prekursorów siRNA przez Dicer-2 wymaga dostępności ATP, nie jest natomiast zależne od obecności 2-nukleotydowego wystającego końca 3'. Sugeruje to, że cięcie tego rodzaju prekursorów nie wymaga silnego wiązania ich końca z domeną PAZ w obrębie kieszeni wiążących koniec pre-miRNA. Odwrotnie, trawienie pre-miRNA w dużym stopniu zależy od struktury jego końca, a nie jest zależne od ATP [152]. Dodatkowo precyzyjne cięcie pre-miRNA wymaga związania pętli spinki prekursora do regionu helikazowego. Razem sugeruje to dwie ważne role regionu helikazowego w białkach typu Dicer: zapewnianie procesywności podczas trawienia długich komplementarnych prekursorów siRNA oraz prawidłowe pozycjonowanie spinek pre-miRNA poprzez oddziaływanie z ich jednoniciową pętlą [153].

Zdecydowana większość roślinnych białek typu Dicer posiada funkcjonalny motyw DExD/H w regionie helikazowym, mimo iż podobnie jak u *Drosophila* poszczególne homologi wyspecjalizowane są w trawieniu prekursorów miRNA (DCL1) bądź siRNA (DCL2-4) [139]. Zachowanie aktywności helikazowej w białkach DCL1 może wiązać się z odmienną strukturą prekursorów roślinnych miRNA, typowo znacznie dłuższych od ich zwierzęcych odpowiedników i wymagających procesywnego działania enzymu. Dla mutantu białka DCL1 z *Arabidopsis* pozbawionego regionu helikazowego zaobserwowano spadek procesywności, a także nieprecyzyjne cięcie tych prekursorów, których dokładność cięcia w niewielkim stopniu zależy od obecności HYL1 [67].

Cechą charakterystyczną, obserwowaną jedynie wśród roślinnych enzymów typu Dicer jest obecność zróżnicowanej liczby dsRBD na końcu karboksylowym białka. DCL1, DCL3 i DCL4 typowo posiadają dwie takie domeny, podczas gdy DCL2, podobnie jak zwierzęce białka Dicer, tylko jedną [133]. W przypadku ludzkiego białka Dicer delecja dsRBD skutkuje jedynie spadkiem wydajności cięcia prekursorów [150]. Tymczasem rośliny *Arabidopsis* produkujące enzym DCL1 pozbawiony obu dsRBD giną na etapie zarodka [154]. U roślin każda z domen wydaje się pełnić odrębną funkcję. Dla białka DCL1 z *Arabidopsis* wykazano, że pierwsza domena (wykazująca większe podobieństwo do zwierzęcych dsRBD) bierze udział w wiązaniu pri-miRNA do enzymu. Druga natomiast uczestniczy w oddziaływaniach z innymi białkami, w tym HYL1, oraz odpowiada za lokalizację DCL1 w obrębie charakterystycznych jądrowych

ziarnistości, w których grupują się różne białka związane z biogenezą miRNA i siRNA [68, 155].

Wciąż najslabiej poznana pozostaje rola DUF283 w roślinnych i zwierzęcych enzymach typu Dicer. W ludzkim białku Dicer DUF283 zlokalizowana jest pomiędzy domeną PAZ, a centrum katalitycznym enzymu, w pobliżu dsRBD [156]. Struktura DUF283 białka DCL4 z *Arabidopsis* określona za pomocą magnetycznego rezonansu jądrowego ujawniła obecność motywu potencjalnie wiążącego dsRNA. Z drugiej strony wykazano możliwość jej oddziaływania z białkiem DRB4, które warunkuje aktywność DCL4 [157].

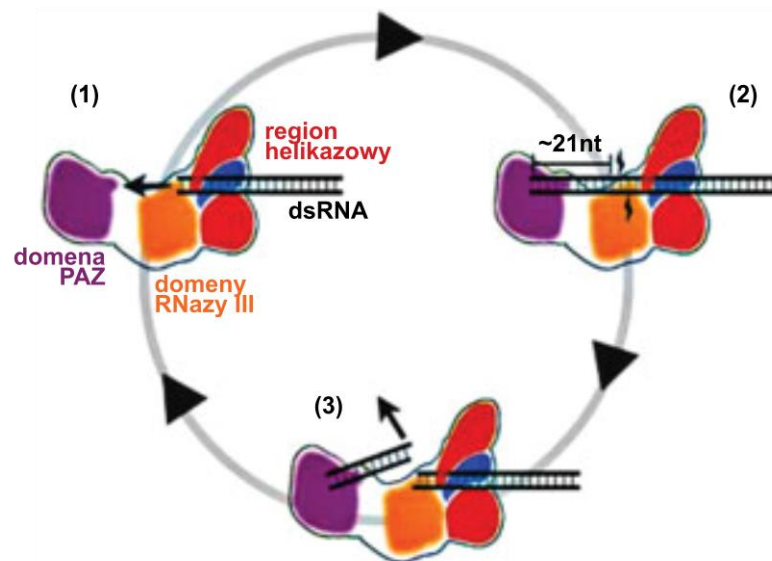
I.3.2. FUNKCJONOWANIE BIAŁEK TYPU DICER

Topologia ludzkiego białka Dicer wraz z danymi dotyczącymi funkcjonalności poszczególnych domen pozwoliły na opracowanie ogólnego mechanizmu działania enzymu (Rys. 1.6). Za wiązanie prekursora odpowiadają domena PAZ oraz region helikazowy, zlokalizowane po przeciwnych stronach białka. Tym samym substrat zajmuje centralny kanał obecny w strukturze enzymu i podlega przecięciu w obrębie jego centrum katalitycznego. Wycięty dupleks RNA zostaje uwolniony z kompleksu, natomiast druga część prekursora może pozostać związana przez region helikazowy białka. Umożliwia to procesywne cięcie długich prekursorów, z których tworzony jest więcej niż jeden produkt [141].

Wciąż najslabiej poznane pozostają szczegóły procesu przyłączenia substratu do białka, szczególnie w przypadku roślinnych białek DCL, które wykazują specjalizację w obróbce poszczególnych typów prekursorów. Porównując aktywność enzymów DCL3 i DCL4 z *Arabidopsis* wykazano pewne różnice w ich funkcjonowaniu. DCL4 trawi preferencyjnie długie substraty, a proces ten zależny jest od obecności ATP. Z kolei DCL3 wykazuje tendencję do preferencyjnego cięcia krótszych dupleksów RNA, które posiadają wystający koniec 3' oraz fosforylowaną adenozyne lub urydynę na końcu 5' [158]. Wpływ na specyficzność substratową mogą mieć także kofaktory DCL, jak na przykład białka wiążące dsRNA: HYL1 i DRB4, specyficznie oddziałujące z odpowiednio DCL1 i DCL4. Wreszcie dostępność poszczególnych białek DCL w różnych tkankach lub typach komórek decydować może o losie danego prekursora. Najbardziej znamienny przykład stanowi proces wyciszania *de novo* transpozonów u *Arabidopsis*. Wykazano, że dsRNA pochodzący z transpozonu

Evadé (EVD) trawiony jest przez enzymy DCL4 i DCL2, a utworzone siRNA kierują procesem potranskrypcyjnego wyciszania EVD. Proces ten nie jest jednak efektywny, a stopniowa akumulacja dsRNA pochodzenia transpozonowego powoduje wysycenie szlaków biogenezy siRNA z udziałem obu enzymów. W konsekwencji białko DCL3 rozpoczyna produkcję transpozonowych hc-siRNA, które wydajnie wyciszają ekspresję transpozonu poprzez RdDM [159].

Przyłączenie substratu umożliwia jego cięcie przez enzym DCL. W szlaku biogenezy miRNA przynajmniej dwa trawienia dokonywane przez DCL1 w obrębie prekursora są niezbędne do wyzwolenia dupletu RNA obejmującego cząsteczkę regulatorową. Z uwagi na długość struktur spinki obecnych w roślinnych prekursorach miRNA często dopiero sekwencja kilku następujących po sobie trawień prowadzi do powstania pre-miRNA: prekursora, z którego bezpośrednio wyzwala się duplet miRNA/miRNA*. Wykazano, że struktury spinki niektórych roślinnych pri-miRNA obejmują więcej niż jeden duplet miRNA/miRNA*. Jak już sygnalizowano



Rysunek 1.6. Mechanizm działania ludzkiego enzymu Dicer, zaproponowany na podstawie topologii białka określonej za pomocą mikroskopii elektronicznej. (1) Prekursor (dsRNA lub pre-miRNA) zostaje przyłączony do białka przy udziale regionu helikazowego. (2) Związanie końca prekursora w obrębie domeny PAZ stabilizuje kompleks i umożliwia trawienie. Długość generowanych produktów determinowana jest przez odległość pomiędzy domeną PAZ, a centrum katalitycznym enzymu. (3) Po cięciu następuje uwolnienie dupletu, pozostały fragment prekursora pozostaje związany w obrębie regionu helikazowego. W ten sposób długie prekursora mogą ulegać procesywnemu cięciu bez odłączenia od enzymu (reprodukcja za zgodą © Macmillan Publishers Ltd) [141].

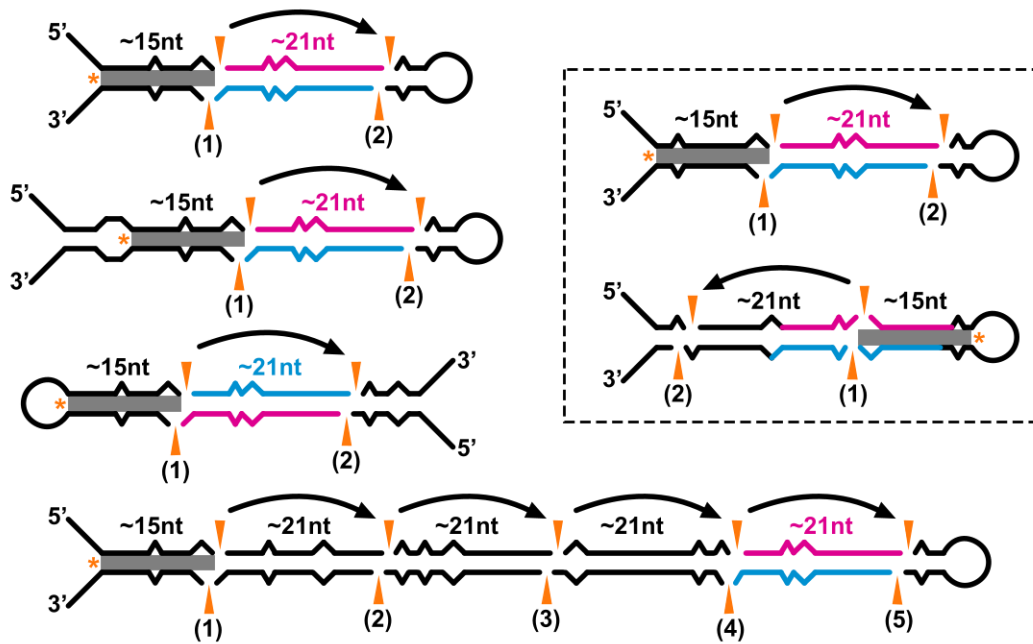
konieczność procesywnego trawienia prekursorów wydaje się być jedną z przyczyn, dla których białka DCL1, wyspecjalizowane podobnie jak Dicer-1 z *Drosophila* w generowaniu miRNA, nie utraciły swej aktywności helikazowej [67, 151].

Białko DCL1 z *Arabidopsis* może rozpoczynać trawienie pri-miRNA zarówno od strony podstawy jak i pętli struktury spinki obejmującej dupleks miRNA/miRNA* (Rys. 1.7) [160, 161]. Wykazano, że niektóre roślinne pri-miRNA podlegają trawieniu w obu kierunkach, co prowadzi do powstawania alternatywnych produktów, a proces ten stanowić może mechanizm regulacji poziomu poszczególnych miRNA w komórce. Typowo miejsce pierwszego trawienia dokonywanego przez DCL1 w obrębie struktury spinki obecnej w pri-miRNA oddalone jest o około 15 nukleotydów od rejonu rozplecenia dsRNA. Prawdopodobnie w tym obszarze struktury spinki następuje przyłączenie białka do pri-miRNA. Zaobserwowano, że obecność rozplecenia warunkuje możliwość trawienia pri-miRNA przez DCL1. Rozplecenie to stanowić może podstawę struktury spinki, być częścią dużej pętli wewnętrznej spinki lub częścią pętli końcowej, jeśli trawienie rozpoczynane jest od tej strony prekursora. Większe wewnętrzne pętle w 15-nukleotydowym fragmencie trzonu poprzedzającym miejsce pierwszego trawienia przeszkadzają w cięciu prekursora [162, 163].

Zaobserwowano, że preferowane miejsce cięcia z reguły także znajduje się w pobliżu rozplecenia dsRNA, powstającego w skutek obecności jednego bądź kilku niesparowań w strukturze spinki prekursora [163]. Trawienie pri-miRNA prowadzi do utworzenia produktu posiadającego 2-nukleotydowy wystający koniec 3'. Koniec ten wiązany jest przez domenę PAZ enzymu. Kolejne cięcia prowadzą do wyzwolenia krótkich dupleksów RNA i następują w odległości około 21 nukleotydów od wygenerowanego wcześniej charakterystycznego końca prekursora. Analiza aktywności białka Dicer z *Giardia* wykazała, że precyzyjne ulokowanie centrum katalitycznego enzymu dokonywane jest głównie w odniesieniu do końca 3' prekursora [145]. Z kolei ludzkie białko Dicer oraz DCL3 z *Arabidopsis* wydają się wyznaczać miejsce trawienia w odniesieniu do fosforylowanego końca 5' prekursora [143, 158]. W przypadku trawienia pre-miRNA przez zwierzęce białka Dicer na lokalizację miejsca cięcia wpływa także odległość od pętli końcowej w strukturze spinki prekursora. Sugeruje to, że wiązanie pętli pre-miRNA przez region helikazowy, podobnie jak wiązanie końca prekursora przez domenę PAZ, umożliwia dokładne pozycjonowanie prekursora w obrębie centrum katalitycznego enzymu i determinuje miejsce trawienia [153].

Co ciekawe u zwierząt wykazano także udział białka Dicer w procesach niezwiązanych z biogenezą małych regulatorowych RNA. Zaobserwowano, że u nicienia *Caenorhabditis elegans* w komórkach ulegających apoptozie białko Dicer podlega trawieniu przez kaspazę CED-3. Cięcie, które ma miejsce w obrębie domeny RNazy IIIa białka prowadzi do powstania dwóch jego fragmentów. Jeden z nich obejmuje fragment RNazy IIIa oraz domeny RNazy IIIb i dsRBD. Zaobserwowano, że tak skrócony enzym pozbawiony jest co prawda aktywności RNazowej, wykazuje natomiast aktywność DNazową pozwalającą na nacinanie jednej nici w dwuniciowych DNA. Ponadto wykazano, że ta aktywność skróconego białka Dicer jest niezbędna dla prawidłowego przebiegu procesu apoptozy u *C. elegans* [164].

Wiedza na temat funkcjonowania białek z rodziny Dicer, szczególnie roślinnego pochodzenia, nadal pozostaje jednak niepełna. Dostępne dane strukturalne ograniczają się obecnie do Dicer z *Giardia intestinalis*, człowieka oraz pojedynczych domen białek z kilku innych organizmów. Wspólny schemat budowy zwierzęcych białek Dicer i roślinnych DCL sugeruje ich zbliżone właściwości biochemiczne. Istnieją jednak wyraźne różnice w funkcjonowaniu obu grup białek. W szlaku biogenezy miRNA roślinne białka DCL łączą w sobie funkcje zwierzęcych rybonukleaz Drosha oraz Dicer. Od białek Dicer człowieka i innych ssaków odróżnia je także daleko posunięta specjalizacja w generowaniu określonego typu małych regulatorowych RNA. Z tego względu pełne zrozumienie roli, jaką białka DCL pełnią w komórkach roślinnych wymaga szczegółowych badań, nieograniczających się jedynie do ich zwierzęcych odpowiedników.



Rysunek 1.7. Sposoby trawienia pri-miRNA przez roślinne białko DCL1. Białko DCL1, w odróżnieniu od zwierzęcych białek Dicer, odpowiada zarówno za trawienie pri-miRNA, jak i pre-miRNA. DCL1 może rozpocząć trawienie prekursora od strony podstawy lub pętli spinki pri-miRNA (po lewej). Pierwsze cięcie (1) następuje w odległości około 15 nukleotydów od miejsca rozplecenia dsRNA (gwiazdka), które znajdują się może u podstawy spinki, w obrębie dużej pętli wewnętrznej lub w obrębie pętli końcowej. Cięcie to prowadzi do powstania pre-miRNA z charakterystycznym 2-nukleotydom wystającym końcem 3'. Kolejne cięcia następują w odległości 21 nukleotydów od miejsca pierwszego cięcia i prowadzą do powstania dupleksów miRNA/miRNA* (różowy/niebieski). Niekiedy do wyzwolenia właściwego dupleksu miRNA/miRNA* wymagana jest sekwencja kilku cięć. Niektóre roślinne pri-miRNA mogą być trawione alternatywnie od strony podstawy lub pętli spinki, co prowadzi do utworzenia z danego prekursora różnych produktów (ramka).

II. CEL PRACY

Podstawowym celem niniejszej pracy doktorskiej była identyfikacja oraz charakterystyka genów kodujących białka typu Dicer u modelowej rośliny bobowatej *Medicago truncatula* oraz wstępna ocena aktywności wybranych białek kodowanych przez te geny.

Zgodnie z przyjętymi na wstępie założeniami osiągnięcie tak zarysowanego celu związane było z realizacją następujących zadań szczegółowych:

- (i) identyfikacja oraz analiza porównawcza genów *DCL* obecnych w genomie *Medicago*;
- (ii) określenie poziomu akumulacji mRNA kodujących poszczególne białka DCL w różnych częściach rośliny;
- (iii) uzyskanie klonów sekwencji komplementarnych do mRNA kodujących białka DCL (cDNA, ang. *complementary DNA*);
- (iv) produkcja wybranych białek DCL – białka zaangażowanego w produkcję miRNA oraz jednego z białek odpowiedzialnych za produkcję siRNA;
- (v) analiza porównawcza aktywności katalitycznej uzyskanych białek.

III. MATERIAŁY I METODY

III.1. MATERIAŁY

III.1.1. OLIGONUKLEOTYDY DNA I RNA

Tabela 3.1. Oligonukleotydy DNA i RNA wykorzystywane w pracy. Kolorem szarym zaznaczono sekwencje dodane do amplifikowanego fragmentu DNA w wyniku zastosowania oligonukleotydu jako startera w PCR.

Oligonukleotydy DNA		
Nazwa	Sekwencja (5`-3`)	Przeznaczenie
Oligo-dT	TTTTTTTTTTTTTTTTTTT	Odwrotna transkrypcja (mieszanina 3 oligonukleotydów w równych proporcjach)
	TTTTTTTTTTTTTTTTTTTC	
	TTTTTTTTTTTTTTTTTTTG	
D1-08-10	TAGAATAGGCGTTGATACACAGCAATAGG	Klonowanie cDNA DCL1
D1-08-13	ACAACCACTGCTTGCTTCTGATTGG	
D2-08-51	CTCACTTCTATTTTCTTAAATTATTACCG	Klonowanie cDNA DCL2
D2-08-19	CCAGGTTCTATCAAGATATTTGATGGTG	
D3-CDNA1	TGAAGCAATTCTTGTCTTCATC	Klonowanie cDNA DCL3
D3-CDNA2	TCGGTTAGTTGGAGGCTAAAA	
D4-CDNA1	ATTTTCCTACGCCGAATTGT	Klonowanie cDNA DCL4
D4-CDNA2	CAATGCCGTACGGTGAATC	
D1-08-12	AATCCTTACTGGCTGGATGCCTG	Sekwencjonowanie cDNA DCL1
D1-08-41	AATGACTTGGTCTTTCGCTCC	
D1-08-42	TTCCAAACAACAGCAGTATTCC	
D1-08-45	GTAAGCTAGATTCAATAGTTTGTACC	
D1-08-46	CTCCTCATCCTTTGTATCCTCATCA	
D1-08-57	TGTTGTGTCTGCATTAGTTCTACC	
D1-pi-61	AATCCGCTTGGCACAACAGG	
D1-pi-62	CCTTCTGTCACCACCTAAAGTTC	
D1-pi-63	AGCATATCTGTTTGTCCCAATGGT	
D1-seq1	TGCTTGTAAGCCAAATCAACA	
D1-seq2	TCTTTTGGTCTTGGAGACTGC	
D1-Hel-seq	CCACAATCTCAGAGGGCATA	
D2-08-17	AGATATGGGTGTTGACTACTGGGAT	
D2-08-21	CTAGAGACTCTATGATGTCTCCAAG	
D2-08-39	ATATACGCTATTCTGAAGAATTAGC	
D2-08-40	GTAACGCCGTCGTTCCCTGG	
D2-08-43	TGCCAGAATGCGAAATTCAGATTACA	
D2-08-44	ATAGGGATTCTAGAAATCTGACATTAAC	

D2-08-47	CATACATCTCCATCCCAGTAGTC	
D2-08-48	AGTGTGATTTCTCAAAGCTTCAGC	
D2-08-59	AACTAGAGTCGAGCTTGATCCA	Sekwencjonowanie cDNA DCL2
D2-08-60	GTCTTTTCGATTGACTATATCCATGACA	
D2-seq	CTTGGAGACATCATAGAGTCTCTAG	
M13-FWD	CAGGAAACAGCTATGAC	Sekwencjonowanie plazmidów pCR
M13-REV	GTAAAACGACGGCCAG	
T7-FWD	TAATACGACTCACTATAGGG	Sekwencjonowanie plazmidów pET
T7-REV	TAGTTATTGCTCAGCGGTGG	
pGEX-FWD	GGGCTGGCAAGCCACGTTTGGTG	Sekwencjonowanie plazmidów pGEX
pGEX-REV	CCGGGAGCTGCATGTGTCAGAGG	
DCL1FL-FW	CACCGAGGATGATGCCGATGCC	Synteza cDNA DCL1 do ligacji z pET-100/D-TOPO
DCL1FL-RE	CTATGACGAGTATAATTTGTTTATTAGTTCC	
DCL2FL-FW	CACCGAAATGGATGACTACGATGAGC	Synteza cDNA DCL2 do ligacji z pET-100/D-TOPO
DCL2FL-RV	TTACTTCCATTGCAAGGAATGC	
DCL83F1	CCGAATTCGATGTAGTGAAAGCTTCAGGATTG	
DCL83R1	GTGAGCCTTGTTCATCGTCGTCCTTGTAGTCTG ACGAGTATAATTTGTTTATTAGTTCC	Synteza cDNA DCL1-CC do ligacji z pGEX-6P-3
DCL83R2	GTAGTCGACTCAGTGTGTTGTGACGGTGGTGT GAGCCTTGTTCATCGTCGT	
FDCL2GST	CCGAATTCGTTGATTGGCTAACAATCAG	Synteza cDNA DCL2-CC do ligacji z pGEX-6P-3
RDCL2GST	CGAGTCGACTTAGTGGTGATGGTGATGATGCT TCCATTGCAAGGAATGC	
qpD1-FW	GCTGCTTTGAAAGAGAAGGAA	Analiza poziomu mRNA DCL1
qpD1-RV	CTCACACACCGGTAGAATGG	
qpD1V2-FW	GGCTACTGGTTCTCTTTTACTGC	Analiza poziomu mRNA DCL1 – wariantu MtDCL1-AS
qpD1V-RV	CATCTTCATGTCCCTCACTGTGT	
qpD2-FW	GATGCTACCAGCGACTTGAA	Analiza poziomu mRNA DCL2a
qpD2-RV	TATTCACCGATGCAGACCTC	
qpD2T2-FW	GATGCTTGTCCAGACAATCTAATC	Analiza poziomu mRNA DCL2b
qpD2T2-RV	AAGTCGCTGATAGCATCTTGG	
qpD2T3-FW	ATCAATTCTTGTCCAGACAATCTTC	Analiza poziomu mRNA DCL2c
qpD2T3-RV	TCAAGTCTCTGATAGCATCTAGGC	
qpD3A-FW	CAGTAGCAATTAAAAAGCCAAAT	Analiza poziomu mRNA DCL3
qpD3A-RV	TGATGAAACAACCTGGAATAGAACC	
qpD4C-FW	AGTAGTTGTGAAGCAACTTTAGCAG	Analiza poziomu mRNA DCL4
qpD4C-RV	ATTTGGAATTGGAGCTTCTGA	
qpACT19	TTCTCTCAGTACTTTCCAGC	Analiza poziomu mRNA aktyny
qpACT20	AAGCATCACAATCACTCC	
T7-TR	AATTTAATACGACTCACTATAGG	Starter do transkrypcji <i>in vitro</i>

Mtr-399-C	CAGGGCAAATCTCCTTTGGCAAACCAGAAACA TTCCTGTTTAAGATTTTTACTCCTCTCTGCCA AAAGAGATTTTCCCTATAGTGAGTCGTATTAA ATT	Matryca do uzyskania pre-miRNA metodą transkrypcji <i>in vitro</i>
Mtr-399-O	CAGGGCAGCTCTCCTTTGGCAAGTGAATTTCT AGAGCCTCTAAGTCAAATGAAATATCACCTGC CAATAGAGACATGCCCTATAGTGAGTCGTATT AAATT	Matryca do uzyskania pre-miRNA metodą transkrypcji <i>in vitro</i>
Oligonukleotydy RNA		
Hsa-33a	GUGCAUUGUAGUUGCAUUGCAUGUUCUGGUGG UACCCAUGCAAUGUUUCCACAGUGCAUC	Substraty do analizy aktywności rybonukleazowej białek
Sens-33a	GUGCAUUGUAGUUGCAUUGCAUGUUCUGGUGA	
Anty-33a	UGGUCUUGUACGUUACGUUGAUGUUACGUGAU	

III.1.2. PLAZMIDY BAKTERYJNE

- pCR-XL-TOPO – plazmid do klonowania DNA zawarty w zestawie TOPO XL PCR Cloning Kit (Life Technologies)
- pET100/D-TOPO – plazmid do produkcji białek w systemie bakteryjnym zawarty w zestawie Champion pET Directional TOPO Expression Kit (Life Technologies)
- pGEX-6P-3 – plazmid do produkcji białek w systemie bakteryjnym (GE Healthcare)

Tabela 3.2. Charakterystyka wykorzystywanych plazmidów bakteryjnych.

	pCR-XL-TOPO	pET100/D-TOPO	pGEX-6P-3
Przeznaczenie	Klonowanie cDNA	Produkcja białek	Produkcja białek
Wielkość	3,5 kpz	5,7 kpz	4,9 kpz
Antybiotyk selekcyjny	kanamycyna	ampicylina	ampicylina
Enzymy stosowane do klonowania	Topoizomeraza I (kowalencyjnie związana z wektorem)	Topoizomeraza I (kowalencyjnie związana z wektorem)	Enzymy restrykcyjne, Ligaza
Wymagania dot. cDNA	Pojedyncze adenozyne na obu końcach 5' cDNA	Sekwencja CACC na jednym z końców cDNA	Miejsca restrykcyjne na obu końcach cDNA
Promotor genu heterologicznego	brak	T7	tac
Charakterystyka	Plazmid wysokokopijny Umożliwia klonowanie długich produktów PCR	Plazmid niskokopijny Produkowane białko posiada ogon histydynowy (6xH) i znacznik Xpress/FLAG (DLYDDDDK) na końcu aminowym	Plazmid niskokopijny Produkowane jest białko fuzyjne z GST na końcu aminowym

III.1.3. SZCZEPY BAKTERII

- *Sinorhizobium meliloti* Rm1021
- *Escherichia coli* TOP10 (Life Technologies)
- *Escherichia coli* BL21-STAR (Life Technologies)

Tabela 3.3. Charakterystyka wykorzystywanych szczepów bakterii.

	<i>S. meliloti</i> Rm1021	<i>E. coli</i> TOP10	<i>E. coli</i> BL21-STAR
Przeznaczenie	Otrzymywanie brodawek korzeniowych <i>M. truncatula</i>	Amplifikacja plazmidowego DNA	Produkcja białek
Metoda transformacji	<i>Nie dotyczy</i>	Szok cieplny	Szok cieplny
Charakterystyka	Szczep wchodzący w symbiozę z <i>M. truncatula</i> . Indukuje powstawanie brodawek korzeniowych.	Szczep do wydajnej transformacji i amplifikacji plazmidowego DNA. Nie posiada żadnych elementów związanych z ekspresją sekwencji zawartych w plazmidzie.	Posiada gen Polimerazy RNA faga T7. Umożliwia ekspresję genów pod kontrolą promotora z faga T7 jak i promotorów bakteryjnych.

III.1.4. POŻYWKI MIKROBIOLOGICZNE

Tabela 3.4. Skład stosowanych podłoży mikrobiologicznych.

Nazwa pożywki	Skład
2YT płynna	1,6% trypton, 1% wyciąg drożdżowy, 0,5% NaCl sterylizowano przez autoklawowanie dodano streptomycynę do końcowego stężenia 50 µg/ml
LB stała	1,5% agar, 1% trypton, 1% NaCl, 0,5% wyciąg drożdżowy sterylizowano przez autoklawowanie dodano ampicylinę lub kanamycynę do końcowego stężenia 100 µg/ml
LB płynna	1% trypton, 1% NaCl, 0,5% wyciąg drożdżowy sterylizowano przez autoklawowanie dodano ampicylinę lub kanamycynę do końcowego stężenia 100 µg/ml
SOC płynna	2% trypton, 0,5% wyciąg drożdżowy, 10 mM NaCl, 2,5 mM KCl, 10 mM MgSO ₄ (sterylizowano przez autoklawowanie 10 mM MgCl ₂ , 20 mM glukoza (sterylizowano przez filtrowanie)

III.1.5. ROZTWORY BUFOROWE

Tabela 3.5. Skład stosowanych roztworów buforowych.

Nazwa buforu	pH	Skład
TBE		89 mM Tris, 89 mM H ₃ BO ₃ , 2 mM EDTA
SB	6,8	250 mM Tris-HCl, 10% SDS, 50% (w/o) glicerol, 0,1% błękit bromofenolowy, 10% 2-merkaptoetanol
TGS	8,3	25 mM Tris-HCl, 192 mM glicyna, 0,1% SDS
TBST	8,0	50 mM Tris-HCl, 138 mM NaCl, 2,7 mM KCl, 0,05% Tween 20
PBS	7,4	10 mM Na ₂ HPO ₄ , 1,8 mM KH ₂ PO ₄ , 137 mM NaCl, 2,7 mM KCl

III.1.6. ZESTAWY ODCZYNNIKÓW

- AP Detection Reagent Kit (Millipore) do prowadzenia reakcji barwnej z alkaliczną fosfatazą
- RNeasy Plant Mini Kit (Qiagen) do izolacji całkowitego RNA z materiału roślinnego
- TOPO XL PCR Cloning Kit (Life Technologies) do klonowania DNA w plazmidzie bakteryjnym pCR-XL-TOPO
- Champion pET Directional TOPO Expression Kit (Life Technologies) do przygotowania wektorów (pochodne plazmidu pET100/D-TOPO) do produkcji białek w systemie bakteryjnym
- BigDye Terminator v1.1 Cycle Sequencing Kit (Life Technologies) do sekwencjonowania DNA
- QIAquick PCR Purification Kit (Qiagen) do oczyszczania DNA
- QIAquick Gel Extraction Kit (Qiagen) do oczyszczania DNA z żelu agarozowego
- QIAprep Spin Miniprep Kit (Qiagen) do izolacji plazmidów z bakterii w małej skali
- MEGAshortscript T7 (Life Technologies) do transkrypcji *in vitro*

III.2. METODY

III.2.1. UNIWERSALNE PROCEDURY LABORATORYJNE

III.2.1.1. Rozdział elektroforetyczny kwasów nukleinowych w warunkach natywnych

Rozdział elektroforetyczny kwasów nukleinowych w warunkach natywnych prowadzono w żelu agarozowym (skład żelu podano w Tabeli 3.6) w buforze 0,5× TBE. Podczas rozdziału preparatów stosowano następujące wzorce długości DNA: GeneRuler 1 kb (Thermo Scientific) lub Perfect Plus 1 kb (EurX).

Rozdzielone cząsteczki kwasów nukleinowych uwidoczniano w świetle UV. Obraz zapisywano przy pomocy systemu do dokumentacji żeli GelLogic 200 (Kodak).

Tabela 3.6. Skład żelu agarozowego do rozdziału elektroforetycznego kwasów nukleinowych w warunkach natywnych.

Żel agarozowy 1%	
Bufor TBE	0,5×
Agaroz	1%
Midori Green (Nippon Genetics)	1 μl/50 ml żelu

III.2.1.2. Rozdział elektroforetyczny kwasów nukleinowych w warunkach denaturujących

Rozdział elektroforetyczny kwasów nukleinowych w warunkach denaturujących prowadzono w żelach poliakryloamidowych z wykorzystaniem systemu do elektroforezy S2 (Biometra). Stosowano żele sekwencyjne o grubości 1 mm i zawartości poliakrylamidu 8% (do celów preparatywnych) lub 15% (do celów analitycznych), skład podano w Tabeli 3.7. Analizowane preparaty przed nałożeniem na żel denaturowano. W tym celu mieszano je z równą objętością 14 M roztworu mocznika z barwnikami (błękitem bromofenolowym i cyjanianem ksylenu) i inkubowano przez 3 min w 90°C, a następnie przez 5 min na lodzie. Rozdział prowadzono w buforze 1×TBE. Rozdzielony DNA/RNA uwidoczniano w świetle UV. Wizualizacji cząsteczek RNA znakowanych radioizotopowo dokonywano przy użyciu skanera FLA-5100 (Fujifilm).

Tabela 3.7. Skład żeli poliakryloamidowych do rozdzielania elektroforetycznego kwasów nukleinowych w warunkach denaturujących.

	Żel PAA 8%	Żel PAA 15%
Bufor TBE	1×	1×
Akrylamid/N,N'-metylenobisakrylamid (stosunek ilościowy 29:1)	8%	15%
Mocznik	7 M	7 M
APS	0,05%	0,05%
TEMED	0,1%	0,1%

Jeśli elektroforezę wykonywano w celach preparatywnych fragmenty żelu zawierające cząsteczki DNA/RNA o pożądanej długości przenoszono do 400 µl 0,3 M octanu sodu (pH 5,0) i inkubowano przez 16 h w 4°C delikatnie mieszając. Następnie krótko wirowano (5000×g, 4°C, 2 min), pobierano supernatant i zawarte w nim DNA/RNA wytrącano przez dodanie 2-3 objętości 96% etanolu (-20°C) i inkubację przez 1 h w -20°C. Preparat wirowano (16000×g, 4°C, 30 min), roztwór dekantowano, a osad suszono i rozpuszczano w wodzie wolnej od RNaz.

III.2.1.3. Rozdział elektroforetyczny białek w warunkach denaturujących (SDS-PAGE)

Rozdział elektroforetyczny białek w warunkach denaturujących prowadzono z wykorzystaniem systemu Mini-Protean (Bio-Rad). Stosowano żele poliakryloamidowe o grubości 0,75 mm złożone z dwóch warstw: zagęszczającej (5% poliakryloamid) i rozdzielającej (12% poliakryloamid). Skład obu warstw podano w Tabeli 3.8. Analizowane preparaty przed nałożeniem na żel mieszano z 0,25 objętości buforu obciążającego SB i inkubowano przez 5 min w 95°C, a następnie przez 5 min na lodzie. Podczas rozdzielania stosowano wzorzec masy cząsteczkowej białek Novex Sharp Pre-Stained (Life Technologies). Rozdział prowadzono w buforze TGS. Wizualizacji białek dokonywano przez barwienie w roztworze błękitu brylantowego Coomassie G250 PageBlue (Life Technologies) lub za pomocą techniki immunodetekcji (Rozdz. III.2.1.4).

Tabela 3.8. Skład żelu poliakryloamidowego do rozdzielania elektroforetycznego białek w warunkach denaturujących.

	Żel rozdzielający	Żel zagęszczający
Bufor Tris-HCl pH 8,8/6,8	300 mM (pH 8,8)	300 mM (pH 6,8)
Akrylamid/N,N'-metylenobisakrylamid (stosunek ilościowy 29:1)	5%	12%
SDS	1%	1%
APS	0,05%	0,05%
TEMED	0,1%	0,1%

III.2.1.4. Immunodetekcja białek

Transfer białek na membranę prowadzono w aparacie do elektrotransferu półsuchego Trans-blot SD (Bio-Rad). Przygotowywano 2 bibuły do elektrotransferu (BioRad) i membranę PVDF (Thermo Scientific) o wymiarach żelu rozdzielającego. Bibuły nasączano buforem TGS z dodatkiem 20% metanolu, a membranę przepłukiwano w metanolu. W aparacie do transferu umieszczano kolejno: bibułę, membranę, żel, bibułę. Transfer prowadzono przez 1 h, przy natężeniu prądu równemu co do wartości (mA) polu powierzchni membrany wyrażonemu w cm^2 , oraz przy napięciu nie przekraczającym 10 V.

Hybrydyzację Western-blot prowadzono w aparacie Snap i.d. (Millipore) zgodnie z opisem producenta. Membranę inkubowano kolejno w roztworze blokującym (30 ml, 10 min), roztworze przeciwciał I-rzędowych (3 ml, 10 min) i roztworze przeciwciał II-rzędowych sprzężonych z alkaliczną fosfatazą (3 ml, 10 min), wykonując trzykrotne płukanie w 30 ml buforu TBST po każdej inkubacji (Tabela 3.9). Reakcję barwną z udziałem alkalicznej fosfatazy wykonywano przy użyciu zestawu odczynników AP Detection Reagent Kit (Millipore) zgodnie z opisem producenta.

Tabela 3.9. Skład roztworów stosowanych do immunodetekcji białek.

	Roztwór blokujący	Roztwór przeciwciał I-rzędowych	Roztwór przeciwciał II-rzędowych
Bufor TBST	1×	1×	1×
BSA	1%	1%	1%
Mysie przeciwciała IgG anty-FLAG (Sigma Aldrich)	-	rozcieńczenie 1:1000	-
Kozie przeciwciała anty-mysie IgG z AP (Abcam)	-	-	rozcieńczenie 1:50000

III.2.2. ANALIZA GENOMU I BADANIE POZIOMU EKSPRESJI GENÓW

III.2.2.1. Identyfikacja genów *in silico*

Sekwencję genomową udostępnioną przez *Medicago truncatula* Genome Project (MTGP) w wersji Mt4.0 [165] przeszukiwano sekwencjami białek DCL z *Arabidopsis thaliana* i *Glycine max* (Załącznik 1) z wykorzystaniem programu tBLASTn opartego o algorytm WU-BLAST 2.0 (ang. *Basic Local Alignment Search Tool*). Analizę prowadzono stosując macierz substytucji BLOSUM62 i limitując parametr wartości oczekiwanej (E, ang. *expectation value*) do maksymalnie 1e-10. W zidentyfikowanych w ten sposób obszarach genomu wyszukiwano sekwencje kodujące typowe domeny obecne w białkach DCL (DExD/H, końca C helikazy, DUF283, PAZ, RNazy III, dsRBD) korzystając z bazy danych NCBI CDD (ang. *Conserved Domain Database*) [166].

III.2.2.2. Analiza bioinformatyczna sekwencji kodujących

Dla zidentyfikowanych genów kodujących białka DCL z *Medicago* (zwanymi dalej *MtDCL*) pobierano przewidywane sekwencje kodujące dostępne za pośrednictwem przeglądarki genomowej JBrowse [167]. Sekwencje te poddawano translacji *in silico*, a dla kodowanych przez nie białek przewidywano masę cząsteczkową (korzystając z serwera SIB Expasy Proteomics [168]) oraz architekturę domenową (korzystając z baz danych NCBI CDD [166] i EMBL SMART [169]). W celu przeprowadzenia analizy filogenetycznej wykonywano porównanie uzyskanych sekwencji białek *MtDCL* z sekwencjami znanych białek DCL z *A. thaliana*, *G. max* i *O. sativa* stosując algorytm ClustalW. Drzewa filogenetyczne konstruowano metodą najwyższej wiarygodności (ang. *maximum likelihood*) z próbkowaniem (ang. *bootstrap*) – 1000 powtórzeń. Analizy filogenetyczne prowadzono przy użyciu oprogramowania MEGA 6 [170].

III.2.2.3. Przygotowanie zawiesiny bakterii symbiotycznych *Sinorhizobium meliloti*

Bakterie *Sinorhizobium meliloti* indukujące powstawanie brodawek symbiotycznych na korzeniach *Medicago* hodowano w pożywce 2YT ze streptomycyną. 10 ml pożywki zaszczerpionej niewielką ilością bakterii inkubowano w 27°C przez 48 h z wytrząsaniem 200 obrotów/min. Bakterie płukano przez trzykrotne wirowanie (5000×g, 4°C, 5 min), dekantację i zawieszenie osadu w 10 ml roztworu MgSO₄ (10 mM). Do inokulacji

korzeni roślinnych stosowano zawiesinę bakterii stukrotnie rozcieńczoną w 10 mM MgSO₄ (względem objętości hodowli).

III.2.2.4. Przygotowanie materiału roślinnego

Nasiona *Medicago truncatula* (ekotyp Jemalong J5) sterylizowano w 95% kwasie siarkowym (15 min), a następnie dokładnie płukano schłodzoną do 4°C wodą. Nasiona umieszczano na szalce na wilgotnym krążku bibuły, inkubowano w ciemności: w 4°C przez 5 dni, a następnie w 23°C do skielkowania. Dalszą uprawę prowadzono w szklarni w następujących warunkach: wilgotność względna 40%, temperatura 22°C/18°C, fotoperiod 16 h/8 h, ilość światła aktywnego fotosyntetycznie (PPFD, ang. *photosynthetic photon flux density*) równa 500 μmol/m²s. Po skielkowaniu rośliny przekładano na szalki ze stałą pożywką MS/B5 zawierającą składniki mineralne wg Murashige i Skooga [171] oraz organiczne wg Gamborga [172]. Na szalkach prowadzono hodowlę 3- i 10-dniowych siewek, które dzielono na część górną (liścienie) oraz dolną (korzenie).

W celu pozyskania dorosłych roślin 7-dniowe siewki przesadzano do doniczek z wermikulitem, uprawiano 2 tygodnie, a następnie przesadzano do doniczek z mieszanką perlitu i piasku (stosunek 3:1). Rośliny w doniczkach podlewano co 48 h wodą, a co tydzień nawozem mineralnym NPK (6:3:6). Z dorosłych roślin pozyskiwano liście, łodygi, merystemy wierzchołkowe, młode nasiona oraz korzenie.

W celu pozyskania dojrzałych brodawek korzeniowych kiełkujące siewki wykładano na szalkę ze stałą pożywką BNM (kompozycja wg Ehrhardta [173]). Korzenie 7-dniowych siewek inokulowano 200 μl zawiesiny bakterii symbiotycznych *S. meliloti* (Rozdz. III.2.2.3). Hodowlę prowadzono 3 tygodnie w warunkach opisanych powyżej. Dla zapewnienia korzeniom właściwej wilgotności oraz braku dostępu światła przykrywano je wilgotną bibułą i folią aluminiową. W każdym przypadku materiał niezwłocznie po zebraniu zamrażano w ciekłym azocie i przechowywano w -80°C.

III.2.2.5. Izolacja RNA i odwrotna transkrypcja

Izolację i oczyszczanie RNA prowadzono wykorzystując materiał roślinny (Rozdz. III.2.2.4) roztarty w moździerz w obecności ciekłego azotu. Stosowano zestaw odczynników RNeasy Plant Mini Kit (Qiagen) zgodnie z opisem producenta. W czasie oczyszczania trawiono obecny w preparacie DNA z wykorzystaniem DNazy wolnej od

RNaz (Qiagen). Jakość otrzymanych preparatów analizowano z pomocą aparatu 2100 Bioanalzyer (Agilent). W dalszych pracach wykorzystywano preparaty, których tzw. współczynnik integralności (RIN, ang. *RNA integrity number*) wynosił co najmniej 8,0. Do odwrotnej transkrypcji używano zestaw odczynników Omniscript Reverse Transcription Kit (Qiagen), inhibitor RNaz RiboLock (Thermo Scientific) oraz mieszaninę starterów oligo-dT. Skład i warunki reakcji podano w Tabeli 3.10. W wyniku reakcji otrzymywano preparaty cDNA odpowiadające puli całkowitego mRNA w badanym materiale.

Tabela 3.10. Skład mieszaniny i warunki reakcji odwrotnej transkrypcji.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O (wolna od RNaz)	-	-	do 20 µl
Bufor RT	10×	1×	2 µl
Mieszanina dNTP	5 mM (każdy)	0,5 mM (każdy)	2 µl
Starter Oligo-dT	10 µM	1 µM	2 µl
Inhibitor RNaz RiboLock	10 U/µl	0,5 U/µl	1 µl
Odwrotna transkryptaza Omniscript	4 U/µl	0,2 U/µl	1 µl
RNA	500 ng/µl	25 ng/µl	1 µl
Odwrotna transkrypcja: 37°C – 1 h			

III.2.2.6. Badanie poziomu akumulacji mRNA

Do określenia poziomu akumulacji poszczególnych mRNA w materiale roślinnym używano system QX200 Droplet Digital PCR (Bio-Rad), wykorzystujący technikę łańcuchowej reakcji polimerazy (PCR, ang. *polymerase chain reaction*) w kroplach (ddPCR, ang. *droplet digital PCR*). Używano gotowej mieszaniny reakcyjnej QX200 ddPCR EvaGreen Supermix (Bio-Rad) zawierającej bufor, nukleotydy, polimerazę DNA oraz barwnik fluorescencyjny EvaGreen służący do detekcji produktów. Reakcje prowadzono wykorzystując dziesięciokrotnie rozcieńczone nieoczyszczone preparaty cDNA otrzymane na drodze odwrotnej transkrypcji (Rozdz. III.2.2.5). Stosowane startery wymieniono w Tabeli 3.1, skład i warunki reakcji podano w Tabeli 3.11. Przygotowane mieszaniny reakcyjne przed umieszczeniem w termocyklerze poddawano emulsyfikacji z olejem QX200 (Bio-Rad) w aparacie QX200 Droplet Generator (Bio-Rad).

Tabela 3.11. Skład mieszaniny i warunki reakcji ddPCR.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	4 µl
Starter FWD	4 µM	0,2 µM	1 µl
Starter REV	4 µM	0,2 µM	1 µl
Preparat po reakcji odwrotnej transkrypcji	0,1× (~100 ng/µl)	0,02× (~20 ng/µl)	4 µl
Mieszanina z enzymem (EvaGreen Supermix)	2×	1×	10 µl

Aktywacja enzymu: 95°C – 5 min
 Synteza produktów, 40 cykli: 95°C – 30 s, 57°C – 30 s, 72°C – 45 s
 Stabilizacja sygnału: 4°C – 5 min, 90°C – 5 min

Po zakończeniu reakcji wykonywano pomiar fluorescencji dla każdej wygenerowanej w danej mieszaninie kropli. Do pomiaru wykorzystywano aparat QX200 Droplet Reader (Bio-Rad). Uzyskane dane analizowano przy użyciu oprogramowania QuantaSoft (Bio-Rad), które na podstawie zmierzonej wartości fluorescencji kwalifikowało krople na tzw. pozytywne (w których nastąpiła amplifikacja docelowej sekwencji) lub tzw. negatywne (nie zawierające sekwencji docelowej). Umożliwiało to obliczenie bezwzględnej ilości badanej cząsteczki w próbce. Dokonywane obliczenia uwzględniały model losowej dystrybucji identyfikowanych cząsteczek w kroplach opisany rozkładem Poissona. Każdą analizę (dany mRNA w danej części rośliny) wykonywano w dwóch powtórzeniach biologicznych, a obliczony poziom akumulacji uśredniano. Do porównania wyników i określenia istotności statystycznej stosowano jednoczynnikową analizę wariancji (ang. *one-way ANOVA*) z testem *post-hoc* Tukey'a. Do analiz statystycznych stosowano oprogramowanie Prism 6 (GraphPad).

III.2.3. KLONOWANIE cDNA

III.2.3.1. Amplifikacja cDNA

Wybrane cDNA amplifikowano z nieoczyszczonej mieszaniny reakcyjnej otrzymanej w procesie odwrotnej transkrypcji RNA izolowanego z merystemów pędu *Medicago* (Rozdz. III.2.2.5). Stosowano metodę PCR z wykorzystaniem polimerazy DNA Expand High Fidelity (Roche) generującej produkty zakończone niesparowaną adenozyną na końcach 5', co było niezbędne w dalszych etapach procedury klonowania. Mieszaninę

reakcyjną przygotowano zgodnie z opisem w Tabeli 3.12, użyte startery wymieniono w Tabeli 3.1.

Tabela 3.12. Skład mieszaniny i warunki reakcji PCR do amplifikacji cDNA z materiału roślinnego.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	do 50 µl
Bufor Expand HF	10×	1×	5 µl
Mieszanina dNTP	2,5 mM (każdy)	200 µM (każdy)	4 µl
Starter FWD	10 µM	0,4 µM	2 µl
Starter REV	10 µM	0,4 µM	2 µl
Preparat po reakcji odwrotnej transkrypcji	1× (~1 µg/µl)	0,1× (~100 ng/µl)	5 µl
Polimeraza DNA Expand HF	3,5 U/µl	0,13 U/µl	0,75 µl

Wstępna denaturacja: 94°C – 2 min; Synteza produktów: 10 cykli: 94°C – 15 s, 55°C – 30 s, 68°C – 5 min, 20 cykli: 94°C – 15 s, 55°C – 30 s, 68°C – 5 min + 5 s w każdym kolejnym cyklu; Końcowe wydłużenie produktów: 68°C – 7 min

Produkty PCR rozdzielano w 1% żelu agarozowym (Rozdz. III.2.1.1). Produkty o oczekiwanej długości izolowano z żelu przy pomocy zestawu odczynników QIAquick Gel Extraction Kit (Qiagen) zgodnie z opisem producenta.

III.2.3.2. Klonowanie cDNA w plazmidzie wysokokopijnym

Oczyszczone produkty PCR ligowano z wektorem pCR-XL-TOPO z wykorzystaniem zestawu odczynników TOPO XL PCR Cloning Kit (Life Technologies). Mieszaninę wektora i insertu (stosunek molowy 1:1) inkubowano przez 5 min w 22°C. 5 µl mieszaniny ligacyjnej transformowano 50 µl komórek kompetentnych *E. coli* TOP10 (Life Technologies) metodą szoku cieplnego. Zawiesinę bakterii delikatnie mieszano i inkubowano kolejno: na lodzie (30 min), w 42°C (30 s) i ponownie na lodzie (2 min). Dodawano 250 µl pożywki SOC i wytrząsano przez 1 h w 37°C z prędkością 225 obrotów/min. Otrzymaną zawiesinę transformowanych bakterii rozprowadzano na szalkach ze stałą pożywką LB z kanamycyną. Szalki inkubowano w 37°C przez 16 h. Z pojedynczych kolonii bakterii wyhodowanych na stałej pożywce LB izolowano plazmidy (pCR-DCL1, pCR-DCL2), stosując zestaw odczynników QIAprep Spin Miniprep Kit (Qiagen) zgodnie z opisem producenta.

III.2.3.3. Analiza konstruktów plazmidowych

Obecność cDNA w przygotowanych plazmidach weryfikowano metodą PCR. W każdym przypadku stosowano parę starterów, z których jeden (FWD) był komplementarny do sekwencji kodującej białko, natomiast drugi (REV) do sekwencji użytego wektora. Skład i warunki reakcji PCR podano w Tabeli 3.13. Produkty reakcji analizowano metodą elektroforezy w 1% żelu agarozowym (Rozdz. III.2.1.1).

Tabela 3.13. Skład mieszaniny i warunki reakcji PCR do analizy konstruktów.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	do 20 µl
Bufor Taq	10×	1×	2 µl
MgCl ₂	25 mM	3 mM	2,4 µl
Starter FWD	10 µM	0,5 µM	1 µl
Starter REV	10 µM	0,5 µM	1 µl
Matryca DNA	50 ng/µl	2,5 ng/µl	1 µl
Mieszanina dNTP	2,5 mM (każdy)	200 µM (każdy)	1,6 µl
Polimeraza DNA Taq	2 U/µl	0,02 U/µl	0,5 µl

Wstępna denaturacja: 95°C – 2 min
 Synteza produktów, 30 cykli: 95°C – 30 s, 52°C – 30 s, 72°C – 2 min
 Końcowe wydłużenie produktów: 72°C – 5 min

Wyselekcjonowane klony poddawane były sekwencjonowaniu. Sekwencjonowano obszar kodujący białko wraz z miejscami ligacji z wykorzystaniem zestawu odczynników BigDye Terminator v1.1 Cycle Sequencing Kit (Life Technologies) oraz starterów wymienionych w Tabeli 3.1. Skład i warunki reakcji podano w Tabeli 3.14.

Tabela 3.14. Skład mieszaniny i warunki reakcji sekwencjonowania.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	do 20 µl
Matryca DNA	50 ng/µl	5 ng/µl	2 µl
Starter	10 µM	0,75 µM	1,5 µl
DMSO	50%	5%	2 µl
BigDye	-	-	3 µl
TRR	-	-	1 µl

Synteza produktów, 30 cykli: 94°C – 10 s, 50°C – 10 s, 60°C – 4 min

Do każdej mieszaniny reakcji sekwencjonowania dodawano 5 µl 125 mM roztworu EDTA i 60 µl 96% etanolu, mieszano, inkubowano przez 15 min i wirowano (16000×g, 4°C, 30 min). Roztwór dekantowano, a powstały osad przemywano 200 µl 70% etanolu i ponownie wirowano (16000×g, 4°C, 10 min). Roztwór dekantowano, a osad suszono i rozpuszczano w 10 µl formamidu. Tak przygotowane próbki poddawano rozdzielowi elektroforetycznemu i odczytowi sekwencji (Wydziałowa Pracownia Techniki Biologii Molekularnej, Wydział Biologii UAM).

III.2.3.4. Szybka amplifikacja końca 5' cDNA (5' RACE)

Metodą szybkiej amplifikacji końca 5' cDNA (5' RACE, ang. *rapid amplification of cDNA end*) identyfikowano fragmenty RNA stanowiące produkty trawienia alternatywnej formy mRNA MtDCL1 (MtDCL1-AS) w obrębie sekwencji odpowiadającej intronowi 14. W procedurze stosowano preparaty RNA izolowane z liścia i korzenia *Medicago* (Rozdz. III.2.2.5) oraz zestaw odczynników 5'/3' RACE Kit (Roche) zgodnie z opisem producenta. Na kolejnych etapach procedury wykorzystywano trzy różne startery, każdy komplementarny do innego fragmentu sekwencji egzonu 15 obecnego w mRNA MtDCL1 (Tabela 3.1). Produkty reakcji analizowano metodą elektroforezy w 1% żelu agarozowym (Rozdz. III.2.1.1).

III.2.4. KONSTRUKCJA WEKTORÓW EKSPRESYJNYCH

III.2.4.1. Synteza cDNA do konstrukcji wektorów ekspresyjnych

Podczas syntezy cDNA metodą PCR stosowano startery wprowadzające na jednym bądź obu końcach produktu sekwencje wymagane w dalszej procedurze tworzenia konstruktów (zob. Tabela 3.1 i Tabela 3.3). Matrycę stanowiły otrzymane wcześniej plazmidy pCR-DCL1 i pCR-DCL2. Mieszaninę reakcyjną przygotowywano zgodnie z opisem w Tabeli 3.15. Otrzymane preparaty cDNA oczyszczano przy pomocy zestawu odczynników QIAquick PCR Purification Kit (Qiagen) zgodnie z opisem producenta.

Tabela 3.15. Skład mieszaniny i warunki reakcji syntezy cDNA (PCR).

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	do 50 µl
Bufor Phusion HF	5×	1×	10 µl
Starter FWD	10 µM	0,5 µM	2,5 µl
Starter REV	10 µM	0,5 µM	2,5 µl
Matryca DNA	10 ng/µl	0,2 ng/µl	1 µl
Mieszanina dNTP	2,5 mM (każdy)	200 µM (każdy)	4 µl
Polimeraza DNA Phusion HS II	2 U/µl	0,02 U/µl	0,5 µl

Aktywacja enzymu: 98°C – 30 s
 Synteza produktów, 35 cykli: 98°C – 10 s, 72°C – 2 min
 Końcowe wydłużenie produktów: 72°C – 5 min

III.2.4.2. Trawienie dsDNA enzymami restrykcyjnymi do celów preparatywnych

Plazmid pGEX-6P-3 i odpowiednie produkty PCR poddawano trawieniu enzymami EcoRI i Sall (Thermo Scientific). Mieszaninę reakcyjną przygotowywano zgodnie z opisem w Tabeli 3.16. Produkt reakcji oczyszczano przy pomocy zestawu odczynników QIAquick PCR Purification Kit (Qiagen) zgodnie z opisem producenta.

Tabela 3.16. Skład mieszaniny i warunki reakcji trawienia dsDNA enzymami restrykcyjnymi.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	do 30 µl
Bufor O	10×	1×	3 µl
Preparat DNA	50 ng/µl	20 ng/µl	12 µl
Enzym EcoRI	10 U/µl	0,5 U/µl	1,5 µl
Enzym Sall	10 U/µl	0,5 U/µl	1,5 µl

Trawienie: 37°C – 4 h
 Inaktywacja enzymu: 65°C – 5 min

III.2.4.3. Klonowanie sekwencji kodujących w wektorach ekspresyjnych

W celu otrzymania wektorów ekspresyjnych na bazie pET100/D-TOPO prowadzono reakcję ligacji wektora i cDNA zgodnie z opisem producenta. Mieszaninę wektora i cDNA (stosunek molowy 1:1) inkubowano przez 5 min w 22°C.

W celu otrzymania wektorów ekspresyjnych opartych o pGEX-6P-3 reakcję ligacji wektora i cDNA trawionych enzymami restrykcyjnymi prowadzono zgodnie z opisem w Tabeli 3.17. Stosunek molowy cDNA do wektora wynosił około 4:1.

Tabela 3.17. Skład mieszaniny i warunki reakcji ligacji wektora pGEX-6P-3 i cDNA.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O	-	-	do 20 µl
Bufor do ligacji T4	10×	1×	2 µl
Plazmid pGEX-6P-3	10 ng/µl	2,5 ng/µl	5 µl
cDNA	10 ng/µl	5 ng/µl	10 µl
Ligaza DNA T4	5 U/µl	0,05 U/µl	0,2 µl
Ligacja: 22°C – 1 h			

5 µl mieszaniny ligacyjnej stosowano do transformacji 50 µl komórek kompetentnych szczepu *E. coli* TOP10 (Life Technologies) metodą szoku cieplnego. Zawiesinę bakterii delikatnie mieszano i inkubowano kolejno: na lodzie (30 min), w 42°C (30 s) i ponownie na lodzie (2 min). Dodawano 250 µl pożywki SOC i wytrząsano 1 h w 37°C z prędkością 225 obrotów/min. Otrzymaną zawiesinę transformowanych bakterii rozprowadzano na dwóch szalkach ze stałą pożywką LB z ampicyliną. Szalki inkubowano w 37°C przez 16 h. Z pojedynczych kolonii bakterii wyhodowanych na stałej pożywce LB izolowano klony wektorów ekspresyjnych, stosując zestaw odczynników QIAprep Spin Miniprep Kit (Qiagen) zgodnie z opisem producenta. Analiza konstruktów prowadzona była zgodnie z wcześniej opisaną procedurą (Rozdz. III.2.3.3).

III.2.5. PRODUKCJA BIAŁEK

III.2.5.1. Przygotowanie i przechowywanie szczepów bakteryjnych

Szczepy bakteryjne przeznaczone do produkcji białek DCL przygotowano w oparciu o szczep *E. coli* BL21-STAR (Life Technologies). Do 50 µl zawiesiny komórek kompetentnych dodawano 1 µl preparatu wektora ekspresyjnego (10 ng/µl) i prowadzono transformację metodą szoku cieplnego. Zawiesinę bakterii delikatnie mieszano i inkubowano kolejno: na lodzie (30 min), w 42°C (30 s) i ponownie na lodzie (2 min). Dodawano 250 µl pożywki SOC i wytrząsano przez 30 min w 37°C z prędkością 225 obrotów/min. Otrzymaną zawiesinę transformowanych bakterii

zaszczepiano 10 ml płynnej pożywki LB z ampicyliną. Bakterie inkubowano przez 16 h w 37°C wytrząsając z prędkością 300 obrotów/min otrzymując zawiesinę bakterii w stacjonarnej fazie wzrostu.

500 µl zawiesiny bakterii stosowano do zaszczepienia 10 ml świeżej pożywki LB z ampicyliną. Hodowlę prowadzono w 37°C wytrząsając z prędkością 300 obrotów/min do momentu osiągnięcia przez zawiesinę bakterii gęstości optycznej OD₆₀₀ równej około 0,6. Do 4 ml zawiesiny dodano 500 µl 80% roztworu glicerolu, delikatnie mieszano, zamrażano w ciekłym azocie i przechowywano w -80°C. W celu wznowienia hodowli niewielką ilością zamrożonego preparatu zaszczepiano płynną pożywkę LB z ampicyliną. Bakterie inkubowano przez 16 h w 37°C wytrząsając z prędkością 300 obrotów/min otrzymując zawiesinę bakterii w stacjonarnej fazie wzrostu.

III.2.5.2. Synteza białka w bakteriach

W celu produkcji białek heterologicznych w bakteriach płynną pożywkę LB z ampicyliną zaszczepiano zawiesiną bakterii danego szczepu ekspresyjnego w stacjonarnej fazie wzrostu w stosunku 20:1. Hodowlę prowadzono w 37°C wytrząsając z prędkością 300 obrotów/min do momentu osiągnięcia przez zawiesinę bakterii gęstości optycznej OD₆₀₀ równej około 0,6. Produkcję indukowano przez dodanie roztworu izopropyl-β-D-tiogalaktopiranozydu (IPTG) do stężenia końcowego 0,5 mM. Produkcję prowadzono przez 4 h w 37°C lub przez 16 h w 18°C wytrząsając hodowlę z prędkością 300 obrotów/min. Po tym czasie zawiesinę bakterii wirowano (7000×g, 4°C, 5 min), pożywkę dekantowano, a osad bakteryjny zamrażano w -20°C.

III.2.5.3. Analiza rozpuszczalności białka

Osad bakteryjny pochodzący z 1,5 ml hodowli zawieszano w 200 µl roztworu CLB (50 mM Tris-HCl pH 8,0, 1× CelLytic B (Sigma Aldrich), 100 mM NaCl). Preparat inkubowano wytrząsając z prędkością 200 obrotów/min przez 10 min i wirowano (16000×g, 5 min). Supernatant zawierający frakcję białek rozpuszczalnych zachowywano do analizy SDS-PAGE, a osad zawieszano w 200 µl roztworu CLB i dodawano 2 µl roztworu lizozymu (20 mg/ml). Preparat inkubowano wytrząsając z prędkością 200 obrotów/min przez 10 min i wirowano (16000×g, 5 min). Supernatant zawierający frakcję białek nierozpuszczalnych zachowywano do analizy SDS-PAGE, a pozostały osad zawieszano w 200 µl 10× rozcieńczonego roztworu CLB. Preparat

wirowano (16000×g, 5 min), supernatant dekantowano, a osad zawierający ciała inkluzyjne zawieszano w 200 µl roztworu 50 mM Tris-HCl pH 8,0 i 100 mM NaCl i zachowywano do analizy SDS-PAGE. Zebrane frakcje analizowano metodą SDS-PAGE (Rozdz. III.2.1.3) oraz przez immunodetekcję białek (Rozdz. III.2.1.4).

III.2.5.4. Izolacja i oczyszczane białka na złożu z immobilizowanym glutationem

Osad bakteryjny (Rozdz. III.2.5.2) zawieszano w roztworze CL przeznaczonym do izolacji danego białka (CL1 lub CL2, zgodnie z Tabelą 3.18) w proporcji 10 ml roztworu na 1 g osadu. Dodawano 0,01 objętości lizozymu (20 mg/ml) i inkubowano wytrząsając z prędkością 200 obrotów/min przez 15 min. Preparat umieszczano na lodzie, poddawano sonikacji (2 s sonikacji co 10 s, 75 cykli), a następnie wirowano (32000×g, 4°C, 30 min). Do supernatantu dodawano złoże zawierające immobilizowany glutation Glutathione Sepharose HP (GE Healthcare) w proporcji 1 ml zawiesiny złoża na 10 ml preparatu. Przed zastosowaniem zawiesina złoża przygotowywana była przez trzykrotne wirowanie (1000×g, 2 min), dekantację i zawieszenie w równej objętości wody. Preparat ze złożem inkubowano na lodzie przez 30 min delikatnie mieszając i nakładano na kolumnę chromatograficzną. Upakowane w kolumnie złoże przemywano zebrany przesączem, a następnie 100 ml roztworu B przeznaczonego do izolacji danego białka (B1 lub B2, zgodnie z Tabelą 3.18). Złoże ze związanym białkiem zawieszano w identycznej objętości buforu B1/B2 zawierającego proteazę ludzkiego rinowirusa 3C (HRV 3C, ang. *human rhinovirus 3C*; Sigma Aldrich) w stężeniu 20 U/ml. Preparat ze złożem inkubowano w 4°C przez 16 h delikatnie mieszając i ponownie nakładano na kolumnę chromatograficzną. Zebrany z kolumny preparat zagęszczano dwustukrotnie przez wirowanie na filtrach membranowych Amicon Ultra 10K (Millipore) w 4°C zgodnie z opisem producenta. W przypadku białka DCL1-CC dokonywano także wymiany roztworu buforującego na B2 przez dziesięciokrotne rozcieńczenie otrzymanego preparatu buforem B2 i ponowne zagęszczenie. Otrzymane preparaty analizowano metodą SDS-PAGE (Rozdz. III.2.1.3) oraz przez immunodetekcję białek (Rozdz. III.2.1.4).

Tabela 3.18. Skład roztworów do izolacji białek w celach preparatywnych.

	Izolacja białka DCL1-CC		Izolacja białka DCL2-CC	
	Roztwór CL1	Roztwór B1	Roztwór CL2	Roztwór B2
Bufor PBS	1×	1×	-	-
Tris-HCl pH 8,0	-	-	50 mM	50 mM
NaCl	-	-	100 mM	100 mM
TCEP	0,5 mM	0,5 mM	0,5 mM	0,5 mM
Cellytic B (Sigma Aldrich)	1×	-	1×	-
Inhibitor proteaz Complete Ultra EDTA-free (Roche)	1 tabl/50 ml	-	1 tabl/50 ml	-

III.2.6. BADANIE AKTYWNOŚCI RYBONUKLEAZOWEJ BIAŁEK

III.2.6.1. Otrzymywanie substratów RNA

Substraty RNA stosowane w reakcjach z białkami DCL lub Dicer otrzymywano na drodze syntezy chemicznej lub transkrypcji *in vitro*. W przypadku cząsteczek otrzymywanych na drodze transkrypcji jako matrycę stosowano oligonukleotydy DNA wymienione w Tabeli 3.1. Transkrypcję przeprowadzano przy użyciu zestawu odczynników MEGAscript T7 (Life Technologies) i krótkiego startera (T7-TR) komplementarnego do fragmentu każdej matrycy. Skład i warunki reakcji podano w Tabeli 3.19.

Tabela 3.19. Skład mieszaniny i warunki reakcji defosforylacji końca 5' RNA.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O (wolna od RNaz)	-	-	2 μl
Matryca DNA	40 ng/μl	8 ng/μl	4 μl
Starter DNA (T7-TR)	20 ng/μl	2 ng/μl	2 μl
Utworzenie dupleksów matryca/starter: 95°C – 5 min, powolne schłodzenie do 22°C			
Bufor do T7	10×	1×	2 μl
Mieszanina rybonukleotydów	18,75 mM (każdy)	7,5 μM (każdy)	8 μl
Mieszanka z Polimerazą RNA T7	-	-	2 μl
Transkrypcja <i>in vitro</i> : 37°C – 8 h			
DNaza Turbo	-	-	1 μl
Trawienie DNA: 37°C – 15 min			

Produkty reakcji rozdzielano metodą elektroforezy w 8% żelu poliakrylamidowym, a właściwe transkrypty oczyszczano zgodnie z procedurą opisaną w Rozdz. III.2.1.2. Otrzymane cząsteczki RNA poddawano defosforylacji końca 5', stosując alkaliczną fosfatazę FastAP (Thermo Scientific). Skład i warunki reakcji podano w Tabeli 3.20.

Tabela 3.20. Skład mieszaniny i warunki reakcji defosforylacji końca 5' RNA.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O (wolna od RNaz)	-	-	do 50 µl
Bufor do FastAP	10×	1×	5 µl
RNA	10 µM	1 µM	10 µl
Alkaliczna fosfataza (FastAP)	1 U/µl	0,2 U/µl	20 µl
Defosforylacja: 37°C – 30 min, Inaktywacja enzymu: 75°C – 5 min			

Produkty reakcji wytrącano przez dodanie 40 µl wody wolnej od RNaz, 10 µl 3 M octanu sodu (pH 5,0) oraz 200 µl 96% etanolu (-20°C) i inkubację przez 1 h w -20°C. Preparat wirowano (16000×g, 4°C, 30 min), roztwór dekantowano, a osad suszono i rozpuszczano w wodzie wolnej od RNaz.

III.2.6.2. Znakowanie radioizotopowe substratów RNA

Cząsteczki RNA stosowane w reakcjach z białkami DCL lub Dicer znakowano radioizotopowo przez fosforylację końca 5' $\gamma^{32}\text{P}$ ATP. Stosowano kinazę polinukleotydową T4 (Thermo Scientific), skład i warunki reakcji podano w Tabeli 3.21.

Tabela 3.21. Skład mieszaniny i warunki reakcji fosforylacji końca 5' RNA.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O (wolna od RNaz)	-	-	do 20 µl
Bufor A do PNK T4	10×	1×	2 µl
RNA	10 µM	0,5 µM	1 µl
$\gamma^{32}\text{P}$ ATP	10 µCi/µl	0,5 µCi/µl	1 µl
Kinaza polinukleotydowa (PNK) T4	10 U/µl	0,5 U/µl	1 µl
Fosforylacja końca 5': 37°C – 15 min			

Produkty reakcji rozdzielano metodą elektroforezy w 8% żelu poliakrylamidowym i oczyszczano zgodnie z procedurą opisaną w Rozdz. III.2.1.2.

III.2.6.3. Analiza aktywności rybonukleazowej białek

Aktywność rybonukleazową białek Dicer i DCL badano prowadząc reakcje trawienia znakowanego radioizotopowo substratu RNA w roztworze buforującym 20 mM Tris, 50 mM NaCl, 2,5 mM MgCl₂ (pH 7,5). Stosowano przygotowane wcześniej preparaty białek DCL1-CC i DCL2-CC oraz preparaty rekombinowanych białek: ludzkiego białka Dicer – Turbo Dicer (Genlantis) oraz białka Dicer z *G. intestinalis* – PowerCut Dicer (Thermo Scientific). Reakcje kontrolne prowadzono w obecności 5 mM EDTA lub bez dodatku białka. Skład i warunki reakcji podano w Tabeli 3.22. Produkty reakcji rozdzielano metodą elektroforezy w 15% żelu poliakrylamidowym zgodnie z procedurą opisaną w Rozdz. III.2.1.2.

Tabela 3.22. Skład mieszaniny i warunki reakcji trawienia RNA.

Odczynnik	Stężenie wyjściowe	Stężenie końcowe	Dodana objętość
H ₂ O (wolna od RNaz)	-	-	2 μl
Substrat RNA	5000 cpm/μl	1000 cpm/μl	2 μl
Utworzenie substratu o homogennej strukturze: 90°C – 3 min, powolne schłodzenie do 4°C			
Bufor do badania aktywności RNazowej	5×	1×	2 μl
EDTA (reakcja kontrolna)/ H ₂ O (wolna od RNaz)/	25 mM/-	5 mM/-	2 μl
Enzym	-	-	1 μl
Trawienie: 37°C – 1 h			

IV. WYNIKI

IV.1. IDENTYFIKACJA GENÓW BIAŁEK DCL W GENOMIE *MEDICAGO TRUNCATULA*

W celu zidentyfikowania genów *DCL* w genomie *Medicago truncatula* wykorzystano bazę danych udostępnionych przez MTGP. Projekt ten realizowany jest od 2003 roku, a jego celem jest poznanie sekwencji części euchromatynowej genomu *Medicago*. W momencie rozpoczęcia prac badawczych opisanych w niniejszej rozprawie publicznie dostępna sekwencja genomu (wersja Mt2.0) obejmowała 214 Mpz, natomiast najnowsza edycja (wersja Mt4.0) zawiera 390 Mpz [165].

Aby zidentyfikować geny kodujące białka DCL u *Medicago* zastosowano algorytm tBLASTn (ang. *basic local alignment search tool*). Algorytm ten umożliwił przeszukanie sekwencji genomowej poddanej translacji *in silico* (w obu kierunkach, w każdym w 3 ramkach odczytu) zadaną sekwencją aminokwasową. Jako zapytanie wykorzystano sekwencje aminokwasowe białek DCL z *Arabidopsis thaliana* oraz *Glycine max* pochodzące z bazy danych NCBI GenBank (Załącznik 1). W ten sposób uzyskano listę aminokwasowych sekwencji homologicznych wraz z koordynatami genomowymi odpowiadających im sekwencji nukleotydowych. Sekwencje zidentyfikowanych w ten sposób regionów genomowych porównano z bazą danych CDD dostępną na serwerze NCBI w celu potwierdzenia obecności sekwencji kodujących zestaw domen typowy dla białek Dicer.

Analiza ta pozwoliła zidentyfikować sześć loci obejmujących geny białek DCL w genomie *Medicago*. Lokalizację chromosomową genów *MtDCL* przedstawiono na Rysunku 4.1a. Trzy spośród zidentyfikowanych genów (*MtDCL1*, 2 i 3) zostały szczegółowo opisane we wcześniejszych badaniach [137]. Identyfikacji trzech pozostałych dokonano w ramach niniejszej pracy. W ramach MTGP tworzony jest kompleksowy opis poznanej sekwencji genomu (tzw. anotacja). Opis ten, dostępny za pośrednictwem przeglądarki JBrowse [167], uwzględnia przewidywaną strukturę i funkcję genów, w tym lokalizację granic pomiędzy egzonami i intronami oraz wygenerowaną na tej podstawie sekwencję cDNA. Struktura genów określana jest w sposób zautomatyzowany, w oparciu o bioinformatyczną analizę sekwencji oraz dostępne dane pochodzące z sekwencjonowania całych transkryptomów, fragmentów mRNA (tzw. EST, ang. *expressed sequence tags*) i białek [165]. Wszystkim trzem nieznanym dotąd genom *MtDCL* automatycznie przypisana została prawidłowa funkcja

w bazie danych MTGP (DCL), jednak bez określenia rodzaju kodowanego białka (DCL1-4).

W celu stwierdzenia, do którego z czterech typów należą białka DCL kodowane przez nowo zidentyfikowane geny ustalono ich podobieństwo do znanych białek DCL pochodzących z *Arabidopsis thaliana*, *Glycine max* i *Oryza sativa*. Analizy filogenetycznej dokonano wykorzystując poddane translacji *in silico* najdłuższe ramki odczytu obecne w cDNA wszystkich sześciu zidentyfikowanych genów *MtDCL*. Pozostałe użyte sekwencje pochodziły z bazy danych NCBI GenBank (Załącznik 1). Otrzymane drzewo filogenetyczne przedstawia Rysunek 4.1b.

Przeprowadzona analiza filogenetyczna potwierdziła prawidłową klasyfikację trzech wcześniej opisanych genów kodujących MtDCL1, 2 i 3. Natomiast wśród trzech nieznanach dotąd genów jeden został zakwalifikowany jako homolog *AtDCL4*, a dwa pozostałe jako homologi *AtDCL2*. Co ciekawe oba nieznanne dotąd geny białek MtDCL2 wykazują ponad 95% podobieństwa w obrębie sekwencji kodującej. Jeden z nich obejmuje jednak kodon STOP, który powoduje, że hipotetyczne białko nie posiada dużej części domeny RNazy IIIb oraz zlokalizowanej na końcu karboksylowym dsRBD. Sekwencja kodująca obie domeny obecna jest w dojrzałym mRNA za kodonem STOP (Rys. 4.2). Trzy nieznanne dotąd białka nazwane zostały zgodnie z wykazanim podobieństwem do białek DCL z *Arabidopsis* oraz przyjętą nomenklaturą jako MtDCL4, MtDCL2b (pełnej długości białko DCL) oraz MtDCL2c (białko skrócone), a kodujące je geny analogicznie jako *MtDCL4*, *MtDCL2b* i *MtDCL2c*. Ponadto przyjęto oznaczenie *MtDCL2a* dla wcześniej zidentyfikowanego homologu *AtDCL2*. Charakterystykę zidentyfikowanych genów kodujących białka MtDCL przedstawiono w Tabeli 4.1.

Rysunek 4.1 (następna strona). Geny białek DCL u *Medicago truncatula*. (a) Lokalizacja chromosomowa sześciu genów kodujących białka DCL. Obraz sporządzono na podstawie sekwencji genomu *Medicago* w wersji Mt4.0 [165]. (b) Drzewo filogenetyczne ilustrujące zależności pomiędzy białkami DCL z *M. truncatula* (Mt), *A. thaliana* (At), *G. max* (Gm) i *O. sativa* (Os). Roślinne białka DCL danego rodzaju (DCL1-4) tworzą wyraźne grupy filogenetyczne. Drzewo skonstruowano stosując metodę najwyższej wiarygodności. (c) Drzewo filogenetyczne ilustrujące zależności pomiędzy znanymi białkami z rodzaju DCL2 z *A. thaliana* (At) oraz trzech gatunków bobowatych: *M. truncatula* (Mt), *G. max* (Gm) i *L. japonicus* (Lj). Białka DCL2 tworzą grupy filogenetyczne zgodne z przynależnością gatunkową. Drzewo skonstruowano stosując metodę najwyższej wiarygodności z próbkowaniem. Na węzłach wskazano odsetek drzew, w których po poróbkowaniu obserwowano dokładnie taki sam węzeł.

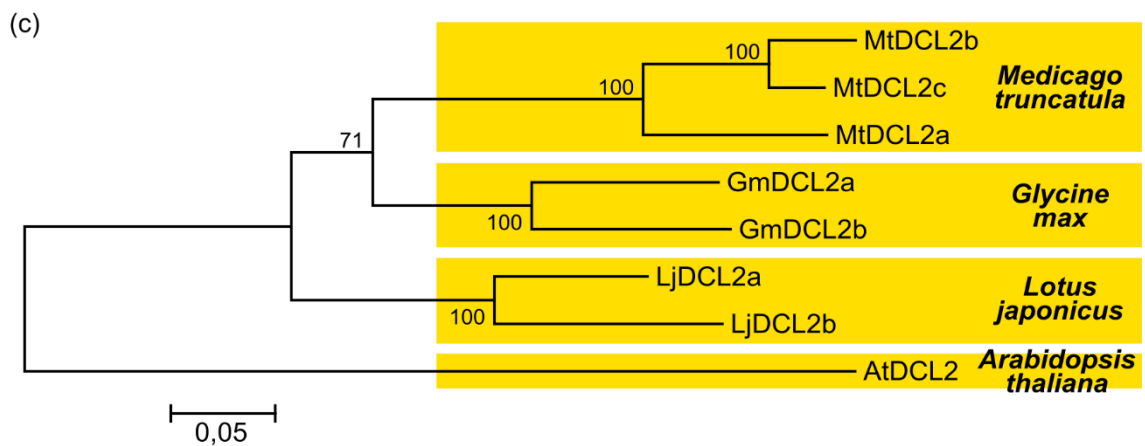
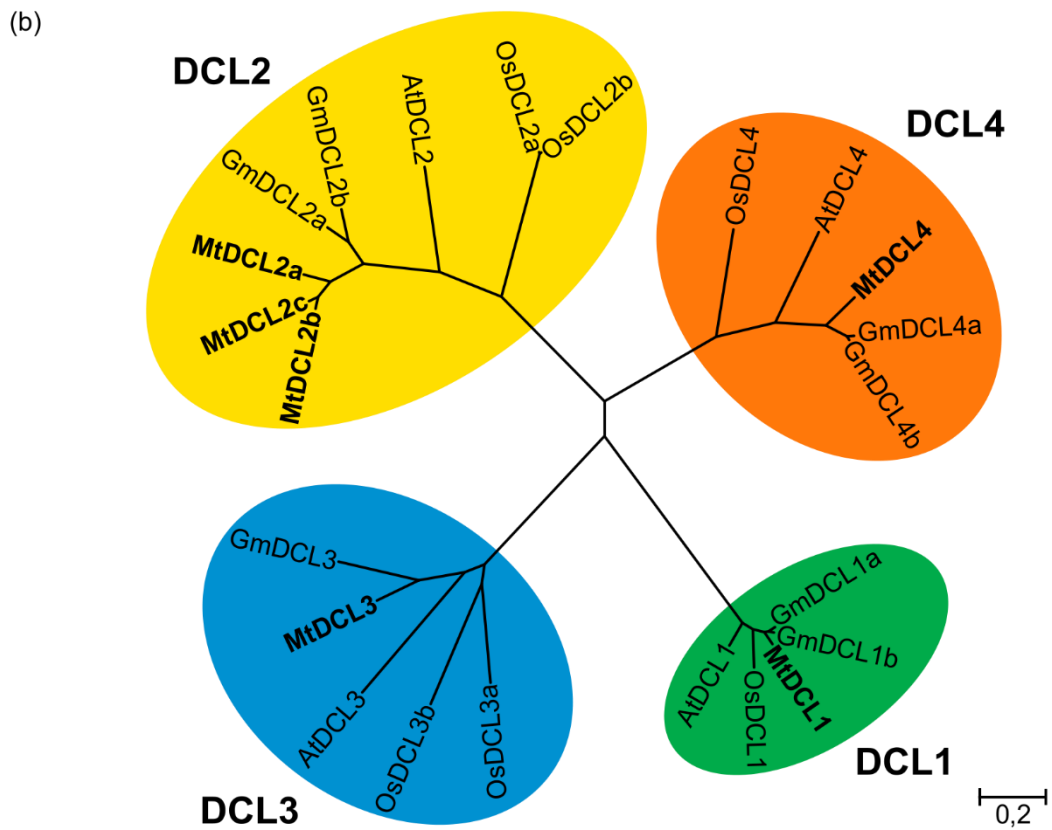
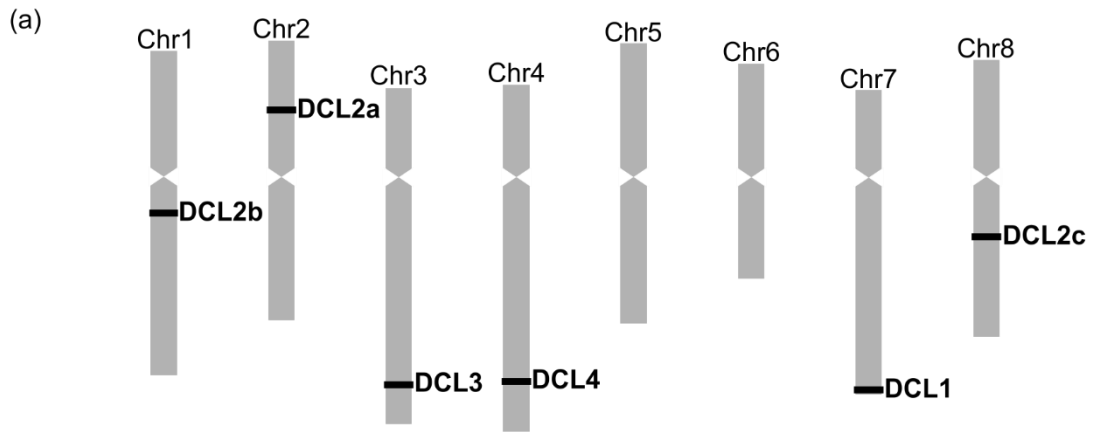


Tabela 4.1. Charakterystyka zidentyfikowanych genów kodujących białka DCL u *Medicago*.

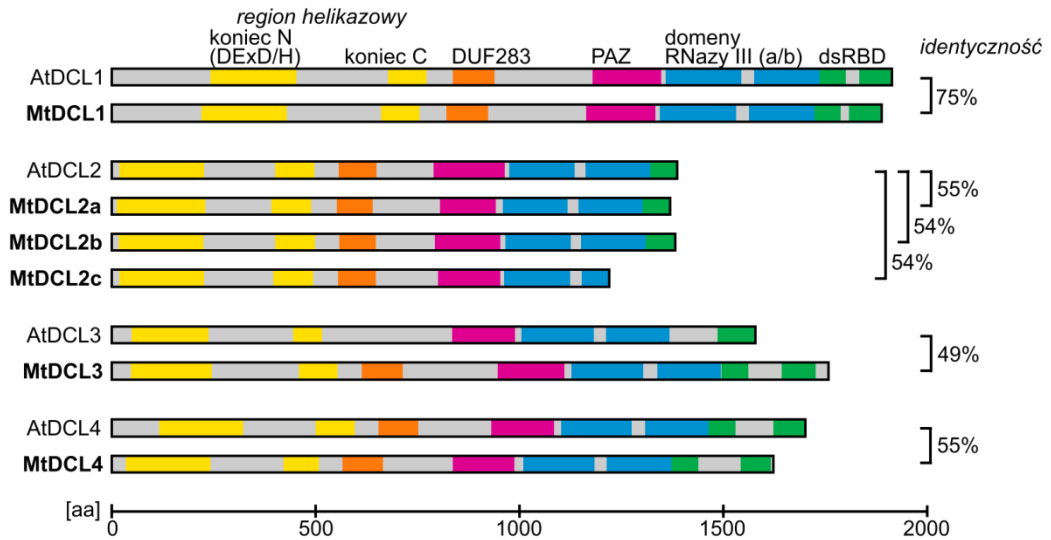
Gen	Nazwa locus	Lokalizacja sekwencji kodującej (Mt4.0)	Długość białka	Masa cząsteczkowa
<i>MtDCL1</i>	Medtr7g118350	Chr7: 49139645-49129819	1880 aa	211,6 kDa
<i>MtDCL2a</i>	Medtr2g030490	Chr2: 11495894-11486899	1377 aa	156,6 kDa
<i>MtDCL2b</i>	Medtr1g060740	Chr1: 26454644-26463238	1385 aa	157,5 kDa
<i>MtDCL2c</i>	Medtr8g069975	Chr8: 29673538-29680711	1219 aa	138,6 kDa
<i>MtDCL3</i>	Medtr3g105390	Chr3: 48587585-48600236	1730 aa	192,6 kDa
<i>MtDCL4</i>	Medtr4g116860	Chr4: 48375041-48390001	1625 aa	182,4 kDa

Dodatkowo przeprowadzono analizę filogenetyczną znanych białek z rodzaju DCL2 z *Arabidopsis thaliana* oraz trzech gatunków roślin bobowatych: *Medicago truncatula*, *Glycine max* i *Lotus japonicus*. *Arabidopsis* posiada jeden gen kodujący białko DCL2, podczas gdy u wszystkich trzech bobowatych zidentyfikowano jego duplikacje. W przeciwieństwie do poprzedniej analizy, gdzie białka danego rodzaju (DCL1-4) z różnych organizmów tworzyły wyraźne grupy filogenetyczne, poszczególne izoformy DCL2 zgrupowały się zgodnie z przynależnością gatunkową (Rys. 4.1c).

Analizę sekwencji aminokwasowych zidentyfikowanych białek przeprowadzono z użyciem baz danych CDD (NCBI) oraz SMART (EMBL). Pięć białek DCL z *Medicago* posiada typowy zestaw domen charakterystyczny dla białek Dicer: domenę obejmującą motyw DExD/H i końca C helikazy tworzące razem region helikazowy białka oraz DUF283, PAZ, domeny RNazy III (a i b) i dwie bądź jedną (DCL2) dsRBD. Jak zaznaczono wcześniej, kodon STOP obecny w mRNA *MtDCL2c* sprawia, że kodowane białko jest skrócone względem pozostałych białek tego rodzaju. Porównanie sekwencji aminokwasowych białek danego rodzaju z *Medicago* i *Arabidopsis* wykazało identyczność pomiędzy *MtDCL1* i *AtDCL1* na poziomie 75%, a w przypadku DCL2, DCL3 i DCL4 na poziomie około 50%. *MtDCL2a* i *MtDCL2b* są identyczne w około 85%, natomiast *MtDCL2b* i *MtDCL2c* – w około 95% (Rys. 4.3).

Rysunek 4.2 (następna strona). Porównanie sekwencji kodujących dwa nieznane dotąd homologii DCL2 z *Medicago* (nazwane *MtDCL2b* i *MtDCL2c*). Obie sekwencje są identyczne w ponad 95%. Prostokątem zaznaczono kodony STOP kończące najdłuższe ramki odczytu obecne w obu cDNA. Lokalizacja kodonu STOP w ramce odczytu kodującej *MtDCL2c* skutkuje powstaniem białka skróconego względem *MtDCL2b*, mimo iż sekwencja kodująca brakujący fragment białka obecna jest w cDNA *MtDCL2c*.

DCL2b ATGGAAGAAGCTC GATGGAATAGATGACGTTATCAGGAGCAGATCA GCCTGATACCTTTCATTTGCTAGAAGCTCAACTTGAA 90
DCL2c 90
DCL2b GCATTGGATAAAGCATT CGCGAAAATACTATAGTGTACTTGGAGACTGGTTGGAAAGACTTTGATAGCCATCATGCTTCTCGAAGC 180
DCL2c 180
DCL2b TACGCTATCACCTCCGAAAGCCTTCTCCTTACATGCTGTTTTTTGGTTCCCAAAGTGTGTTGGTTCTCAACAAGCTAAGGCTTTG 270
DCL2c 270
DCL2b AGAAATCACACTGATTTGAAAGTTGGAACGTTATGGGGAGAGAAAGGTGTGACAACTGGGATGGAGCTATATGAAAGAACAAATGGAG 360
DCL2c 360
DCL2b AAACATGAGGTGCTTGTATGACTCCTGCAATATGCTTTCTGCTTGAGGCATAGCTTTATCAAAGTAAAGTAAAGTTTAAATA 450
DCL2c 450
DCL2b ATGGATGAATGCCATCATGCTGCTGGTAGACACCATATGCTGTATCATGAGGGAAATTTATCACCATCAATTAAGATCTGGTATCACC 540
DCL2c 540
DCL2b GAGCTTCTCGAATTTTGGGATGACTGCATCCCAATTAAGTCGAAAGCGGCAAATCTAAGTTGACCTTGTGCGAAAAATATTCGGGAA 630
DCL2c 630
DCL2b CTAATGACTTAAATGCATTCAAAGGTTTATACATGTGTGAGTGATGCTGTCATCTCCAATTCATACCAACCTCAACCAAAATTCAG 720
DCL2c 720
DCL2b GTTACAATGAAATGTAATTTTATATGCGCAATTCGAGAATTAGCCAATAAACTCAGCAGTTTAAAACAACAGCAGGCTTTACGTC 810
DCL2c 810
DCL2b ACAAGTCAAATTTTCAAAATCACTGTTGACTCTGCACACAAGAATAAGCAAGGATTACACTCACTCAATTTTTGCTTGGAAAGAG 900
DCL2c 900
DCL2b CTGGTGTGTTGGTTGGCTTTGAAGGCCGCGGAGTCTTATCATCCATGAAATGAAACTC-----TTTATGGGTAACCTCA 978
DCL2c 990
DCL2b GGTGATCAAATTTGAAAGATTTTGTTCGCTGCTATTCTGACACTGCAAAAGTTACATACCATCTGATCCTCAGTGGATTATTGGTGAC 1068
DCL2c 1080
DCL2b AATATGAATCCGATATGAAAGGGGGCTATTGACCTCAAAGTGTGTGCTTATGGACTGTCTACTTGAGTACAGGAATTTTACTGAA 1158
DCL2c 1170
DCL2b ATGAGATGATAGTTTTGTGAAAGGGTATTACAGCTATTGCTTGGAGTCTTTTGAACAATTTGCTTCCGAAATACAATAGCTGG 1248
DCL2c 1260
DCL2b ACGGCTAAATTCATTGCAGGAAACAGTAGTAAATACAAAACCAATCAAGGAAAAATCAAATCAAATTTTGAAGATTTCCGATAGGGA 1338
DCL2c 1350
DCL2b TTGGTCAACGTCATTGTTGCGACATCAATCTGAGGAGGGTTAGATGTTCAAAGTGAATTTGGTTATTAGATTGACCCATCTCCC 1428
DCL2c 1440
DCL2b ACCGTTGTCAGTTTTGTACAGTCCCGTGGCCGTGCCAGAATGCAAAATCAGATTATATATTGATGGTTAAGAGTGGGATTCAGTTACA 1518
DCL2c 1530
DCL2b CGTCTCGATTAGAGAAATACCTTGATGGTGTGAAATGATGAGGAAGGAGTCCCTGCATCATCTTCCCTCCATGCGAAAGTCTGAA 1608
DCL2c 1620
DCL2b AGTGATAAATTAATGAGCAAACCTACCGTGTGCAAGCACGGAGCTGTTGTCAATCTTAGTTC TAGCATTACTTGTATACTTGTAC 1698
DCL2c 1710
DCL2b TGCTCACGACTTCCAGCAGATGGGTACTTTAAACCGACTCCAAGTGGGACAAACAGACTGGAACGTTGATCTTCTTAAAGATTTGCTC 1788
DCL2c 1800
DCL2b CTCAAAGTACAAGCTGTTAATGTACATGGTGAATCAAAGTACTTAAAAAATATGCGATGCTTGAAGCATGCAAAACGACTTCATATGAT 1878
DCL2c 1884
DCL2b GGAGCTTGGACAGATAAATCTGTTCTTCTATAGTCTGTGAAAAAGCAGAGGTTGGAGGACTTTGGGAATGAACCATATGAGGAGGATCAA 1968
DCL2c 1974
DCL2b CCAAGTATCGTACCGCTTGAATTTGGTGAATCGTATGCAAAATAGTAAACACAATATACTATTGTTATTTATAGAGCTAAACAAAAAT 2058
DCL2c 2064
DCL2b TTTAGCTACGATATCACGTGCAAGACATTTTCTGCAACTAGAGTCAAGCTTGATCTAGAAACGAAATGATGAAATTTACATGAGT 2148
DCL2c 2154
DCL2b TTCGATAGAGGTAGCTTGTCTGTAATTTGAATATACAAAGGATCCATTAATCTCTCCCGGATCAGGTTCTTTTGTGTAAGAAATTCCAA 2238
DCL2c 2244
DCL2b GTTAATGTCCTTGGAAATCTTATGAACCATAAAACAGATATGGAACCTTTCCGCAAGTTCTGTTTGGAAAGATGATCTGAGTTGAT 2328
DCL2c 2334
DCL2b TATCTTCTGCTGCCATCTATTGCTATAGAGGAAACCATCTGTTGATGGATAACATTAATCCATACATCCATCTATAGTTAAGTGT 2418
DCL2c 2424
DCL2b CTGCATCATGAAGCAATATATGACTGAAAGGGTTAGTTTGGCCCTGCATATTGCAAAATGCTTTGATTTGACTCTCACAATGGC 2508
DCL2c 2514
DCL2b CGTACCTACATCACCATGTTAATGGAATGGATGGAATTCACCTCTGGAAGTAGGGGATGGTGAAGTATCACATACAGAAAGTAC 2598
DCL2c 2604
DCL2b TTTGACAAAAACATGGCATTCAATTCGCTTTGAGCATCAACGCCACTTAAGGCAAGACATGTTTTCCAGTAAAAAATATTGTCAC 2688
DCL2c 2694
DCL2b GGATATAGGCAAGCGAAGGACAGAGATGTGAGCAAGACCTTGTGAGTTACCTCCTGAACCTTGCCTATAATCATGACCAATATCA 2778
DCL2c 2784
DCL2b GTCGAACTATTTACTCATTTCCTTTATCCATCGATCATGCATAGACTTGAGTCAATGCTTGTAGCTTGAACCTTAAAAAGATGCAT 2868
DCL2c 2874
DCL2b TTGATCTTTTGGCCGAGAACAGTATTCAAACATTCAAGGCTTTGGAAGCAATGACCACAAAGTCAATGCAAGGAGAAGCTTCAATTCAGAG 2958
DCL2c 2964
DCL2b TCTCTGGAGACTTAGGAGATCTTTTTTAAATAAGCTGTAGTCAACAGCTTTTTAATATGATGAAAAATCATCAGGAGGCTCTCTT 3048
DCL2c 3054
DCL2b AGTGTGAAGAGGGAGAAGATCAATTTAAGCTGCCCTCTGTAAGTTAGGTGCGAGTCTAAACTCCGGGCTCATACGGAATGAGCCT 3138
DCL2c 3144
DCL2b TTTGATCTTAAGACCTGGATCACTTCTGGAGTTAAATCAAGATGTTTTAAATTAGAAGAGACAGTTTTCAAGGGGACAAAAATATATACT 3228
DCL2c 3234
DCL2b CGCGGAATAGAAAAATGAAACGGAAAATTCGCTGATGTTGTTGAAGCACTAATTTGCTGCTTCTTAGCCTGGTGGTGGATGGCT 3318
DCL2c 3324
DCL2b GCATTGCTGTTTTGGATTGGGTGGTATCAAAGTGAATTTCAATATACACCTATGAGAGGCAATTCGATCTTGTCTCAGACAATCTA 3408
DCL2c 3414
DCL2b ATCAATGTCAGCTTCTAGAATCCCTCTGAAAGACTCGTTTAGGACCGTTCACTGTTAGTAGAAGCTATGACACAGGTTCTTACAGT 3498
DCL2c 3504
DCL2b CTGCCGATGTGCCAGATGCTATCAGGACTTGAATACTTAGGAGACCGGTTGGATATCTCATTACTACACATTTGTAATGAA 3588
DCL2c 3594
DCL2b CATCCTGGCATGACACCTGGCGAGTTAAGTGATATGCGGTCAGCTCTGTGAATAACGATGTTATGCAATGTCTGCCATAAGGTTCAA 3678
DCL2c 3684
DCL2b CTGCATAAACATGTACTCCATGCCCTCAGGAGTTACATAAACATATGCAAGTACCTTGATAAAATGAGCAACAGCTCTCATCACAA 3768
DCL2c 3774
DCL2b ACTTTTGGATGGGAGTCTGAGGCTTCAATCCCAAGGACTTGGAGACATCATAGAGTCTTAGCTGGTGTATATTGTTGATTGAGGA 3858
DCL2c 3864
DCL2b TACAACAAGGAGGTTGTGGCAAAGCATAAAGCCTCTTTTGAACCCCTTGTGACACCTGATACATTAAAGCATCCATCCGATCAGAGAA 3948
DCL2c 3954
DCL2b CTAACGAACTGTGCAGAAAAATGAAATATACATGGAGAAGACTTGTCCAGGAACGACGGAGTTACTAGCTGTAGAATGAGGTAAT 4038
DCL2c 4044
DCL2b GCAGATGGCATAAATTCATCAATACGAGTACAAAGGTTCTACTGACAAGAAAAACAGCTACAAGACTAGCCTGCAAGAGAGTTTGAATTC 4128
DCL2c 4134
DCL2b TTGCAGTTGAAGGAAACACAGACCAGTGA 4158
DCL2c 4164

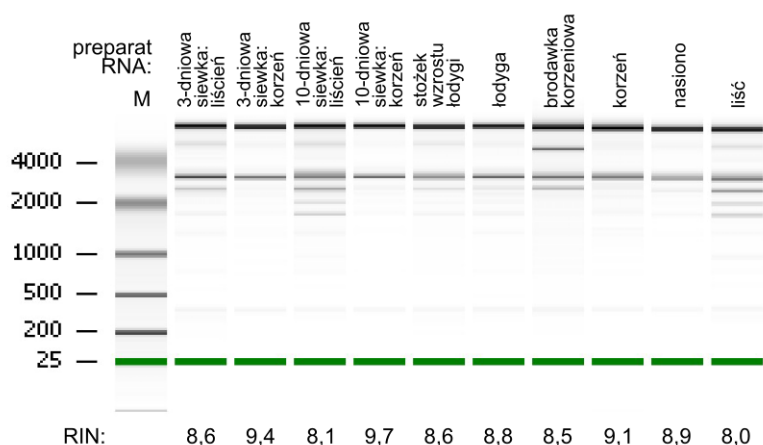


Rysunek 4.3. Porównanie organizacji domenowej sześciu białek DCL kodowanych przez genom *Medicago* i białek DCL z *Arabidopsis*. W ramach sekwencji każdego białka (kolor szary) zaznaczono kolejno lokalizację domeny DEXD/H i domeny końca C helikazy (obie kolor żółty), tworzących razem region helikazowy, domeny DUF283 (pomarańczowy), PAZ (różowy), RNazy IIIa i b (niebieski) oraz dsRBD (zielony). Lokalizację domen określono na podstawie analizy sekwencji aminokwasowych w bazach danych CDD (NCBI) i SMART (EMBL). Obok wskazano poziom identyczności sekwencji poszczególnych białek DCL z *Medicago* względem odpowiadającego im homologu z *Arabidopsis*.

IV.2. ANALIZA PROFILU EKSPRESJI GENÓW *MtDCL*

W celu określenia poziomu ekspresji sześciu zidentyfikowanych genów kodujących białka MtDCL dokonano izolacji całkowitego RNA z rośliny. Wykorzystany do izolacji RNA materiał obejmował: nasiona, 3- i 10-dniowe siewki (w obu przypadkach podzielone na liście oraz korzeń) oraz poszczególne części dorosłej rośliny (liść, łodygę, stożek wzrostu łodygi, korzeń i brodawkę korzeniową). Celem uniknięcia zanieczyszczenia genomowym DNA izolacja obejmowała etap inkubacji preparatu z DNazą. Otrzymane preparaty RNA poddano analizie ilościowej i jakościowej przy użyciu aparatu 2100 Bioanalyzer (Agilent). Do dalszych analiz wykorzystano preparaty, dla których tzw. współczynnik integralności (RIN, ang. *RNA integrity number*) wynosił co najmniej 8,0 (Rys. 4.4). W kolejnym etapie przygotowano preparaty cDNA odpowiadające puli całkowitego mRNA w badanym materiale roślinnym. Preparaty te otrzymano na drodze odwrotnej transkrypcji RNA izolowanego z poszczególnych części *Medicago* z użyciem startera oligo-dT.

Do ilościowej analizy akumulacji cząsteczek mRNA kodujących białka MtDCL w badanym materiale roślinnym wykorzystano metodę ddPCR. Pozwala ona na pomiar bezwzględnego stężenia cząsteczek o określonej sekwencji w badanym preparacie. Jak wykazano metodą tą, w porównaniu z analizami opartymi na PCR w czasie rzeczywistym (ang. *real-time PCR*), charakteryzuje wyższą precyzją pomiarów i

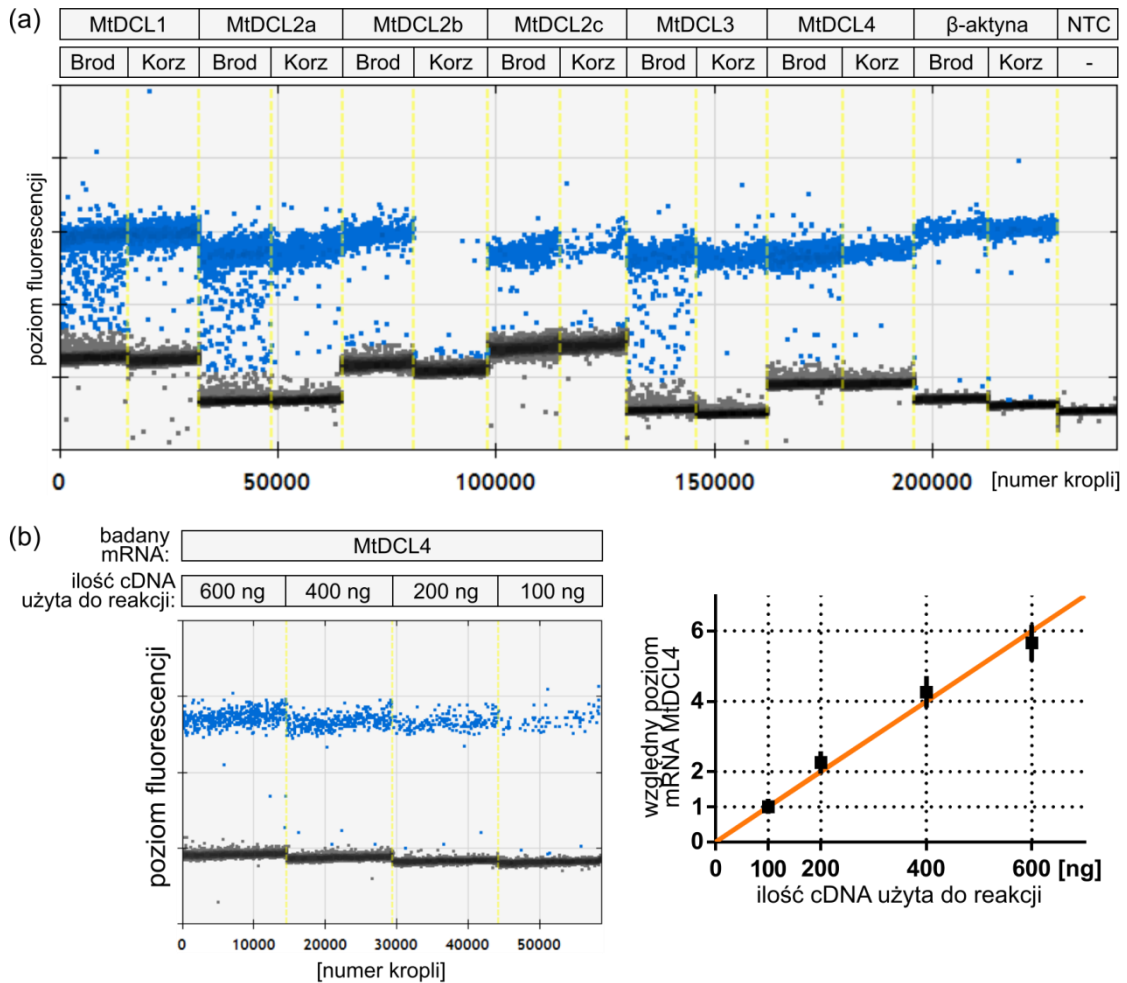


Rysunek 4.4. Analiza jakościowa otrzymanych preparatów RNA przy użyciu systemu 2100 Bioanalyzer (Agilent). Przedstawiono elektroforogramy uzyskane dla całkowitego RNA wyizolowanego z różnych części rośliny oraz RIN określony dla każdego z preparatów.

wyższa powtarzalność wyników [174]. Pary starterów stosowane w ddPCR, specyficzne wobec analizowanych sekwencji docelowych zaprojektowano z uwzględnieniem następujących kryteriów zalecanych przez producenta wykorzystywanego urządzenia (Bio-Rad): (i) długość amplikonu mieści się w przedziale 120-180 nukleotydów, (ii) temperatura topnienia dupleksu DNA utworzonego przez starter i sekwencję komplementarną wynosi 59°C (± 1) oraz (iii) co najmniej jeden starter z pary jest komplementarny do miejsca złączenia dwóch egzonów w mRNA. Stosowanie ostatniego kryterium miało na celu wyeliminowanie możliwości amplifikacji badanej sekwencji na matrycy genomowego DNA stanowiącego zanieczyszczenie preparatu RNA użytego w reakcji odwrotnej transkrypcji.

Przygotowane mieszaniny reakcyjne ddPCR podlegały frakcjonowaniu na około 15 tysięcy kropeł, w których zachodziły niezależne reakcje z udziałem polimerazy DNA. Losowy rozkład matrycy w kroplach powodował, że jedynie część kropeł zawierała docelową cząsteczkę, identyfikowaną i amplifikowaną za pomocą pary specyficznych starterów. Krople, w których nastąpiła amplifikacja były identyfikowane na podstawie fluorescencji obecnego w mieszaninie reakcyjnej i specyficznego wobec dsDNA barwnika (EvaGreen) i zliczane. Stanowiło to podstawę do ilościowej analizy badanej cząsteczki. Przykładowy wynik ddPCR ukazujący poziom akumulacji wszystkich badanych transkryptów w dwóch różnych częściach rośliny przedstawiono na Rysunku 4.5. Wyniki wszystkich pomiarów zamieszczono w Załączniku 2. Względny poziom akumulacji każdego z badanych mRNA znormalizowano względem poziomu akumulacji mRNA β -aktyny. Założono, że ekspresja genu β -aktyny zachodzi na jednolitym poziomie we wszystkich komórkach rośliny. Dzięki temu mRNA β -aktyny mógł zostać wykorzystany jako wzorzec pozwalający na korektę różnic pomiędzy porównywanymi próbkami. Różnice te mogły wynikać przykładowo z niedokładności pomiarów stężeń bądź innych niedokładności o charakterze technicznym.

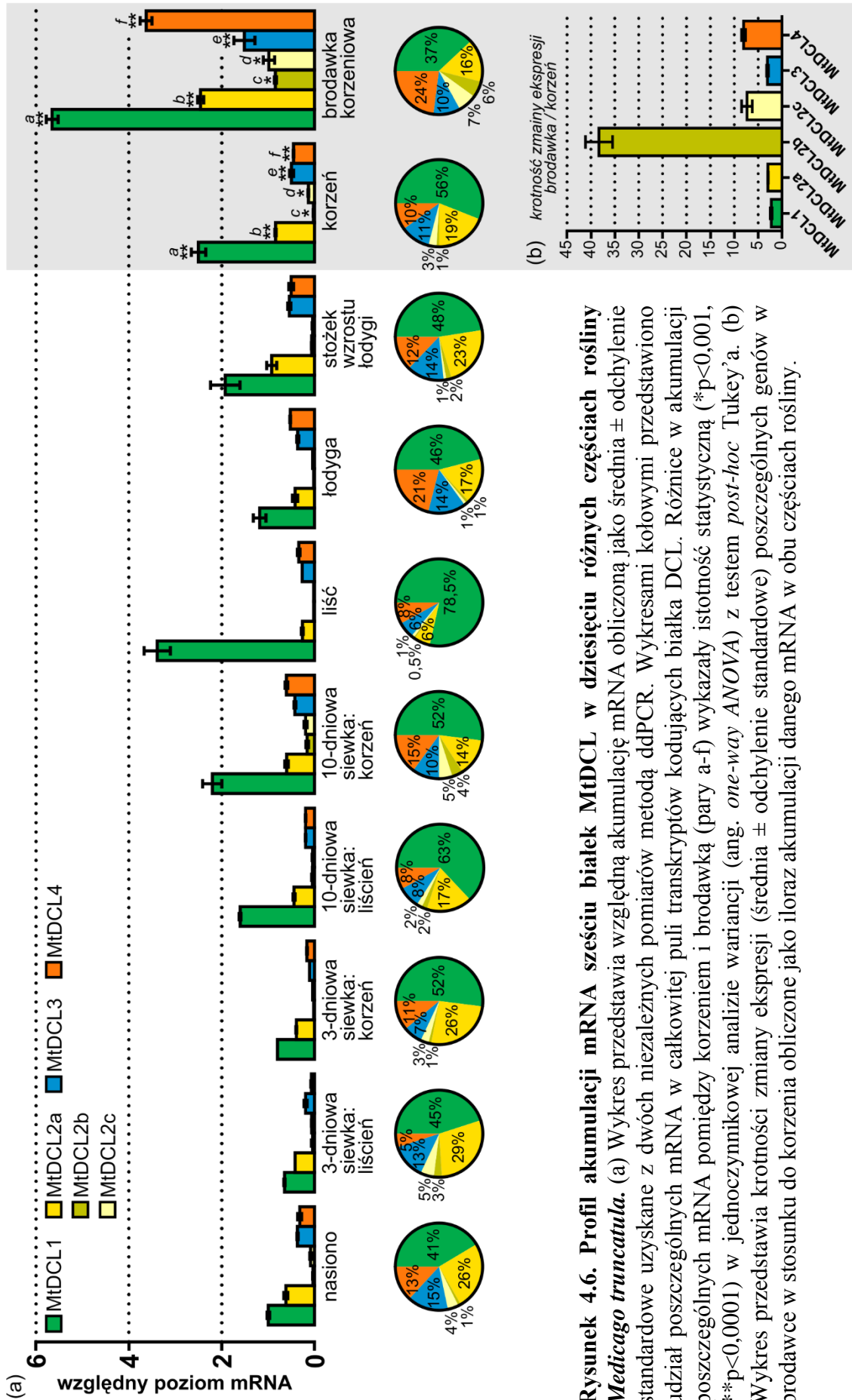
Analizy z wykorzystaniem ddPCR pozwoliły na identyfikację wszystkich transkryptów *MtDCL* w każdej badanej części rośliny. We wszystkich preparatach zaobserwowano podobny profil akumulacji mRNA *MtDCL*: najwyższy obserwowano zawsze dla mRNA *MtDCL1*, natomiast najniższy dla mRNA *MtDCL2b* i *MtDCL2c* (Rys. 4.6).



Rysunek 4.5. Analiza poziomu akumulacji mRNA białek DCL z *Medicago truncatula* z wykorzystaniem ddPCR. (a) Przykładowy zestaw wyników ddPCR uzyskanych dla wszystkich analizowanych mRNA (MtDCL1, 2a, 2b, 2c, 3, 4 i β -aktyny) w odniesieniu do dwóch różnych części rośliny: korzenia (Korz) i brodawki korzeniowej (Brod) *Medicago*. Wykresy przedstawiają poziom fluorescencji każdej kropli zmierzony po przeprowadzeniu ddPCR z zastosowaniem par starterów specyficznych wobec sekwencji danego mRNA. Na niebiesko zaznaczono tzw. krople pozytywne (w których nastąpiła amplifikacja docelowej sekwencji), a na szaro tzw. krople negatywne (uznane za nie zawierające sekwencji docelowej). NTC – reakcja kontrolna bez dodatku cDNA. (b) Przykładowy zestaw wyników ddPCR ilustrujący liniową zależność pomiędzy ilością cDNA użytego w reakcji, a zmierzonym względny poziomem badanego mRNA. Po lewej stronie przedstawiono wykres analogiczny jak powyżej, uzyskany w wyniku analizy akumulacji mRNA MtDCL4 metodą ddPCR z zastosowaniem różnej ilości cDNA uzyskanego na drodze odwrotnej transkrypcji całkowitego mRNA z liścia *Medicago* (odpowiednio 100, 200, 400 i 600 ng). Na wykresie po prawej stronie punktami zaznaczono względny poziom mRNA MtDCL4 określony na podstawie wykonanej analizy metodą ddPCR. Poziom mRNA w ddPCR określany jest z uwzględnieniem modelu losowej dystrybucji identyfikowanych cząsteczek w kroplach opisanej rozkładem Poissona. Słupkami zaznaczono 95% przedział ufności dla każdego oszacowania opartego o rozkład Poissona. Pomarańczowa linia na wykresie wyznacza rzeczywistą liniową zależność pomiędzy ilością preparatu cDNA użytego w reakcji, a względny poziomem badanego mRNA.

Porównując wyniki uzyskane dla poszczególnych części rośliny zaobserwowano, że najwyższy poziom akumulacji mRNA *MtDCL* cechuje brodawkę korzeniową. Największy, niemal 40-krotny wzrost akumulacji mRNA w brodawce względem korzenia zaobserwowano dla *MtDCL2b*. W przypadku *MtDCL2c* i *MtDCL4* zaobserwowano ponad 7-krotny wzrost akumulacji odpowiadających im mRNA, a dla pozostałych DCL około 3-krotny (Rys. 4.6b). Znacznie podwyższona ekspresja genów *MtDCL2b*, *MtDCL2c* i *MtDCL4* skutkuje wzrostem udziału tych mRNA w całkowitej puli transkryptów kodujących białka *MtDCL* w brodawce. W przypadku *MtDCL2b*, który ulega najniższej ekspresji we wszystkich badanych częściach rośliny udział odpowiadającego mu mRNA w całkowitej puli mRNA genów *MtDCL* wzrasta z 1% w korzeniu do 7% w brodawce. W przypadku genów *MtDCL2c* i *MtDCL4* zaobserwowano około dwukrotny wzrost udziału poszczególnych mRNA, odpowiednio z 3 do 7% i z 10 do 24%. Jednocześnie udział transkryptów genów *MtDCL2a* i *MtDCL3* pozostał niemal niezmienny, a w przypadku *MtDCL1* zmalał z 56 do 37%.

W dojrzałym liściu zaobserwowano natomiast znacznie podwyższoną akumulację mRNA *MtDCL1* względem pozostałych transkryptów. W efekcie mRNA *MtDCL1* stanowi ponad trzy czwarte puli wszystkich transkryptów *MtDCL* obecnych w liściu (Rys. 4.6).



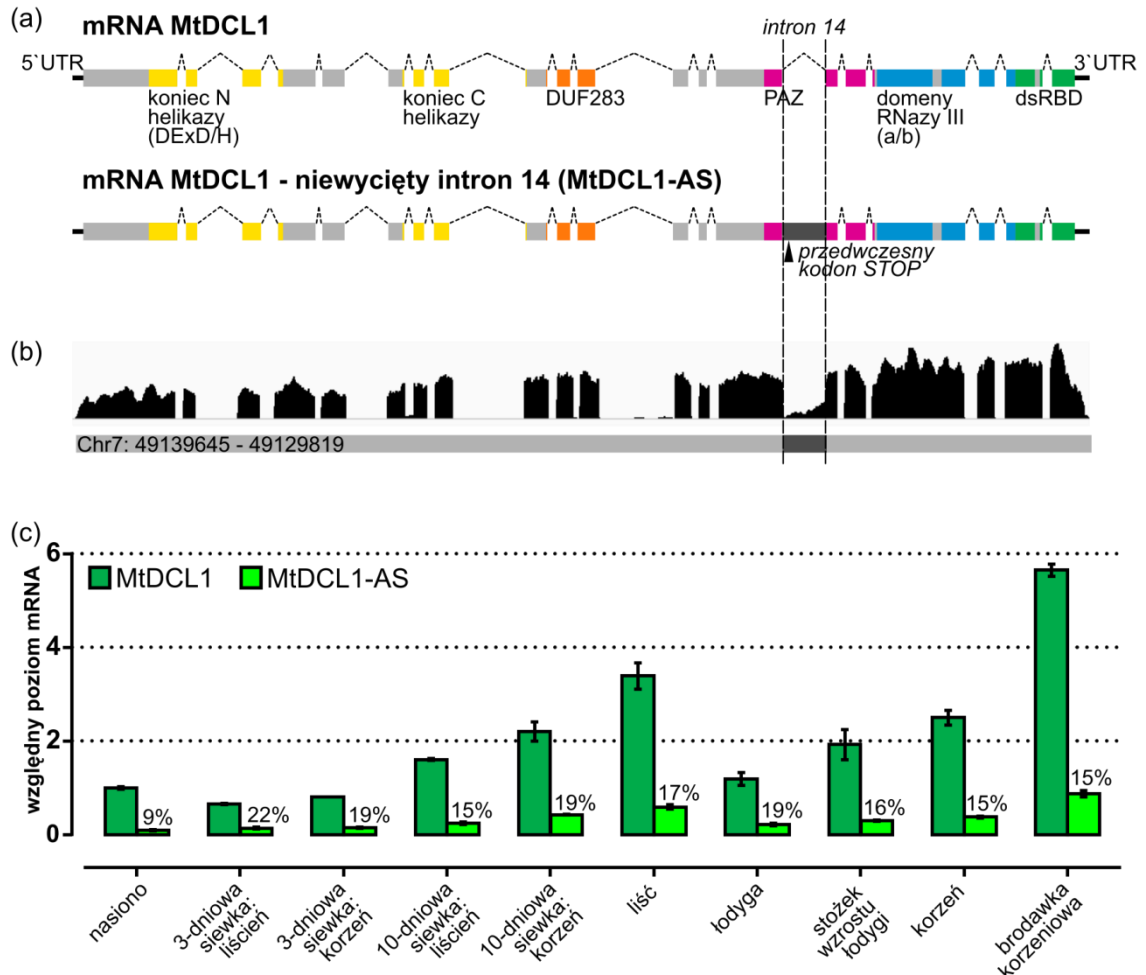
Rysunek 4.6. Profil akumulacji mRNA sześciu białek MtDCL w dziesięciu różnych częściach rośliny *Medicago truncatula*. (a) Wykres przedstawia względną akumulację mRNA obliczoną jako średnia \pm odchylenie standardowe uzyskane z dwóch niezależnych pomiarów metodą ddPCR. Wykresami kołowymi przedstawiono udział poszczególnych mRNA w całkowitej puli transkryptów kodujących białka DCL. Różnice w akumulacji poszczególnych mRNA pomiędzy korzeniem i brodawką (pary a-f) wykazały istotność statystyczną (* $p < 0,001$, ** $p < 0,0001$) w jednoczynnikowej analizie wariancji (ang. *one-way ANOVA*) z testem *post-hoc* Tukey'a. (b) Wykres przedstawia krotności zmiany ekspresji (średnia \pm odchylenie standardowe) poszczególnych genów w brodawce w stosunku do korzenia obliczone jako iloraz akumulacji danego mRNA w obu częściach rośliny.

IV.3. KLONOWANIE I CHARAKTERYSTYKA cDNA WYBRANYCH BIAŁEK MtDCL

Sekwencje kodujące białka MtDCL1 i MtDCL2a, wybrane do analiz ze względu na ich udział w biogenezie odmiennych typów małych regulatorowych RNA w roślinie (odpowiednio miRNA i siRNA), uzyskano metodą PCR. Jako matrycę wykorzystano otrzymany wcześniej preparat cDNA odpowiadający całkowitej puli mRNA obecnego w stożku wzrostu łodygi *Medicago* (Rozdz. IV.2). Produkty PCR wprowadzono do wektora wysokokopijnego pCR-XL-TOPO. Konstrukty przygotowano z wykorzystaniem enzymu topoiomerazy I kowalencyjnie przyłączonego do obu końców zliniowanego plazmidu. Otrzymano w ten sposób plazmidy pCR-DCL1 i pCR-DCL2a. Obecny w nich cDNA poddano sekwencjonowaniu, które w obu przypadkach potwierdziło obecność pełnej długości sekwencji kodującej właściwe białko DCL. Sekwencjom kodującym towarzyszyły fragmenty nieulegających translacji końców 5' i 3' mRNA (5' i 3' UTR, ang. *untranslated region*). Otrzymane sekwencje (Załącznik 3) porównano z sekwencjami pochodzącymi z bazy danych MTGP (Mt4.0) [165]. Stwierdzono, że przeprowadzone w ramach MTGP analizy struktury genów *MtDCL1* i *MtDCL2a* doprowadziły do poprawnego określenia granic pomiędzy intronami i egzonami. Sklonowane sekwencje kodujące MtDCL1 i MtDCL2a posiadały strukturę identyczną jak proponowana przez MTGP. Najistotniejszą zaobserwowaną różnicą było wydłużenie ciągu powtórzeń CAA w sklonowanej sekwencji kodującej białko MtDCL1 o dziewięć tripletów, co skutkuje obecnością dziewięciu dodatkowych asparagin: Asn139-147. Ponadto w przypadku obu białek występują pojedyncze substytucje nukleotydowe, niekiedy skutkujące substytucjami aminokwasowymi w poszczególnych pozycjach. Niemniej identyczność pomiędzy sekwencjami sklonowanymi, a obecnymi w bazie danych MTGP w przypadku zarówno MtDCL1 jak i MtDCL2 wynosiła około 99% niezależnie czy porównywano sekwencje nukleotydowe czy odpowiadające im sekwencje aminokwasowe.

Podczas klonowania sekwencji kodującej MtDCL1 zidentyfikowano także wariant cDNA zawierający 457-nukleotydowy intron 14 genu tego białka. Wskazało to na możliwość występowania w roślinie alternatywnej formy mRNA, oznaczonej jako MtDCL1-AS. Analiza sekwencji zidentyfikowanego wariantu cDNA wykazała, że obecność intronu 14 powoduje pojawienie się przedwczesnego kodonu STOP w ramce odczytu, w obrębie sekwencji kodującej domenę PAZ (Rys. 4.7a). Dane pochodzące z sekwencjonowania RNA z *Medicago*, stanowiące część bazy danych MTGP [165]

wskazują, że intron 14 stosunkowo często pozostaje niewycięty z pre-mRNA w czasie obróbki potranskrypcyjnej (Rys. 4.7b).



Rysunek 4.7. Alternatywne składanie mRNA MtDCL1. (a) Struktura dojrzałego mRNA MtDCL1 oraz mRNA z niewyciętym intronem 14 (MtDCL1-AS). W ramach sekwencji kodującej białko MtDCL1 zaznaczono sekwencje kodujące domenę DExD/H i domenę końca C helikazy (obie kolor żółty), domenę DUF283 (pomarańczowy), PAZ (różowy), RNazy IIIa i b (niebieski), dsRBD (zielony) oraz rejonów międzysdomenowych (kolor szary). Obecność intronu 14 skutkuje powstaniem przedwczesnego kodonu STOP w ramce odczytu kodującej MtDCL1 – w obrębie sekwencji kodującej domenę PAZ (strzałka). (b) Przykład ilustrujący pokrycie sekwencji genomowej w obszarze genu MtDCL1 danymi pochodzącymi z sekwencjonowania RNA (dane zdeponowane w bazie danych MTGP) [165]. Obecność odczytów odpowiadających sekwencji intronu 14 sugeruje występowanie alternatywnej formy mRNA MtDCL1 obejmującej ten intron. (c) Profil akumulacji MtDCL1-AS oraz całkowitej frakcji mRNA MtDCL1 (transkryptów posiadających jak i pozbawionych intronu 14) w dziesięciu różnych częściach rośliny *Medicago truncatula*. Wykres przedstawia względną akumulację mRNA obliczoną jako średnia \pm odchylenie standardowe uzyskane z dwóch niezależnych pomiarów metodą ddPCR. Wskazano także procentowy udział formy MtDCL1-AS w całkowitej frakcji mRNA MtDCL1.

Metodą ddPCR potwierdzono występowanie mRNA MtDCL1-AS we wszystkich analizowanych wcześniej częściach rośliny i określono poziom jego akumulacji. Analizę przeprowadzono analogicznie jak miało to miejsce w przypadku pozostałych badanych mRNA (Rozdz. IV.2). Najwyższą akumulację mRNA MtDCL1-AS zaobserwowano w brodawce korzeniowej. Wynik ten jest zgodny z obserwacjami dotyczącymi pozostałych transkryptów kodujących białka MtDCL. Określono także procentowy udział MtDCL1-AS w całkowitej frakcji mRNA MtDCL1, która obejmowała zarówno transkrypty pozbawione jak i posiadające intron 14. W nasionach udział ten wyniósł 9%, co stanowiło najniższy obserwowany wynik, natomiast w pozostałych analizowanych częściach rośliny wahał się w granicach 15-22%. Średni udział dla wszystkich pomiarów wyniósł około 16% (Rys. 4.7c).

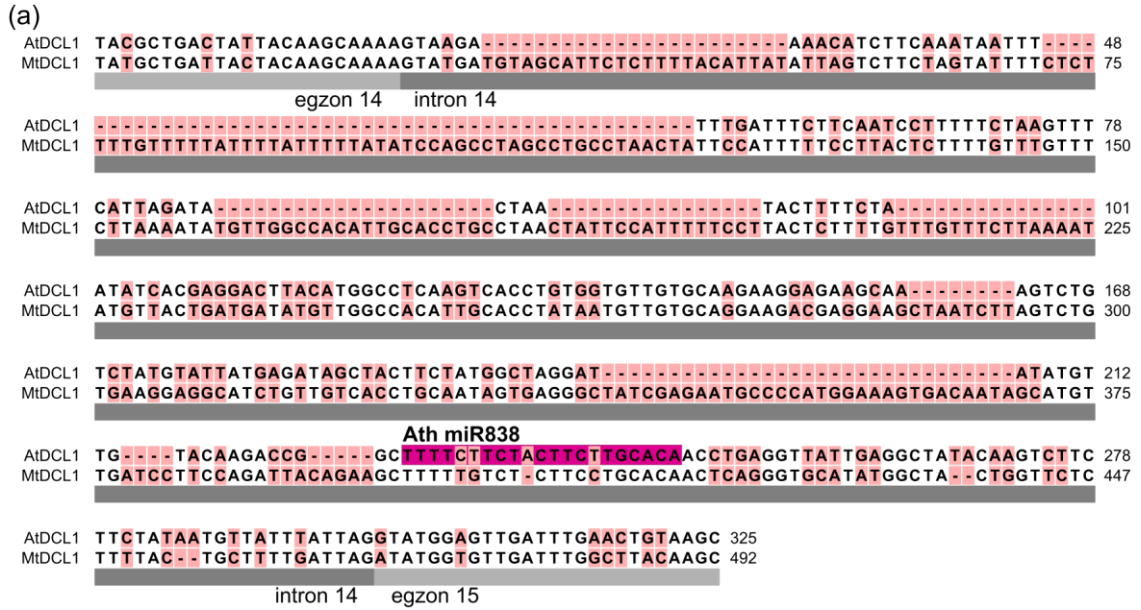
Występowanie alternatywnej formy mRNA DCL1 opisano także u *Arabidopsis thaliana*. Forma ta, podobnie jak w przypadku wariantu mRNA zidentyfikowanego u *Medicago* obejmuje niewycięty intron 14, co skutkuje pojawieniem się kodonu STOP w sekwencji kodującej domenę PAZ białka. Jak wykazano, ta alternatywna forma mRNA AtDCL1 stanowi także prekursor roślinnego miRNA miR838. Dupleks miRNA/miRNA* wycinany jest przez białko AtDCL1 ze struktury spinki kodowanej przez intron 14. Prowadzi to do powstania dwóch fragmentów mRNA, co potwierdzono techniką 5` RACE i 3` RACE. Ponadto obecność dojrzałego miR838 u *Arabidopsis* została potwierdzona przez głębokie sekwencjonowanie RNA [70, 175]. Uważa się, że poprzez biogenezę miR838 białko AtDCL1 bierze udział w regulacji poziomu własnego mRNA, jak dotąd nie jest natomiast znany żaden transkrypt podlegający bezpośredniej regulacji z udziałem miR838.

Aby przekonać się, czy podobne zjawisko występuje u *M. truncatula* porównano introny 14 genów DCL1 z *Medicago* i *Arabidopsis*. Wykazano, że pomimo istotnych różnic sekwencyjnych intron z *Medicago* zawiera sekwencję homologiczną do Ath miR838, która także ulokowana jest w obszarze tworzącym dużą strukturę typu spinka do włosów (Rys. 4.8). W celu sprawdzenia czy intron 14 obecny w MtDCL1-AS obejmuje sekwencję jakiegokolwiek znanego miRNA z *Medicago* przeszukano bazy danych miRBase (wydanie 20) [176]. Wśród zgromadzonych w miRBase miRNA z *Medicago* nie znaleziono jednak takiego, który potencjalnie mógłby pochodzić z sekwencji intronu 14.

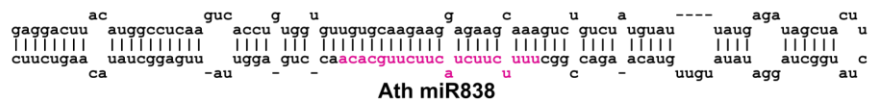
Podjęto także próbę identyfikacji cząsteczek RNA, które mogłyby stanowić produkty trawienia mRNA MtDCL1-AS w obrębie sekwencji odpowiadającej intronowi

14. W tym celu przeprowadzono analizę 5' RACE z wykorzystaniem preparatu całkowitego RNA izolowanego z liścia *Medicago*. Preparat ten wybrano ze względu na fakt, iż w liściu zaobserwowano wysoki poziom akumulacji mRNA MtDCL1-AS, w porównaniu do innych części rośliny. Analiza 5' RACE obejmowała syntezę cDNA z użyciem startera specyficznego wobec mRNA MtDCL1, poliadenylację otrzymanego cDNA, a następnie PCR ze starterami specyficznymi wobec MtDCL1 oraz ogona poli-A. Otrzymanie produktów o długości około 70-520 par zasad (pz) świadczyłoby o obecności w analizowanym preparacie cząsteczek RNA pochodzących z transkryptu MtDCL1-AS, których koniec 5' znajdowałby się w obrębie sekwencji odpowiadającej intronowi 14 (Rys. 4.8c). Wykonana analiza nie doprowadziła jednak do identyfikacji transkryptów będących produktami trawienia MtDCL1-AS w obrębie sekwencji odpowiadającej intronowi 14.

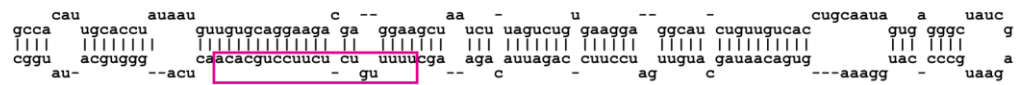
Rysunek 4.8 (następna strona). Alternatywny wariant mRNA DCL1 występujący u *Medicago truncatula* i *Arabidopsis thaliana*. (a) Porównanie sekwencji intronu 14, która obecna jest w alternatywnej formie mRNA DCL1 u *Medicago* i *Arabidopsis*. W intronie pochodzącym z *Arabidopsis* zaznaczono sekwencję kodującą miR838. (b) Porównanie struktury drugorzędowej fragmentu RNA pochodzącego z intronu 14 genu *AtDCLI* (obejmującego sekwencję miR838) oraz homologicznego fragmentu intronu 14 genu *MtDCLI*. W obrębie struktur zaznaczono sekwencję miR838 z *Arabidopsis* i homologiczną sekwencję z *Medicago*. Struktury określono za pomocą programu RNAfold [177]. (c) Schemat reakcji 5' RACE prowadzonej w celu identyfikacji cząsteczek RNA, które mogłyby stanowić produkty trawienia mRNA MtDCL1-AS w obrębie sekwencji odpowiadającej intronowi 14. Na kolejnych etapach reakcji wykorzystywano trzy startery komplementarne do sekwencji egzonu 15 oraz dwa startery pozwalające na amplifikację końca 5' identyfikowanej cząsteczki RNA (oznaczone jako RACE 1 i RACE 2). Procedura obejmowała kolejno reakcję odwrotnej transkrypcji, degradację RNA, poliadenylację końca 3' otrzymanych produktów oraz dwie serie PCR z wykorzystaniem różnych par starterów. Otrzymanie produktów o długości około 70-520 par zasad (pz) świadczyłoby o obecności cząsteczek RNA, których koniec 5' znajdowałby się w obrębie sekwencji intronu 14.



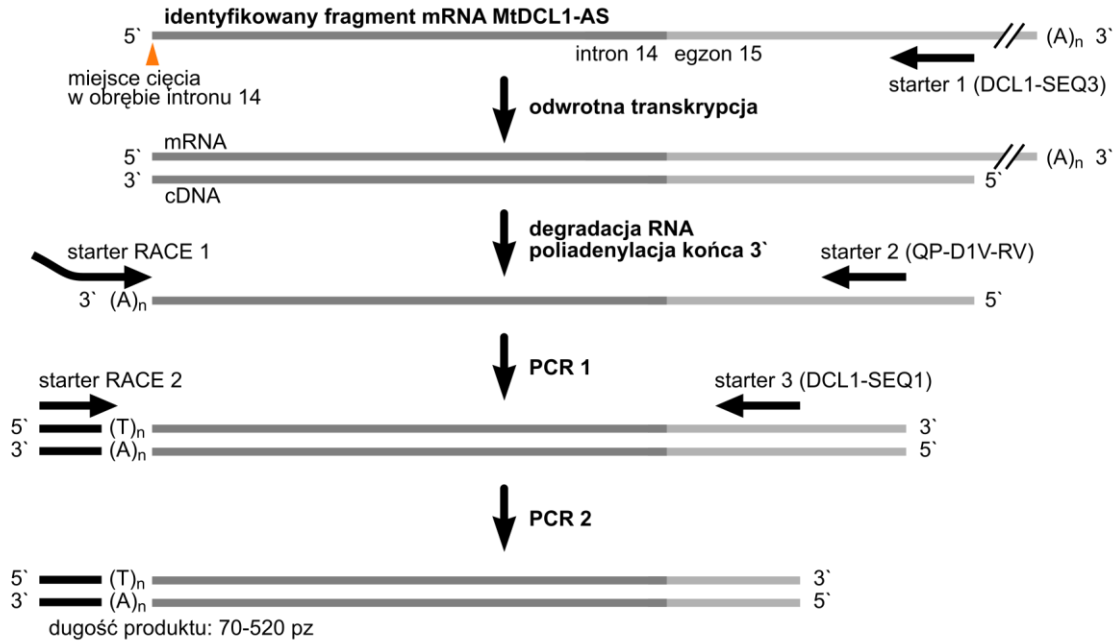
(b) struktura spinki w intronie 14 obecnym w alternatywnym wariantcie mRNA AtDCL1



struktura spinki w intronie 14 obecnym w MtDCL1-AS



(c) 5'RACE



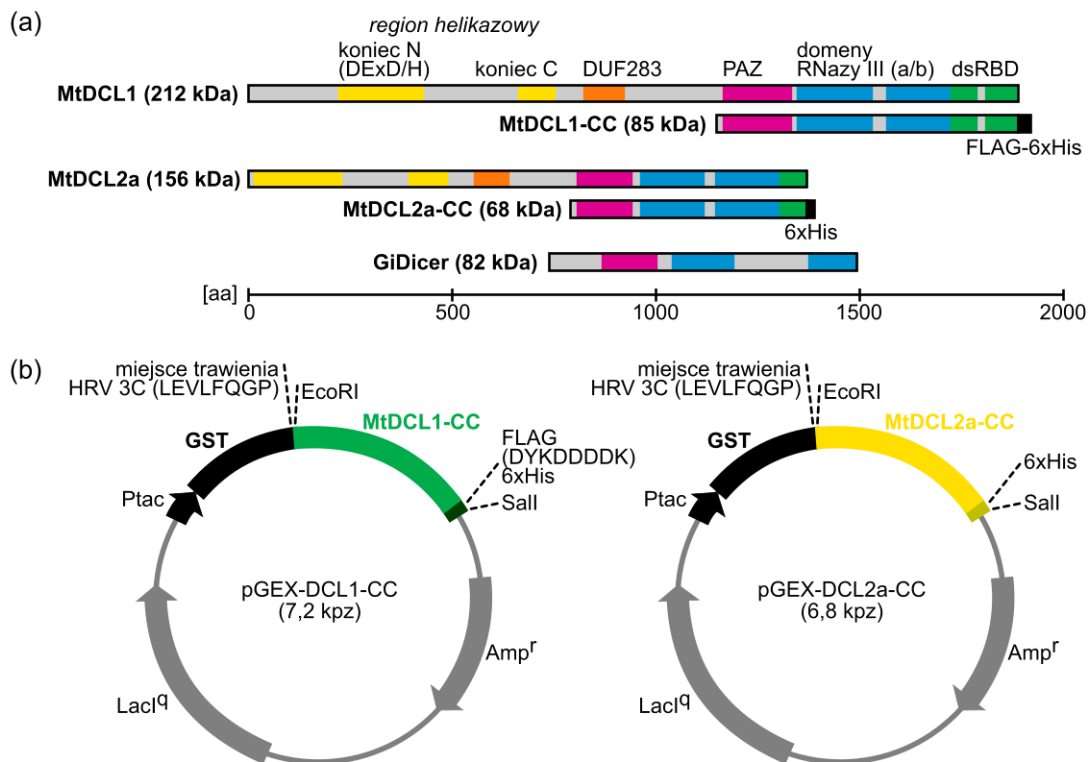
IV.4. PRODUKCJA BIAŁEK MtDCL1 I MtDCL2A

W pierwszym etapie badań podjęto próbę produkcji pełnej długości białek MtDCL1 i MtDCL2a w bakteryjnym systemie ekspresyjnym. Konstrukty do ekspresji przygotowano w oparciu o wektor pET100/D-TOPO przeznaczony do produkcji białek fuzyjnych posiadających 6-aminokwasowy znacznik histydynowy oraz 8-aminokwasowy znacznik FLAG na końcu aminowym. Obecność znacznika histydynowego w białku fuzyjnym pozwala na zastosowanie metody oczyszczania opartej o powinowactwo do jonów niklu, natomiast znacznik FLAG pozwala na specyficzną identyfikację białka metodą immunodetekcji. Sekwencje kodujące białka DCL amplifikowano metodą PCR, jako matrycę stosując wcześniej otrzymane wektory pCR-DCL1 i pCR-DCL2a. Konstrukty ekspresyjne pET-DCL1 i pET-DCL2a przygotowano, analogicznie jak w przypadku pCR-XL-TOPO, z wykorzystaniem enzymu topoizomerazy I. Następnie poddano je sekwencjonowaniu, potwierdzając poprawność obecnego w nich cDNA: zachowanie ramki odczytu i brak mutacji wynikających z działania polimerazy DNA. Oba konstrukty wykorzystano do transformacji komórek *Escherichia coli* szczepu przeznaczonego do produkcji białek (BL21-STAR). W obu przypadkach podjęto próbę uzyskania białka w bakteriach, testując szereg warunków hodowli. Warunki podlegające modyfikacji obejmowały temperaturę i czas hodowli oraz stężenie IPTG tj. czynnika indukującego ekspresję heterologicznego genu. Z bakterii izolowano następnie frakcje białek rozpuszczalnych, nierozpuszczalnych oraz tworzących ciała inkluzyjne, które analizowano metodą SDS-PAGE i immunodetekcji z wykorzystaniem monoklonalnych przeciwciał anti-FLAG. Pomimo przetestowania różnych warunków hodowli nie udało się wyprodukować pełnej długości białek MtDCL1 i MtDCL2a.

Wobec trudności z produkcją MtDCL1 i MtDCL2a w bakteriach prowadzone prace skupiono na opracowaniu wydajnej metody produkcji białek skróconych. Postanowiono zaprojektować białka pozbawione niektórych domen funkcjonalnych, a wciąż wykazujące podstawową aktywność katalityczną charakterystyczną dla rybonukleaz typu Dicer. Wzorzec stanowiło pełnej długości białko Dicer z *Giardia intestinalis*, które w swej strukturze posiada jedynie domeny PAZ i RNazy III, a przy tym jest aktywne katalitycznie [140]. Zaprojektowano białka stanowiące tzw. trzony katalityczne: posiadające domenę PAZ, dwie domeny RNazy III oraz natywny koniec karboksylowy obejmujący jedną (MtDCL1) bądź dwie (MtDCL2a) domeny dsRBD

(Rys. 4.9a). Przewidywana masa cząsteczkowa tak zaprojektowanych białek (nazwanych MtDCL1-CC i MtDCL2a-CC) wyniosła odpowiednio 86 i 69 kDa.

Aby uzyskać konstrukty do produkcji trzonów katalitycznych białek DCL wykorzystano wektor pGEX-6P-3. Wektor ten służy do uzyskiwania białek fuzyjnych zawierających S-transferazę glutationu (GST, ang. *glutathione S-transferase*) na końcu aminowym, która może być odcięta z zastosowaniem proteazy HRV 3C. Obecność GST w białku fuzyjnym zwiększa jego rozpuszczalność, wydajność produkcji i pozwala na stosowanie metody oczyszczania opartej na powinowactwie do glutationu. W celu poszerzenia możliwości identyfikacji i oczyszczania białek, do sekwencji kodującej trzon katalityczny białek DCL dodano odcinki kodujące znacznik FLAG i znacznik histydynowy (MtDCL1-CC) lub sam znacznik histydynowy (MtDCL2a-CC) na końcu karboksylowym (Rys. 4.9a).

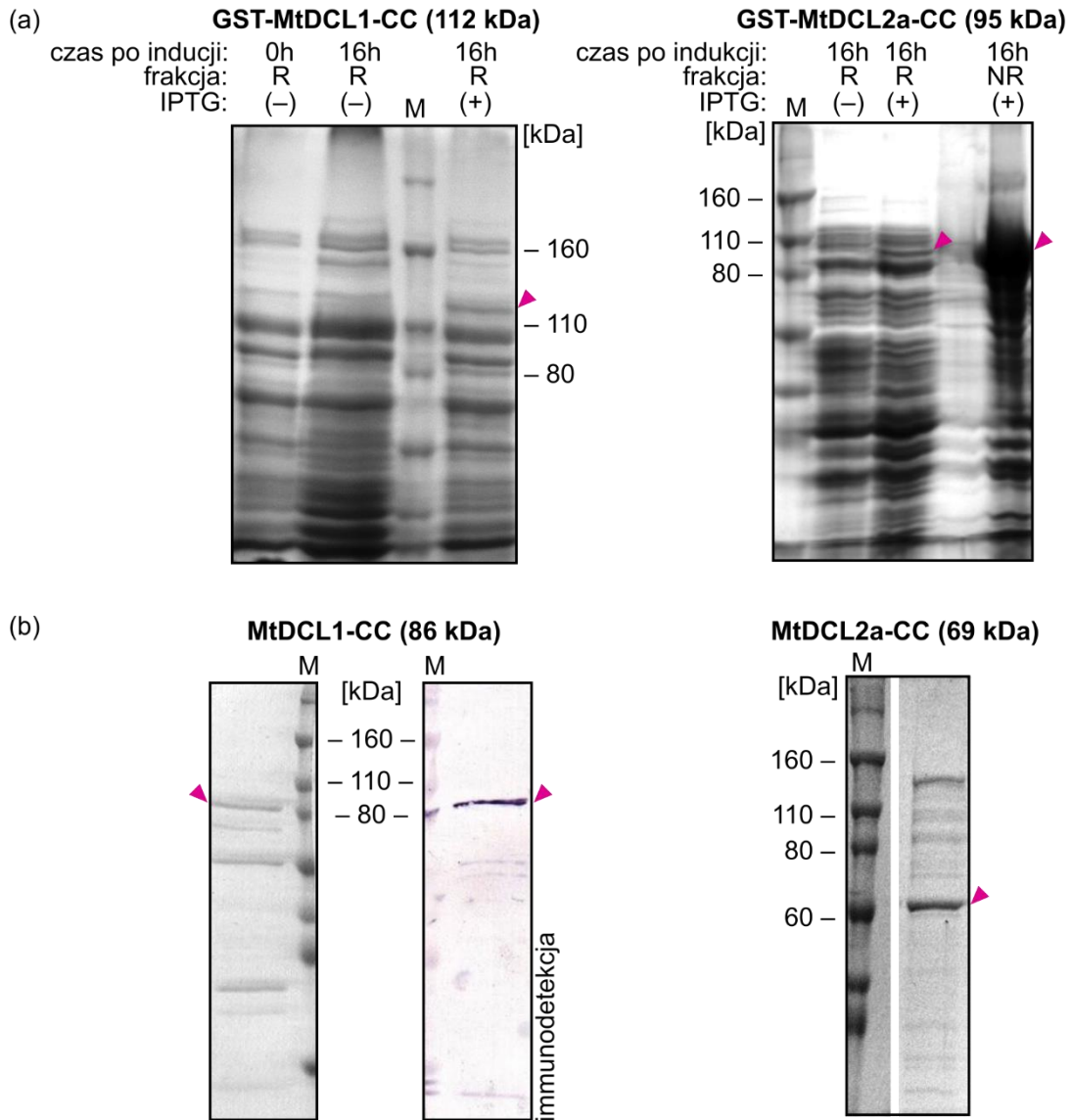


Rysunek 4.9. (a) Porównanie kompozycji domenowej pełnej długości białek MtDCL1 i MtDCL2a, zaprojektowanych trzonów katalitycznych MtDCL1-CC i MtDCL2a-CC oraz białka Dicer z *Giardia intestinalis* (GiDicer). (b) Struktura wektorów ekspresyjnych pGEX-DCL1-CC i pGEX-DCL2a-CC do produkcji białek MtDCL1-CC i MtDCL2a-CC. Oznaczono: promotor *tac* (Ptac), gen oporności na ampicylinę (Amp^r), gen represora promotora *tac* (LacI^Q), elementy kodowanego białka fuzyjnego i wykorzystywane miejsca restrykcyjne (EcoRI i Sall).

Sekwencje kodujące MtDCL1-CC i MtDCL2a-CC amplifikowano metodą PCR na matrycy wcześniej otrzymanych wektorów pCR-DCL1 i pCR-DCL2a. Zastosowano startery wprowadzające na końcach cDNA miejsca restrykcyjne (EcoRI lub SalI) oraz sekwencje znaczników: FLAG i histydynowego (MtDCL1-CC) lub tylko histydynowego (MtDCL2a-CC). Do konstrukcji wektorów pGEX-DCL1-CC i pGEX-DCL2a-CC (Rys. 4.9b) wykorzystano restryktazy EcoRI i SalI oraz ligazę DNA. Przygotowane konstrukty do ekspresji poddano sekwencjonowaniu, a następnie wykorzystano do transformacji komórek *E. coli* BL21-STAR.

W celu uzyskania wydajnej produkcji białek GST-MtDCL1-CC i GST-MtDCL2a-CC w przygotowanych szczepach, dokonano optymalizacji warunków hodowli. W przypadku GST-MtDCL1-CC głównym celem optymalizacji było podwyższenie wydajności produkcji białka. Białko to ulegało bowiem syntezie ze stosunkowo niską wydajnością, niemniej jednak cała jego pula obecna była we frakcji białek rozpuszczalnych w komórkach. Natomiast w przypadku GST-MtDCL2a-CC optymalizacja warunków hodowli prowadzona była w celu uzyskania jak największej puli białka rozpuszczalnego w komórkach, gdyż główny problem stanowiła akumulacja białka w formie nierozpuszczalnej (Rys. 4.10a). Ostatecznie w obu przypadkach wybrano następujące warunki hodowli: zawiesina bakterii przed dodaniem IPTG schładzana była do 18°C, produkcję białka prowadzono w 18°C przez 16 godzin.

Po izolacji przeprowadzono wiązanie białka fuzyjnego do złoża zawierającego immobilizowany glutation. Związane na złożu białko poddano trawieniu z wykorzystaniem rekombinowanej proteazy HRV 3C, posiadającej domenę GST, podobnie jak GST-MtDCL1-CC i GST-MtDCL2a-CC. Po odcięciu od partnera fuzyjnego rekombinowane białka MtDCL1-CC i MtDCL2a-CC wypłukano ze złoża, do którego związana pozostała proteaza oraz sam partner fuzyjny (GST). Otrzymane w ten sposób preparaty obu białek zagęszczono przez wirowanie na filtrze membranowym. Wyniki analizy SDS-PAGE i immunodetekcji otrzymanych preparatów MtDCL1-CC i MtDCL2a-CC przedstawiono na Rysunku 4.10b. Obecność w preparatach białek DCL potwierdzono także przez ich identyfikację techniką spektrometrii mas.



Rysunek 4.10. Produkcja rekombinowanych białek MtDCL1-CC i MtDCL2a-CC.

(a) Wynik analizy SDS-PAGE frakcji białek rozpuszczalnych (R) i nierozpuszczalnych (NR) izolowanych ze szczepów ekspresyjnych przeznaczonych do produkcji MtDCL1-CC i MtDCL2a-CC w formie białek fuzyjnych z GST. Białka izolowane były z hodowli w momencie dodania induktora (0h) oraz po 16 godzinach od indukcji ekspresji (16h). Hodowle prowadzono w 18°C, równoległe z i bez dodatku IPTG. (b) Wynik analizy SDS-PAGE oczyszczonych i zagęszczonych preparatów białek MtDCL1-CC i MtDCL2a-CC oraz analizy preparatu MtDCL1-CC metodą immunodetekcji z wykorzystaniem przeciwciał anti-FLAG. M – wzorec masy cząsteczkowej białek. Strzałkami zaznaczono prążki odpowiadające produkowanym białkom DCL.

IV.5. BADANIE AKTYWNOŚCI KATALITYCZNEJ BIAŁEK MTDCL1 I MTDCL2A

Otrzymane białka rekombinowane, stanowiące trzony katalityczne białek MtDCL1 i MtDCL2a, poddano testom enzymatycznym, w celu scharakteryzowania ich aktywności rybonukleazowej. W prowadzonych analizach wykorzystano substraty odpowiadające pre-miRNA powstającym w procesie biogenezy dwóch roślinnych miRNA (miR399c i miR399o z *Medicago*) oraz ludzkiego miR33a. Dodatkowo na bazie sekwencji Hsa pre-miR33a zaprojektowano substrat będący w pełni komplementarnym dwuniciowym RNA. Jedna z nici zawierała sekwencję miRNA oraz 11-nukleotydowy fragment, który w Hsa pre-miR33a tworzy trzon spinki. Druga nić substratu była w pełni komplementarna. Na obu końcach dupleksu występowały 2-nukleotydowe wystające końce 3'. Substrat ten uzyskano między innymi w celu określenia wpływu niesparowań obecnych w strukturze prekursora na pulę generowanych produktów.

Jak zaznaczono wcześniej, w biogenezie zwierzęcych miRNA rybonukleaza Drosha trawi pri-miRNA generując pre-miRNA o charakterystycznej strukturze typu spinki. Trawienie to następuje w obrębie końca dupleksu miRNA/miRNA* ulokowanego bliżej podstawy spinki prekursora. Natomiast u roślin za generowanie pre-miRNA, jak i dojrzałych miRNA, odpowiada rybonukleaza DCL1. Enzym ten rozpoczynać może trawienie struktury spinki obecnej w pri-miRNA zarówno u jej podstawy jak i w obszarze pętli końcowej, co determinuje strukturę powstającego pre-miRNA (prekursora, z którego bezpośrednio wycinany jest dupleks miRNA/miRNA*). W pierwszym przypadku powstający pre-miRNA utworzony jest przez jedną nić RNA przyjmującą strukturę spinki, podobnie jak ma to miejsce u zwierząt. W drugim przypadku pre-miRNA utworzony jest przez dwie częściowo komplementarne nici RNA (Rys. 4.11).

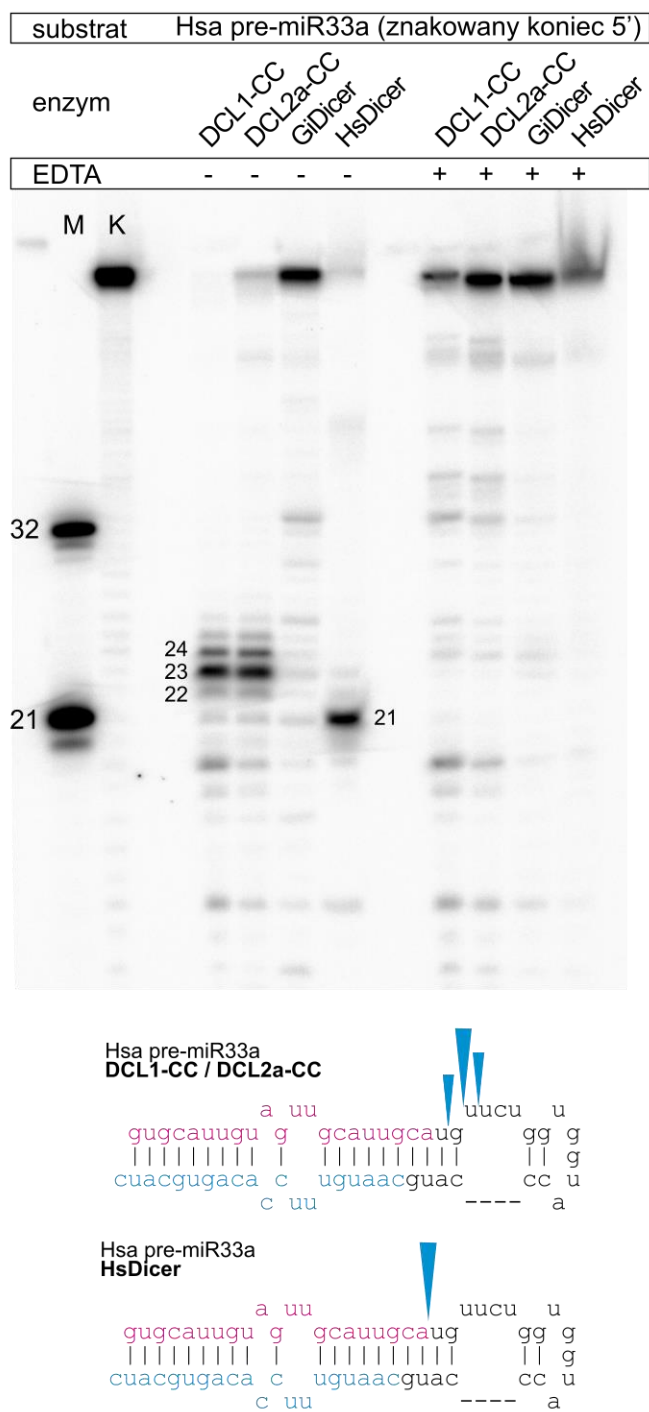
Jak wykazano u *Arabidopsis* trawienie pri-miRNA z rodziny miR399 następuje w kierunku od podstawy do pętli struktury spinki obecnej w prekursorze [162]. Oznacza to, że pre-miR399 utworzony jest przez jedną cząsteczkę RNA, która przyjmuje strukturę spinki obejmującą dupleks miRNA/miRNA*. Zgodnie z tym założeniem zaprojektowano substraty Mtr pre-miR399c i Mtr pre-miR399o.

Cząsteczki RNA użyte do analiz otrzymano na drodze syntezy chemicznej (Hsa pre-miR33a i dupleks RNA 33a) lub transkrypcji *in vitro* (Mtr pre-miR399c i Mtr pre-miR399o) i znakowano przez fosforylację końca 5' RNA z zastosowaniem

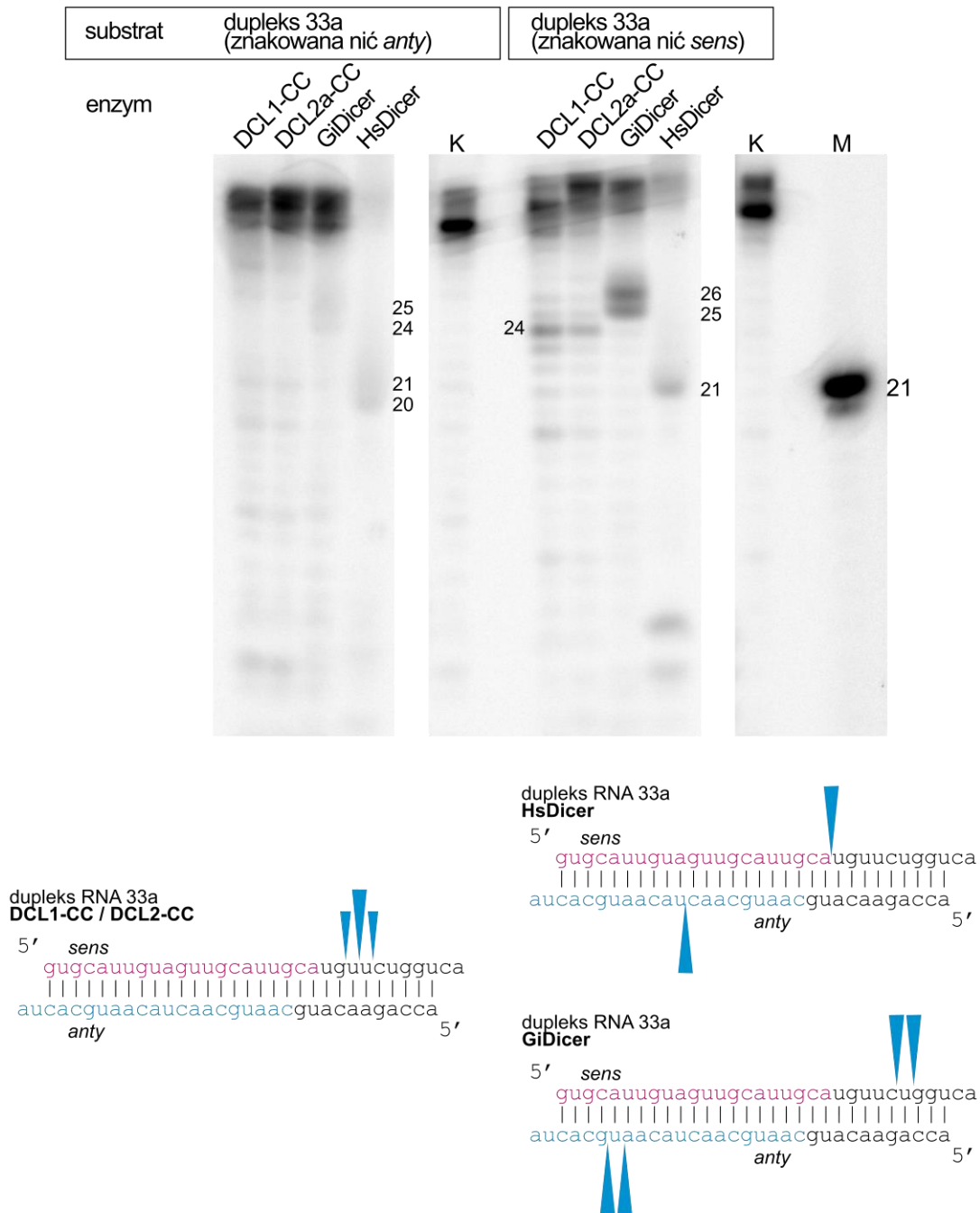
radioizotopu fosforu ^{32}P . Do przygotowania substratu w postaci dupleksu RNA 33a stosowano komplementarne cząsteczki RNA nazwane *sens* i *anty*, z których zawsze tylko jedna była znakowana izotopem. Strukturę stosowanych substratów przedstawia Rysunek 4.11.

Reakcje z rekombinowanymi białkami MtDCL1-CC i MtDCL2-CC prowadzono w obecności jonów Mg^{2+} , które warunkują aktywność katalityczną enzymów typu rybonukleaza III [178]. W celach kontrolnych analogiczne reakcje prowadzono z dodatkiem EDTA, który tworzy kompleksy z jonami Mg^{2+} . Zapobiega to ich wiązaniu przez enzym, blokując tym samym jego aktywność katalityczną. Dla porównania przeprowadzono także reakcje z wykorzystaniem preparatów rekombinowanych białek Dicer z człowieka (HsDicer) oraz *Giardia intestinalis* (GiDicer), jedynych preparatów białek Dicer dostępnych w sprzedaży. W celu analizy znakowanych radioizotopowo produktów reakcji rozdzielano je w denaturującym żelu poliakrylamidowym i wizualizowano metodą autoradiografii. Analizy z wykorzystaniem poszczególnych substratów RNA powtarzano przynajmniej dwukrotnie. Otrzymane wyniki przedstawiono na Rysunkach 4.12-4.15.

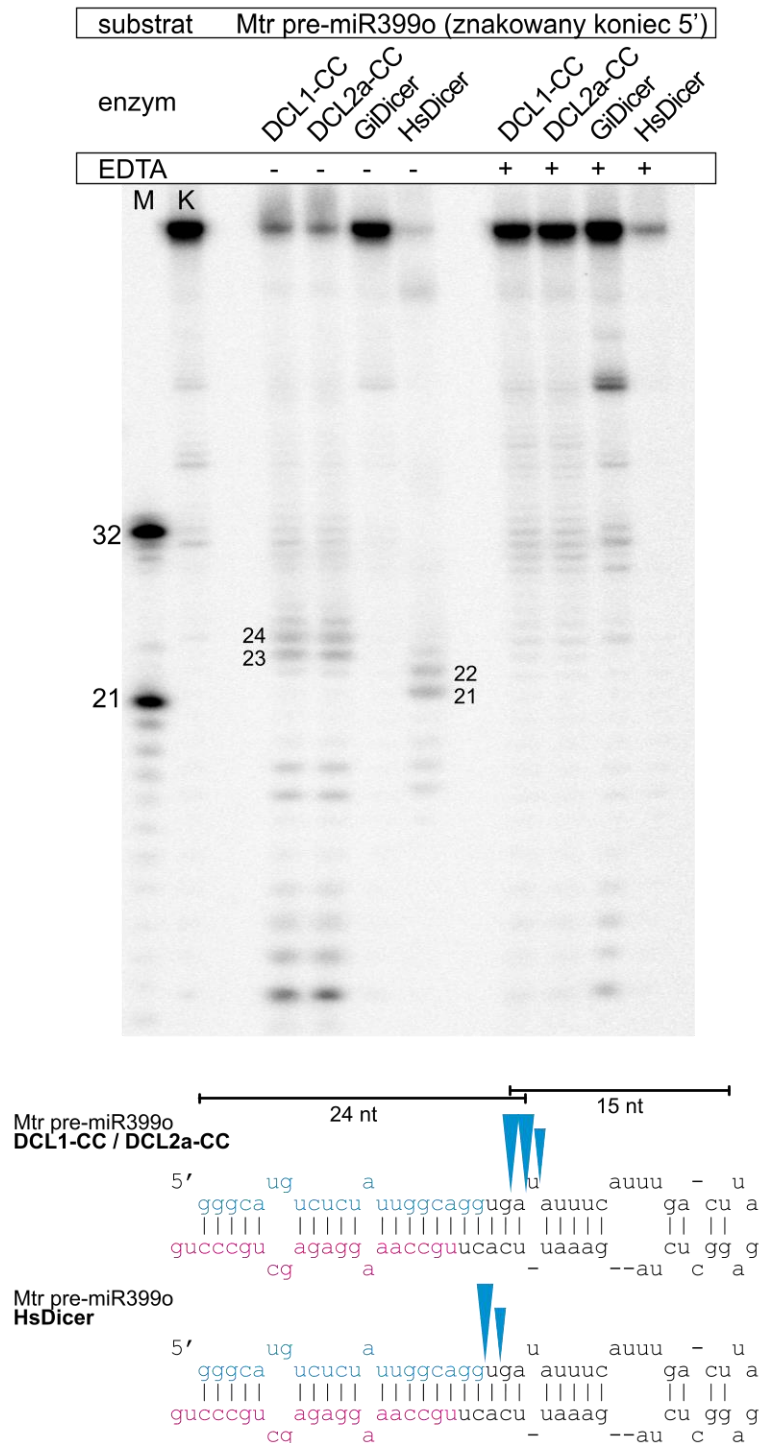
Spodziewany główny produkt reakcji stanowiły dupleksy miRNA/miRNA*. Z uwagi na sposób znakowania substratów tylko jedna nić dupleksu podlegała wizualizacji autoradiograficznej. Jak wcześniej wspomniano roślinne białka DCL1 uczestniczą w biogenezie 21-nukleotydomi miRNA, natomiast DCL2 zaangażowane są głównie w produkcję 22-nukleotydomi NAT-siRNA [64]. Stąd oczekiwane produkty reakcji z udziałem DCL1-CC i DCL2a-CC stanowiły cząsteczki o różnej długości, odpowiednio 21 i 22 nukleotydów. Z doniesień literaturowych wiadomo, że stosowane w celach porównawczych enzymy HsDicer i GiDicer generują w warunkach *in vitro* produkty o długości odpowiednio około 21 i 25 nukleotydów [140, 179]. W obu przypadkach odpowiada to sytuacji obserwowanej *in vivo* [176, 180].



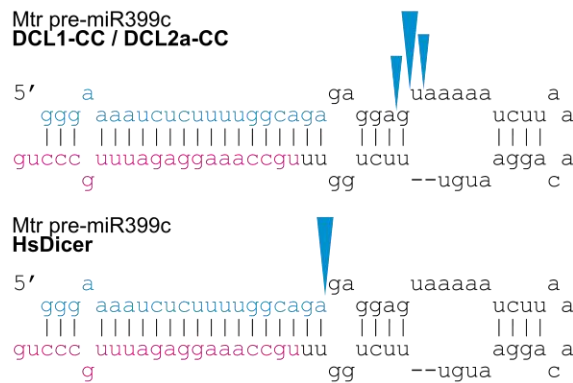
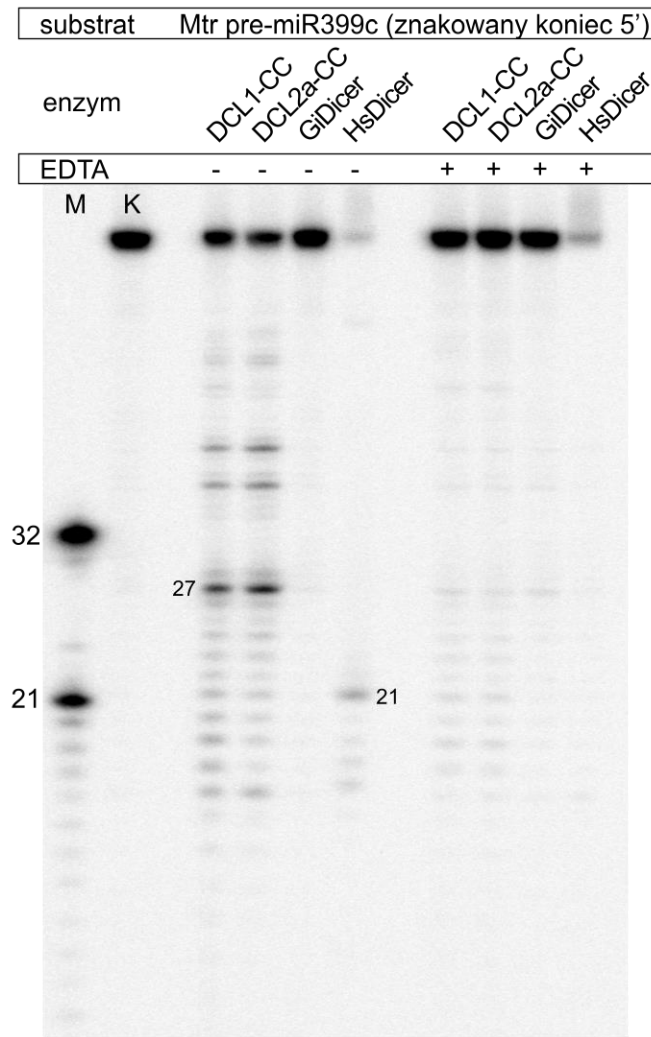
Rysunek 4.12. Wynik analizy produktów reakcji trawienia substratu RNA Hsa pre-miR33a z udziałem trzonów katalitycznych białek MtDCL1 i MtDCL2a (MtDCL1-CC i MtDCL2a-CC) oraz białek GiDicer i HsDicer. Substrat był znakowany radioizotopowo na końcu 5'. Produkty reakcji rozdzielono metodą elektroforezy w denaturującym żelu poliakrylamidowym i wizualizowano autoradiograficznie. W celach kontrolnych równolegle prowadzono reakcje z dodatkiem EDTA oraz reakcję bez dodatku enzymu (K). M – wzorzec długości cząsteczek RNA. Na podstawie wzorca oszacowano długość głównych produktów reakcji. Określone na ich podstawie miejsca cięcia substratu zaznaczono na modelu jego struktury drugorzędowej.



Rysunek 4.13. Wynik analizy produktów reakcji trawienia dupleksów RNA 33a z udziałem trzonów katalitycznych białek MtDCL1 i MtDCL2a (MtDCL1-CC i MtDCL2a-CC) oraz białek GiDicer i HsDicer. Dupлексы utworzono przez hybrydyzację dwóch cząsteczek RNA (*sens* i *anty*), z których tylko jedna była znakowana radioizotopowo na końcu 5'. Produkty reakcji rozdzielono metodą elektroforezy w denaturującym żelu poliakrylamidowym i wizualizowano autoradiograficznie. Jako kontrolę równolegle prowadzono reakcję bez dodatku enzymu (K). M – wzorzec długości cząsteczek RNA. Na podstawie wzorca oszacowano długość głównych produktów reakcji. Określone na ich podstawie miejsca cięcia substratu zaznaczono na modelu jego struktury drugorzędowej.



Rysunek 4.14. Wynik analizy produktów reakcji trawienia substratu RNA Mtr pre-miR399o z udziałem trzonów katalitycznych białek MtDCL1 i MtDCL2a (MtDCL1-CC i MtDCL2a-CC) oraz białek GiDicer i HsDicer. Substrat był znakowany radioizotopowo na końcu 5'. Produkty reakcji rozdzielono metodą elektroforezy w denaturującym żelu poliakrylamidowym i wizualizowano autoradiograficznie. Jako kontrolę równoległe prowadzono reakcje z dodatkiem EDTA oraz reakcję bez dodatku enzymu (K). M – wzorzec długości cząsteczek RNA. Na podstawie wzorca oszacowano długość głównych produktów reakcji. Określone na ich podstawie miejsca cięcia substratu zaznaczono na modelu jego struktury drugorzędowej.



Rysunek 4.15. Wynik analizy produktów reakcji trawienia substratu RNA Mtr pre-miR399c z udziałem trzonów katalitycznych białek MtdCL1 i MtdCL2a (MtdCL1-CC i MtdCL2a-CC) oraz białek GiDicer i HsDicer. Substrat był znakowany radioizotopowo na końcu 5'. Produkty reakcji rozdzielono metodą elektroforezy w denaturującym żelu poliakrylamidowym i wizualizowano autoradiograficznie. Jako kontrolę równoległe prowadzono reakcje z dodatkiem EDTA oraz reakcję bez dodatku enzymu (K). M – wzorzec długości cząsteczek RNA. Na podstawie wzorca oszacowano długość głównych produktów reakcji. Określone na ich podstawie miejsca cięcia substratu zaznaczono na modelu jego struktury drugorzędowej.

Analizy przeprowadzone z wykorzystaniem czterech wyżej opisanych substratów (Hsa miR33a, dupleks RNA 33a, Mtr miR399c i Mtr miR399o) pozwoliły stwierdzić, że trzono katalityczne MtDCL1 i MtDCL2a posiadają aktywność rybonukleazową. Zaobserwowano także, że oba analizowane białka DCL generują pule produktów o identycznej długości, mimo iż pochodzą z enzymów wyspecjalizowanych w produkcji małych regulatorowych RNA różniących się długością.

W przypadku obu białek DCL zaobserwowano nieprecyzyjny wzór cięcia substratów. Głównemu produktowi trawienia towarzyszyły zawsze produkty o 1 nukleotyd dłuższe oraz krótsze. Ponadto główny produkt przeprowadzonych reakcji miał długość od około 23 do 27 nukleotydów, a więc większą od długości typowych produktów MtDCL1 i MtDCL2a obserwowanych *in vivo*. Dla porównania białko HsDicer z reguły generowało w analogicznych reakcjach jeden główny produkt o długości 21 nukleotydów, identycznej jak w warunkach *in vivo*. Natomiast w przypadku białka GiDicer powstawanie produktów o oczekiwanej długości (około 25 nukleotydów) zaobserwowano jedynie w reakcji z w pełni komplementarnym dupleksem (Rys. 4.13).

Długość produktów obserwowanych w reakcjach MtDCL1-CC i MtDCL2a-CC z pre-miR33a i pre-miR399c sugeruje, że trawienie nastąpiło po związaniu się wystającego końca 3' pre-miRNA w obrębie domeny PAZ. Miejsce trawienia zlokalizowane jest wtedy w odległości około 21 nukleotydów od związanego końca prekursora (Rys. 4.12 i 4.15). Natomiast produkty obserwowane w wyniku reakcji z pre-miR399o mogły powstać zarówno w sposób analogiczny, jak i w wyniku związania pre-miRNA w orientacji odwrotnej, prowadzącej do trawienia prekursora w kierunku od pętli do podstawy spinki prekursora (Rys. 4.14). Miejsce pierwszego cięcia zlokalizowane jest wtedy w odległości około 15 nukleotydów od pętli końcowej.

W reakcjach z dupleksem 33a zaobserwowano powstawanie produktów trawienia zachodzącego tylko w jednej orientacji, prowadzącej do wycięcia miRNA. Działo się tak, pomimo że przygotowany dupleks RNA posiada z obu stron końce charakterystyczne dla pre-miRNA, które wiązane są w obrębie domeny PAZ enzymu. Białka HsDicer i GiDicer trawiły dupleks RNA 33a w obu możliwych orientacjach. W efekcie obserwowano zarówno wycinanie dupleksu miRNA/miRNA*, jak i alternatywnego dupleksu pochodzącego z drugiego końca prekursora. W drugim przypadku następowało jednocześnie rozcięcie sekwencji odpowiadającej miR33a (Rys. 4.13).

W przypadku obu białek DCL i wszystkich trzech stosowanych pre-miRNA zaobserwowano, że generowane produkty mają zróżnicowaną długość w zależności od użytego pre-miRNA. W przypadku Mtr pre-miR399c główny produkt reakcji zarówno z MtDCL1-CC jak i MtDCL2a-CC miał długość 27 nukleotydów (Rys. 4.15), w przypadku Hsa pre-miR33a i Mtr pre-miR399o 23-24 nukleotydów (Rys. 4.12 i 4.14). Jednocześnie zaobserwowano, że cięcie w wyniku którego generowany jest główny produkt reakcji następowało zawsze pomiędzy sparowanym i niesparowanym nukleotydem. Natomiast białko HsDicer generowało w analogicznych reakcjach produkty o długości 21 nukleotydów, niezależnie od tego w jakim kontekście strukturalnym następowało cięcie prekursora.

Brak niesparowań w substracie skutkowało wyzwoleniem przez oba białka DCL produktów o długości około 24 nukleotydów, jak wykazały reakcje z dupleksem RNA 33a (Rys. 4.11). Natomiast w reakcjach z odpowiadającym mu pre-miR33a miejsce cięcia było przesunięte i następowało w pozycji pomiędzy sparowanym i niesparowanym nukleotydem, co skutkowało wyzwoleniem głównego produktu o długości 23 nukleotydów (Rys. 4.12).

V. DYSKUSJA

V.1. BIAŁKA DCL KODOWANE PRZEZ GENOM *MEDICAGO TRUNCATULA*

Bobowate stanowią drugą po wiechlinowatych najszerzej uprawianą na świecie rodzinę roślin. Szacuje się, że zaspokajają one około jedną trzecią światowego zapotrzebowania na azot pochodzenia białkowego u ludzi [181]. Unikalną cechą roślin bobowatych stanowi możliwość asymilacji azotu atmosferycznego poprzez symbiozę z bakteriami glebowymi należącymi do kilku rodzajów (między innymi *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*) nazywanych wspólnie ryzobiami. Proces ten zachodzi w wyspecjalizowanych strukturach zwanych brodawkami, które tworzone są przez komórki wewnętrznych warstw kory pierwotnej korzeni. *Medicago truncatula* stanowi modelowy organizm do badań nad całą rodziną bobowatych z racji posiadania stosunkowo małego diploidalnego genomu, samopylności, krótkiego cyklu rozwojowego oraz dobrze opracowanych metod transformacji [182]. Stąd badania nad *Medicago* mają istotne znaczenie zarówno z naukowego jak i gospodarczego punktu widzenia.

Analizy sekwencji genomu *Medicago* w wersji Mt3.5 pozwoliły na identyfikację jedynie trzech genów kodujących białka DCL: *MtDCL1*, *MtDCL2* i *MtDCL3* [137]. Natomiast dwa inne niemal w pełni zsekwencjonowane i dobrze opisane genomy bobowatych: *Glycine max* i *Lotus japonicus*, kodują wszystkie cztery rodzaje białek DCL. U *Glycine max* zidentyfikowano siedem genów *dcl*: *GmDCL1a*, *GmDCL1b*, *GmDCL2a*, *GmDCL2b*, *GmDCL3*, *GmDCL4a* i *GmDCL4b* [135, 136], a u *Lotus japonicus* pięć: *LjDCL1*, *LjDCL2a*, *LjDCL2b*, *LjDCL3* i *LjDCL4* [183].

W ramach niniejszej pracy, bazując na zaktualizowanej sekwencji genomu *Medicago* w wersji Mt4.0, zidentyfikowano gen kodujący białko rodzaju DCL4, oznaczony jako *MtDCL4*. Jak wykazano u *Arabidopsis*, DCL4 pełni istotną rolę w biogenezie ta-siRNA oraz obronie rośliny przed infekcją wirusową. Wiadomo jednak, że w przypadku braku DCL4 inne białka DCL mogą przejmować jego funkcje [76, 77], co mogło częściowo usprawiedliwiać nieobecność DCL4 u *Medicago*. Niemniej zaprezentowane w pracy wyniki dowodzą, że genom *Medicago* koduje wszystkie cztery rodzaje białek DCL, podobnie jak ma to miejsce u pozostałych bobowatych.

Analiza filogenetyczna wykazała, że dwa dalsze zidentyfikowane i nieznanne dotąd geny białek DCL stanowią paralogi *MtDCL2*. Duplikacje genu kodującego DCL2

zostały zaobserwowane u wielu innych okrytonasiennych, w tym z rodziny bobowatych, natomiast *Arabidopsis* posiada tylko jeden gen tego rodzaju. Jak zaznaczono wcześniej uważa się, że duplikacje prowadzące do powstania czterech rodzajów białek DCL miały miejsce w krótkim odstępie czasu, we wczesnym etapie ewolucji roślin [139]. W analizie filogenetycznej roślinne białka DCL z różnych gatunków roślin tworzą wyraźne grupy monofiletyczne obejmujące wyłącznie białka danego rodzaju (DCL1-4). Prawdopodobnie odzwierciedla to zachowaną we wszystkich gatunkach specjalizację poszczególnych białek DCL w generowaniu określonych klas małych regulatorowych RNA. Natomiast w analogicznej analizie obejmującej wyłącznie izoformy DCL2 z trzech roślin bobowatych (*Medicago truncatula*, *Glycine max* i *Lotus japonicus*) oraz *Arabidopsis* białka formują grupy filogenetyczne zgodne z przynależnością gatunkową. Izoforny DCL2 z *Glycine max* i *Lotus japonicus* kodowane są przez geny, które tworzą duplikacje tandemowe na jednym z chromosomów, natomiast każdy z trzech genów *MtDCL2* zlokalizowany jest na innym chromosomie. Szacuje się, że ostatni wspólny przodek *Medicago truncatula* i *Glycine max* istniał około 54 miliony lat temu, a *Medicago truncatula* i *Lotus japonicus* – około 51 milionów lat temu [184]. Sekwencje aminokwasowe MtDCL2a i MtDCL2b wykazują 85% identyczności, podobnie jak dwie izoformy DCL2 u *Populus trichocarpa*. Duplikacja prowadząca do powstania dwóch izoform DCL2 u *Populus trichocarpa* prawdopodobnie miała miejsce około 8-13 milionów lat temu [133]. Razem może to sugerować, że pomimo obecności przynajmniej dwóch genów *DCL2* u wszystkich trzech analizowanych roślin bobowatych, powstały one w wyniku niezależnych duplikacji, już po rozdzieleniu się linii ewolucyjnych tych trzech gatunków.

Interesującą obserwacją była identyfikacja trzeciego genu *MtDCL2*, kodującego skrócone białko DCL. U *Glycine max* poza genami *GmDCL2a* i *GmDCL2b* zidentyfikowano także dwa inne paralogi oznaczone jako *GmDCL2c* i *GmDCL2d*. Zostały one jednak uznane za pseudogeny, z racji obecności wielu przedwczesnych kodonów STOP w obrębie sekwencji kodującej odpowiadające im białka DCL [135]. Natomiast *MtDCL2c* posiada tylko jeden przedwczesny kodon STOP, który skutkuje skróceniem kodowanego białka względem typowej architektury białek DCL2 o fragment końca karboksylowego obejmujący część domeny RNazy IIIb oraz dsRBD. Gen *MtDCL2c* ulega ekspresji we wszystkich badanych organach *Medicago*, a więc nie

można wykluczyć, że jego dojrzałe transkrypty podlegają translacji, co prowadzi do powstawania skróconego wariantu białka DCL2 w roślinie.

Funkcja hipotetycznego białka MtDCL2c pozostaje w sferze domysłów, znany jest jednak przykład funkcjonowania w komórkach skróconego białka Dicer. Jak już wspomniano białko Dicer z *Caenorhabditis elegans* (DCR-1) podlega może trawieniu przez kaspazę CED-3, co prowadzi do utworzenia białka obejmującego fragment RNazy IIIa oraz domenę RNazy IIIb i dsRBD. Jak wykazano utworzony enzym posiada aktywność DNazy i pełni ważną rolę w procesie apoptozy komórek nicienia [164]. Analizy sekwencji wykazały, że hipotetyczne białko MtDCL2c posiada region helikazowy oraz domeny DUF283, PAZ, RNazy IIIa i fragment RNazy IIIb. Tak więc MtDCL2c, podobnie jak skrócone DCR-1, posiada jedną w pełni zachowaną i jedną skróconą domenę RNazy III. Jak wykazano u człowieka brak dsRBD na końcu karboksylowym nie musi zaburzać funkcjonalności białka Dicer. Domena ta warunkuje aktywność rybonukleazową ludzkiego białka Dicer jedynie w sytuacji braku domeny PAZ [144]. Oznacza to, że MtDCL2c może potencjalnie posiadać aktywność nukleazową i pełnić jakąś funkcję biologiczną w komórkach *Medicago*.

Analizy poziomu akumulacji mRNA MtDCL pozwoliły na identyfikację wszystkich badanych transkryptów w różnych częściach młodych oraz dojrzałych roślin *Medicago*. Sugeruje to, że wszystkie sześć genów *MtDCL* ulega konstytutywnej ekspresji w czasie całego cyklu życiowego rośliny. Najwyższy poziom akumulacji odnotowano zawsze dla mRNA MtDCL1. Podobne obserwacje opisano w przypadku analizy akumulacji transkryptów genów *DCL* u *Arabidopsis*. Badania te także dotyczyły wielu różnych części rośliny, w tym liści i łodyg na różnych etapach rozwoju oraz korzeni roślin poddanych działaniu różnych czynników stresowych [185]. Wysoki poziom ekspresji *DCL1* na tle pozostałych genów *DCL* podkreśla ważną rolę białka DCL1 w regulacji ekspresji genów u roślin.

Analizując poziom ekspresji genów *MtDCL* w poszczególnych częściach rośliny zaobserwowano, że jest on istotnie podwyższony w brodawce korzeniowej. Wiele roślinnych miRNA zaangażowanych jest zarówno w proces morfogenezy brodawki jak i symbiozy pomiędzy rośliną i bakterią [186]. Do grupy tej należą między innymi miRNA miR166 i miR169 z *M. truncatula* [187, 188] oraz miR482, miR1512 i miR1515 z *G. max* [189]. Zróznicowany profil akumulacji pomiędzy brodawką, a korzeniem *Medicago* wykazano dla dużej grupy miRNA, w tym ich 22- i 24-nukleotydomowych wariantów, oraz 24-nukleotydomowych małych regulatorowych RNA

obejmujących sekwencje miRNA* [190]. Sugeruje to udział wielu białek DCL w trawieniu prekursorów tych miRNA i może wyjaśniać podwyższony poziom ekspresji wszystkich genów *MtDCL* w brodawce.

Wiadomo także, że wiele roślinnych miRNA, które biorą udział w odpowiedzi na infekcję patogenem, jest jednocześnie zaangażowanych w nawiązywanie relacji symbiotycznej w korzeniach roślin bobowatych [191]. Jak wykazano do grupy tej należą między innymi miRNA związane z nieswoistą odpornością roślin opartą o tzw. molekularne wzorce związane z patogenami (PAMP, ang. *pathogen-associated molecular pattern*), w tym miRNA regulujące roślinne geny odporności (R, ang. *resistance*) [192]. Precyzyjna regulacja roślinnych systemów obronnych wydaje się być zatem kluczowym czynnikiem warunkującym możliwość nawiązania symbiozy pomiędzy rośliną, a bakterią. Znaczenie tego procesu podkreśla także fakt, iż dużą część obecnych w glebie ryzobiów stanowią szczepy pasożytnicze, które mogą infekować roślinę, lecz nie biorą udziału w wiązaniu azotu [193].

Jak wykazano u *Arabidopsis* białka DCL4 i DCL1 biorą udział w obronie rośliny przed infekcją bakteryjną [194], jednocześnie DCL4 i DCL2 pełnią ważną rolę w odpowiedzi antywirusowej organizmu [77]. Zaobserwowana w brodawce istotna indukcja ekspresji wszystkich genów *MtDCL* zaangażowanych w odpowiedź rośliny na infekcję patogenem świadczyć może o ich ważnej roli w procesie nawiązywania relacji symbiotycznej.

U *Glycine max* wykazano, że miR1515 zaangażowany w proces morfogenezy brodawki reguluje gen *GmDCL2b*, kodujący jedną z dwóch funkcjonalnych izoform DCL2 obecnych w roślinie [189]. Specjalizacja funkcjonalna poszczególnych izoform białek DCL danego rodzaju została wykazana u *Oryza sativa*, gdzie białka OsDCL3a i OsDCL3b biorą udział w biogenezie różnych klas 24-nukleotydowych siRNA [195]. Niemal 40-krotny wzrost poziomu ekspresji *MtDCL2b* w brodawce względem korzenia świadczyć może zatem o specjalizacji funkcjonalnej białek DCL2 u *Medicago*, oraz wskazywać na szczególną rolę *MtDCL2b* w funkcjonowaniu brodawki.

W ramach niniejszej pracy zidentyfikowano także alternatywny wariant mRNA *MtDCL1*, oznaczony jako *MtDCL1-AS*. Wariant ten obejmuje niewycięty intron 14 genu *MtDCL1*, obecny jest we wszystkich analizowanych organach rośliny i stanowi istotną część całkowitej frakcji mRNA *MtDCL1*. Jak zaznaczono wcześniej, u *Arabidopsis* analogiczny alternatywny wariant mRNA *AtDCL1* stanowi prekursor *Ath mir838*, a sekwencja tego miRNA zlokalizowana jest w intronie 14 [70, 175].

Nie powiodła się identyfikacja produktów trawienia MtDCL1-AS w obrębie sekwencji intronu 14, ani żadnego miRNA potencjalnie pochodzącego z tego intronu (wśród zdeponowanych w bazie danych miRBase). Nie można jednak wykluczyć, że intron 14 genu *MtDCL1* koduje funkcjonalny miRNA, gdyż zawiera sekwencję tworzącą dużą strukturę typu spinki charakterystyczną dla prekursorów miRNA, a w jej obrębie znajduje się sekwencja homologiczna do Ath miR838. Z drugiej strony możliwość sklonowania pełnej długości MtDCL1-AS wskazuje, że przynajmniej część mRNA obejmujących intron 14 akumuluje w formie nieprzeciętej i koduje skrócone białko MtDCL1. W badaniach nad Ath miR838 zaproponowano, że wiązanie białka DCL1 w obrębie spinki kodującej to miRNA może być przyczyną zaburzeń w składaniu pre-mRNA AtDCL1 [70]. Ponieważ u *Medicago* zaobserwowano podobne zaburzenia składania pre-mRNA, nie zidentyfikowano natomiast produktów trawienia MtDCL1-AS, możliwe jest, że u *Arabidopsis* proces biogenezy Ath miR838 nie wpływa na powstawanie alternatywnego wariantu mRNA AtDCL1. Z drugiej strony, istnieje także możliwość, że białko MtDCL1 wiąże się do struktury spinki obecnej w intronie 14, co powoduje zaburzenie w wycinaniu tego intronu z pre-mRNA, ale do trawienia dochodzi bardzo rzadko lub nie dochodzi wcale. Hipotezę tą wspiera obserwowany bardzo niski poziom Ath miR838 u *Arabidopsis* [70]. Zróżnicowany udział MtDCL1-AS w całkowitej puli mRNA MtDCL1 w różnych częściach rośliny może sugerować, że powstawanie tego alternatywnego transkryptu ma znaczenie w regulacji poziomu pełnej długości białka MtDCL1 w komórce, które jest najważniejszym dla zachowania homeostazy spośród wszystkich DCL. Ponieważ zjawisko zaobserwowano u dwóch odlegle spokrewnionych gatunków roślin, wydaje się, że ma ono charakter konserwatywny i może być powszechne wśród roślin.

MtDCL1-AS koduje białko obejmujące region helikazowy, domenę DUF283 oraz fragment domeny PAZ MtDCL1. Akumulacji tej formy mRNA sprzyja zapewne brak trawienia związanego z obecnością miRNA w obrębie intronu 14. Istnieje więc możliwość, że transkrypty te podlegają translacji umożliwiając syntezę kolejnego skróconego wariantu białka DCL w komórce. Przez brak obu domen RNazy III hipotetyczne białko byłoby pozbawione aktywności rybonukleazowej, co nie wyklucza jednak jego innej roli biologicznej.

W komórkach zwierzęcych funkcjonuje grupa helikaz określanych jako ATPazy aktywowane dupleksami RNA (DRA, ang. *duplex RNA activated ATPase*), w ramach której klasyfikowane są także zwierzęce białka Dicer [196]. Do grupy białek

DRA należą między innymi zidentyfikowane u *Caenorhabditis elegans* helikazy spokrewnione z białkami Dicer (DRH, ang. *Dicer-related helicase*). Wykazano, że białka DRH-1 i DRH-3 w różny sposób wpływają na prawidłowe funkcjonowanie endogennych i egzogennych siRNA w komórkach nicienia [197]. DRH-1 uczestniczy w identyfikacji dsRNA pochodzenia wirusowego [198]. Natomiast DRH-3 wpływa na biogenezę drugorzędowych siRNA, w tym vsiRNA, zależnych od białek RDR [199] oraz wiąże się specyficznym z dupleksami RNA o długości minimum 22 nukleotydów [200]. Ponadto DRH-1 i DRH-3 oddziałują z białkiem DCR-1 nicienia [201]. Stąd jedną z proponowanych ról obu białek DRH jest identyfikacja i wiązanie określonych typów prekursorów siRNA oraz rekrutacja białek uczestniczących w ich trawieniu. Ważna rola helikaz z grupy DRA w mechanizmach obronnych opartych o małe regulatorowe RNA u *Caenorhabditis elegans*, oraz fakt, że podobne mechanizmy są na szeroką skalę wykorzystywane w obronie roślin przed patogenami, pozwalają przypisać duży potencjał funkcjonalny skróconemu białku MtDCL1. Nie jest wykluczone, że analogiczne białka obecne są u innych gatunków roślin, w tym *Arabidopsis thaliana*.

V.2. PORÓWNANIE AKTYWNOŚCI KATALITYCZNEJ BIAŁEK MtDCL1 I MtDCL2A

Badania nad roślinnymi i zwierzęcymi białkami Dicer pozwoliły wskazać wiele czynników determinujących miejsce cięcia w obrębie prekursora, a co za tym idzie ostateczną sekwencję generowanych miRNA i siRNA. Zakłada się, że za pozycjonowanie prekursora w obrębie enzymu odpowiadają dwie domeny zlokalizowane po przeciwnych stronach wydłużonej struktury białka, a samo trawienie następuje w centralnie położonym centrum katalitycznym. Domena PAZ posiada dwie kieszenie, które wiążą końce 5' i 3' nici tworzącej prekursor miRNA [146], natomiast region helikazowy prawdopodobnie oddziałuje z pętlą struktury spinki prekursora [153]. W efekcie miejsce cięcia substratu wyznaczone jest względem tych dwóch charakterystycznych elementów struktury prekursora, innymi słowy determinowane jest przez odległość dzielącą miejsca wiązania prekursora w obrębie domeny PAZ i regionu helikazowego oraz centrum katalityczne enzymu.

W przypadku ludzkiego białka Dicer określono także wpływ innych elementów struktury pre-miRNA na długość generowanych produktów. Obecność asymetrycznych wypętleń w strukturze wycinanego dupletu powoduje, że nici odpowiadające miRNA i miRNA* mogą mieć różną długość. Natomiast symetryczne wypętlania (tworzące wewnętrzne pętle) nie zmieniają istotnie długości generowanych produktów [202]. Ponadto obecność różnych kofaktorów zwierzęcych białek Dicer wpływa na lokalizację miejsca cięcia w obrębie prekursora [203, 204]. Razem ilustruje to złożoność procesu biogenezy miRNA i siRNA z udziałem białek Dicer.

Jednym z celów prowadzonych badań było porównanie aktywności białka DCL biorącego udział w biogenezie miRNA oraz jednego z białek zaangażowanych w produkcję siRNA. W analizach planowano wykorzystać zarówno pełnej długości białka MtDCL1 i MtDCL2a, jak i odpowiadające im fragmenty trzonów katalitycznych, warunkujące podstawową aktywność rybonukleazową obu enzymów. Porównanie aktywności czterech tak zaprojektowanych białek mogłoby dostarczyć wielu informacji na temat czynników warunkujących specjalizację białek DCL w generowaniu różnych klas małych regulatorowych RNA. Niepowodzeniem zakończyły się jednak próby produkcji pełnej długości białek w bakteriach. Prawdopodobną przyczyną mogła być zbyt duża masa cząsteczkowa obu białek (odpowiednio 212 i 156 kDa). Jak wykazano szansa na uzyskanie wydajnej produkcji rekombinowanego białka w prokariotycznym systemie ekspresyjnym znacząco maleje wraz ze wzrostem jego wielkości [205].

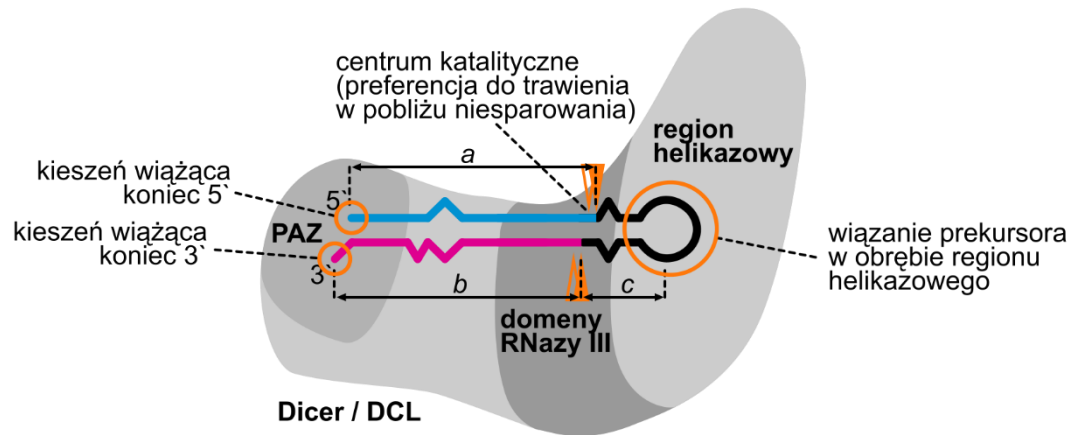
W analizach aktywności trzonów katalitycznych białek MtDCL1 i MtDCL2a zaobserwowano, że trawienie poszczególnych pre-miRNA prowadziło do powstania puli kilku produktów, różniących się długością o 1 nukleotyd. Ponadto w każdym przypadku odnotowano inną długość głównego produktu reakcji: od 23 do 27 nukleotydów. Ponieważ w testowanych trzonach katalitycznych brak było regionu helikazowego, głównym czynnikiem determinującym długość generowanych produktów była prawdopodobnie odległość pomiędzy miejscem wiązania końców prekursora w obrębie domeny PAZ, a centrum katalitycznym enzymu. Dokonane obserwacje wskazywać więc mogą na dużą elastyczność struktury testowanych trzonów katalitycznych enzymów DCL. W przypadku ludzkiego białka Dicer wykazano, że domena DUF283 łączy usytuowane przeciwległe w strukturze białka domeny PAZ oraz helikazy, a jej lokalizacja sugeruje także możliwość dimeryzacji z dsRBD [156]. Wpływa to zapewne na stabilizację struktury trzonu katalitycznego w pełnej długości białku HsDicer. Brak dużego fragmentu łączącego region helikazowy z domeną PAZ, w tym domeny DUF283, tłumaczyć może domniemaną elastyczność struktury testowanych trzonów katalitycznych białek DCL.

W przeprowadzonych analizach trzony katalityczne białek MtDCL1 i MtDCL2a wykazały identyczną aktywność rybonukleazową, mimo że pochodzą z enzymów wyspecjalizowanych w produkcji odpowiednio 21-nukleotydowych miRNA i 22-nukleotydowych siRNA. Obserwację tą także tłumaczyć można dużą elastycznością struktury obu testowanych trzonów katalitycznych. Prawdopodobnie fragment białka łączący region helikazowy z domeną PAZ, poza stabilizacją struktury trzonu katalitycznego, zapewnia takie ułożenie domen PAZ i RNazy III by faworyzowane było powstawanie produktów o określonej długości, charakterystycznej dla danego białka DCL. Te istotne, lecz zapewne niewielkie różnice strukturalne będą możliwe do określenia jedynie na podstawie wysokorozdzielczych struktur krystalicznych poszczególnych białek DCL. Natomiast na podstawie przeprowadzonych analiz wysunąć można ogólny wniosek, że zarówno specjalizacja białek DCL w generowaniu produktów o odmiennej długości, jak i precyzja cięcia, uwarunkowane są czynnikami zlokalizowanymi poza trzonem katalitycznym enzymu.

Jak zaznaczono wcześniej, roślinne białka DCL1 i DCL2 różnią się liczbą zlokalizowanych na końcu karboksylowym domen typu dsRBD: białka z rodzaju DCL1 typowo posiadają dwie takie domeny, a białka z rodzaju DCL2 – jedną. Różnica ta zachowana była także w analizowanych trzonach katalitycznych MtDCL1 oraz

MtDCL2a i, jak się okazało, pozostała bez wpływu na aktywność katalityczną obu białek. Nie zaobserwowano żadnych różnic zarówno w reakcjach z analogami pre-miRNA jak i dupleksem RNA 33a przypominającym prekursor siRNA. Sugeruje to, że domeny te nie mają bezpośredniego wpływu na selektywność białek DCL wobec określonego typu prekursorów. Z badań nad drugą domeną dsRBD z białka AtDCL1 wiadomo, że wykazuje ona podobne powinowactwo do dupleksów RNA posiadających jak i pozbawionych niesparowań i wypętleń [206], oraz że odpowiada ona głównie za oddziaływania z innymi białkami [155]. Nie można więc wykluczyć, że dsRBD wpływają na specjalizację funkcjonalną białek DCL i ich specyficzność wobec określonego typu substratów. Funkcja ta może być bowiem realizowana przez specyficzne oddziaływania pomiędzy dsRBD, a określonymi kofaktorami. Wiadomo, że HYL1 oddziałuje z drugą domeną dsRBD w AtDCL1 [155], oraz że wpływa na aktywność katalityczną tego enzymu *in vivo* [207]. Rola domen dsRBD w specjalizacji funkcjonalnej roślinnych białek DCL została zaproponowana już na wczesnych etapach badań nad tymi enzymami [208].

Co ciekawe, w analizowanym układzie istotnym czynnikiem determinującym miejsce cięcia wydaje się być struktura samego prekursora. Na podstawie wykonanych analiz wnioskować można, że cięcie z udziałem trzonów katalitycznych białek DCL zachodzi preferencyjnie, nie w określonej odległości od końca prekursora, lecz w pozycji pomiędzy sparowanym i niesparowanym nukleotydem. Jednocześnie prowadzone w celach porównawczych reakcje z HsDicer wykazały, że to pełnej długości białko trawi prekursor w sposób odmienny, zwykle generując niemal homogeny produkt o długości około 21 nukleotydów. Tendencję do wyboru miejsca cięcia w pobliżu niesparowania zaobserwowano wcześniej w testach *in vitro* z użyciem kompleksu białek DCL1 (pełnej długości), HYL1 i SE z *Arabidopsis* [163]. Razem sugeruje to, że oba analizowane trzony katalityczne białek DCL, posiadając stosunkowo elastyczną strukturę, mogą generować produkty o pewnym zakresie długości. W takiej sytuacji wybór miejsca cięcia determinuje struktura samego prekursora, w tym lokalizacja obecnych w nim niesparowań. Jednocześnie wskazuje to, że obserwowana preferencja do cięcia substratu w pobliżu niesparowania jest właściwością centrum katalitycznego roślinnych białek DCL i razem z innymi czynnikami wpływa na ostateczną pozycję, w której następuje przecięcie nici RNA prekursora *in vivo* (Rys. 5.1).



Rysunek 5.1. Znane czynniki wpływające na wybór miejsca trawienia prekursora miRNA przez białka Dicer i DCL. Na precyzyjną lokalizację miejsca cięcia wpływa odległość pomiędzy kieszeniami wiążącymi koniec 5' i 3' prekursora a centrum katalitycznym enzymu (odpowiednio *a* i *b*) jak również pomiędzy centrum katalitycznym a miejscem wiązania prekursora przez region helikazowy (*c*). Dodatkowo w przypadku białek DCL wykazano preferencję do wyboru miejsca trawienia w pobliżu niesparowania obecnego w trzonie struktury spinki prekursora.

VI. WNIOSKI

Wyniki prac, które opisane zostały w niniejszej rozprawie podsumować można następującymi wnioskami:

- Genom *Medicago truncatula* obejmuje sześć genów kodujących białka DCL: *MtDCL1*, *MtDCL2a*, *MtDCL2b*, *MtDCL2c*, *MtDCL3*, *MtDCL4*
- Wszystkie sześć genów *MtDCL* ulega konstytutywnej ekspresji w różnych częściach młodych i dorosłych roślin *Medicago*
- Znacząca indukcja ekspresji wszystkich sześciu genów *MtDCL* w brodawce korzeniowej wskazuje na ich ważną rolę w procesie nawiązywania symbiozy z bakteriami glebowymi
- Szczególnie wysoka indukcja ekspresji genu *MtDCL2b* w brodawce w porównaniu z pozostałymi paralogami sugeruje specjalizację funkcjonalną poszczególnych izoform *MtDCL2*
- Gen *MtDCL2c* koduje skrócone białko DCL, potencjalnie mogące jednak posiadać aktywność rybonukleazową
- W komórkach *Medicago* obecny jest alternatywny transkrypt genu *MtDCL1* (*MtDCL1-AS*), który koduje białko pozbawione aktywności rybonukleazowej, potencjalnie mogące jednak posiadać aktywność helikazową
- Zaburzenie składania mRNA kodującego białko DCL1 może być powszechne u roślin i wpływać na regulację poziomu tego białka w komórkach
- Domeny PAZ, RNazy III a i b oraz dsRBD tworzą tzw. trzon katalityczny białek *MtDCL* i zapewniają ich podstawową aktywność rybonukleazową
- Trzony katalityczne białek *MtDCL1* i *MtDCL2a* charakteryzuje duża elastyczność strukturalna, która w pełnej długości białku DCL ograniczona jest prawdopodobnie przez fragment łączący region helikazowy z domeną PAZ, obejmujący domenę DUF283
- Trzony katalityczne *MtDCL1* i *MtDCL2a* wykazują preferencję do trawienia substratu RNA w pozycji pomiędzy sparowanym, a niesparowanym nukleotydem, co w pełnej długości białkach wpływa na ostateczną pozycję, w której następuje przecięcie nici RNA prekursora
- Obecność różnej liczby domen dsRBD na końcu karboksylowym trzonów katalitycznych *MtDCL1* i *MtDCL2a* nie wpływa bezpośrednio na ich aktywność katalityczną *in vitro*, w tym specyficzność wobec określonego typu substratów

VII. STRESZCZENIE / ABSTRACT

Małe regulatorowe RNA są jednym z najważniejszych czynników kształtujących przebieg różnorodnych procesów zachodzących w organizmach eukariotycznych. Cząsteczki te wpływają między innymi na namnażanie i różnicowanie komórek, stabilizację struktury genomu, są także wykorzystywane jako element specyficznej obrony przed działaniem patogenów. Kluczową rolę w procesie biogenezy małych regulatorowych RNA pełnią rybonukleazy Dicer, u roślin znane jako białka typu Dicer (DCL). Białka te odpowiadają za wycięcie z prekursora dojrzałej cząsteczki regulatorowej (miRNA bądź siRNA). Genom *Arabidopsis thaliana* koduje 4 białka typu Dicer (DCL1-4), reprezentujące 4 wyspecjalizowane funkcjonalnie grupy białek DCL występujące u roślin. DCL1 bierze udział głównie w biogenezie miRNA i generuje produkty o długości około 21 nukleotydów, z kolei białka DCL2, DCL3 i DCL4 zaangażowane są przede wszystkim w biogenezę siRNA i generują produkty, z których większość ma długość odpowiednio 22, 24 i 21 nukleotydów.

Celem niniejszej pracy była identyfikacja oraz charakterystyka genów kodujących białka DCL u modelowej rośliny bobowatej *Medicago truncatula* oraz wstępna ocena aktywności wybranych białek kodowanych przez te geny. W ramach przeprowadzonych badań stwierdzono obecność sześciu genów *DCL* w genomie *Medicago*, kodujących wszystkie cztery typy białek DCL (DCL1-4), w tym trzy izoformy białka DCL2. Określono profil ekspresji wszystkich zidentyfikowanych genów DCL w różnych częściach młodych i dorosłych roślin *Medicago*. Ponadto we wszystkich badanych częściach rośliny zaobserwowano akumulację alternatywnej formy mRNA DCL1, podobnej do opisanej wcześniej u *Arabidopsis*. W dalszym etapie badań dokonano porównania aktywności trzonów katalitycznej dwóch wybranych białek DCL z *Medicago* – białka zaangażowanego w produkcję miRNA (DCL1) oraz jednego z białek odpowiedzialnych za produkcję siRNA (DCL2).

In most eukaryotes a group of small RNAs, known as miRNAs and siRNAs, participates in sequence-specific regulation of numerous biological processes, including developmental and stress response, genome stabilization and defense against pathogens. Biogenesis of small regulatory RNAs depends on a group of endonucleolytic enzymes, members of the RNase III family, which cleave longer RNA precursors to generate mature molecules. In animals these enzymes are called Dicer proteins, while in plants they are known as Dicer-like (DCL) proteins. *Arabidopsis thaliana* possesses four DCL proteins (DCL1-4), which represent four distinct DCL groups typically found in plants. DCL1 mainly produces 21 nt long miRNAs, while DCL2, DCL3, and DCL4 specifically produce various classes of siRNAs: 22, 24 and 21 nt long respectively.

The aim of this study was identification and characterization of DCL genes in *Medicago truncatula*, a model legume plant, and evaluation of the catalytic activity of the encoded proteins. As a result six putative DCL genes were identified in the *Medicago truncatula* genome. They encode DCL proteins of all four types (DCL1-4) including three DCL2 isoforms. Expression profiles of all six genes in various plant tissues and organs were determined. In addition, accumulation of an alternative splicing variant of MtDCL1 mRNA, similar to the one identified in *Arabidopsis*, was observed. Activity of the catalytic core fragments of two DCL enzymes, involved either in miRNA (DCL1) or siRNA biogenesis (DCL2), was demonstrated and compared.

VIII. FINANSOWANIE



Program Międzynarodowe Projekty Doktoranckie „*Structural biology of plants and microbes*” realizowany dzięki wsparciu Unii Europejskiej, Europejskiego Funduszu Rozwoju Regionalnego, w ramach Programu Operacyjnego Innowacyjna Gospodarka.



Projekt „*Wsparcie stypendialne dla doktorantów na kierunkach uznanych za strategiczne z punktu widzenia rozwoju Wielkopolski*” realizowany dzięki wsparciu Unii Europejskiej, Europejskiego Funduszu Społecznego, w ramach Programu Operacyjnego Kapitał Ludzki.



Projekt „*Ochrona patentowa peptydu o aktywności RNazy typu Dicer oraz innowacyjnej metody generowania przy pomocy tego peptydu krótkich cząsteczek RNA o właściwościach regulatorów ekspresji genów*” realizowany dzięki wsparciu Unii Europejskiej, Europejskiego Funduszu Rozwoju Regionalnego, w ramach Programu Operacyjnego Innowacyjna Gospodarka.

IX. LITERATURA

1. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR: **RNA maps reveal new RNA classes and a possible function for pervasive transcription.** *Science* 2007, **316**:1484–1488.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al.: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860–921.
3. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal R, Ainscough R, Alexandersson M, An P, Al E: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520–562.
4. Kapranov P, Willingham AT, Gingeras TR: **Genome-wide transcription and the implications for genomic organization.** *Nat Rev Genet* 2007, **8**:413–423.
5. Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR: **Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays.** *Genome Res* 2005, **15**:987–997.
6. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, et al.: **Landscape of transcription in human cells.** *Nature* 2012, **489**:101–8.
7. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR: **Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution.** *Science* 2005, **308**:1149–1154.
8. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, et al.: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799–816.
9. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, et al.: **GENCODE: The reference human genome annotation for The ENCODE Project.** *Genome Res* 2012:1760–1774.

10. Stephen S, Pheasant M, Makunin I V, Mattick JS: **Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock.** *Mol Biol Evol* 2008, **25**:402–408.
11. Mercer TR, Dinger ME, Mattick JS: **Long non-coding RNAs: insights into functions.** *Nat Rev Genet* 2009, **10**:155–9.
12. Taft RJ, Pheasant M, Mattick JS: **The relationship between non-protein-coding DNA and eukaryotic complexity.** *Bioessays* 2007, **29**:288–299.
13. Mercer TR, Mattick JS: **Structure and function of long noncoding RNAs in epigenetic regulation.** *Nat Struct Mol Biol* 2013, **20**:300–7.
14. Mattick JS: **RNA regulation: a new genetics?** *Nat Rev Genet* 2004, **5**:316–323.
15. Mattick JS: **The central role of RNA in human development and cognition.** *FEBS Lett* 2011, **585**:1600–1616.
16. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*.** *Nature* 1998, **391**:806–811.
17. Krol J, Loedige I, Filipowicz W: **The widespread regulation of microRNA biogenesis, function and decay.** *Nat Rev Genet* 2010, **11**:597–610.
18. Furuno M, Pang KC, Ninomiya N, Fukuda S, Frith MC, Bult C, Kai C, Kawai J, Carninci P, Hayashizaki Y, Mattick JS, Suzuki H: **Clusters of internally primed transcripts reveal novel long noncoding RNAs.** *PLoS Genet* 2006, **2**:e37.
19. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammana H, Helt G, Struhl K, Gingeras TR: **Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs.** *Cell* 2004, **116**:499–509.
20. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.** *Nature* 2009, **458**:223–227.
21. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhata R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R: **The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression.** *Genome Res* 2012:1775–1789.

22. Wu J, Okada T, Fukushima T, Tsudzuki T, Sugiura M, Yukawa Y: **A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in Arabidopsis.** *RNA Biol* 2012;302–313.
23. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS: **RNA polymerase V transcription guides ARGONAUTE4 to chromatin.** *Nat Genet* 2009, **41**:630–634.
24. Yin Q-F, Yang L, Zhang Y, Xiang J-F, Wu Y-W, Carmichael GG, Chen L-L: **Long Noncoding RNAs with snoRNA Ends.** *Mol Cell* 2012:219–230.
25. Wang KC, Chang HY: **Molecular Mechanisms of Long Noncoding RNAs.** *Mol Cell* 2011:904–914.
26. Geisler S, Collier J: **RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts.** *Nat Rev Mol Cell Biol* 2013, **14**:699–712.
27. Tay Y, Rinn J, Pandolfi PP: **The multilayered complexity of ceRNA crosstalk and competition.** *Nature* 2014, **505**:344–52.
28. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J: **Target mimicry provides a new mechanism for regulation of microRNA activity.** *Nat Genet* 2007, **39**:1033–1037.
29. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J: **Natural RNA circles function as efficient microRNA sponges.** *Nature* 2013, **495**:384–8.
30. Matera AG, Terns RM, Terns MP: **Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs.** *Nat Rev Mol Cell Biol* 2007, **8**:209–220.
31. Siomi MC, Sato K, Pezic D, Aravin AA: **PIWI-interacting small RNAs: the vanguard of genome defence.** *Nat Rev Mol Cell Biol* 2011, **12**:246–258.
32. Allen E, Xie Z, Gustafson AM, Sung G-H, Spatafora JW, Carrington JC: **Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana.** *Nat Genet* 2004, **36**:1282–1290.
33. Kim VN, Han J, Siomi MC: **Biogenesis of small RNAs in animals.** *Nat Rev Mol Cell Biol* 2009, **10**:126–139.
34. Merchan F, Boualem A, Crespi M, Frugier F: **Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins.** *Genome Biol* 2009, **10**:R136.
35. Lu C, Jeong D-H, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, Liu B, Cao X, Meyers BC, Green PJ: **Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs).** *Proc Natl Acad Sci U S A* 2008, **105**:4951–4956.

36. Zhao X, Zhang H, Li L: **Identification and analysis of the proximal promoters of microRNA genes in Arabidopsis.** *Genomics* 2013, **101**:187–94.
37. Kim YJ, Zheng B, Yu Y, Won SY, Mo B, Chen X: **The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana.** *EMBO J* 2011, **30**:814–822.
38. Kruszka K, Pieczynski M, Windels D, Bielewicz D, Jarmolowski A, Szweykowska-Kulinska Z, Vazquez F: **Role of microRNAs and other sRNAs of plants in their changing environments.** *J Plant Physiol* 2012:1664–1672.
39. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC: **Expression of Arabidopsis MIRNA genes.** *Plant Physiol* 2005, **138**:2145–2154.
40. Szarzynska B, Sobkowiak L, Pant BD, Balazadeh S, Scheible W-R, Mueller-Roeber B, Jarmolowski A, Szweykowska-Kulinska Z: **Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs.** *Nucleic Acids Res* 2009, **37**:3083–3093.
41. Cuperus JT, Fahlgren N, Carrington JC: **Evolution and functional diversification of MIRNA genes.** *Plant Cell* 2011, **23**:431–442.
42. Zong J, Yao X, Yin J, Zhang D, Ma H: **Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups.** *Gene* 2009, **447**:29–39.
43. Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Carrington JC: **Genome-wide profiling and analysis of Arabidopsis siRNAs.** *PLoS Biol* 2007, **5**:e57.
44. Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS: **The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center.** *Cell* 2006, **126**:79–92.
45. Mosher RA, Schwach F, Studholme D, Baulcombe DC: **PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis.** *Proc Natl Acad Sci U S A* 2008, **105**:3145–3150.
46. Haag JR, Pikaard CS: **Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing.** *Nat Rev Mol Cell Biol* 2011, **12**:483–492.
47. Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS: **In Vitro Transcription Activities of Pol IV, Pol V, and RDR2 Reveal Coupling of Pol IV and RDR2 for dsRNA Synthesis in Plant RNA Silencing.** *Mol Cell* 2012, **48**:811–818.
48. Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, Herr AJ, Pikaard CS, Baulcombe DC: **An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis.** *Plant Cell* 2007, **19**:1507–1521.

49. Si-Ammour A, Windels D, Arn-Boulidoires E, Kutter C, Ailhas J, Meins F, Vazquez F: **miR393 and Secondary siRNAs Regulate Expression of the TIR1/AFB2 Auxin Receptor Clade and Auxin-Related Development of Arabidopsis Leaves.** *PLANT Physiol* 2011:683–691.
50. Yoshikawa M: **Biogenesis of trans-acting siRNAs, endogenous secondary siRNAs in plants.** *Genes Genet Syst* 2013:77–84.
51. Axtell MJ, Jan C, Rajagopalan R, Bartel DP: **A two-hit trigger for siRNA biogenesis in plants.** *Cell* 2006, **127**:565–577.
52. Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, Wu S-H: **22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants.** *Proc Natl Acad Sci U S A* 2010, **107**:15269–15274.
53. Yoshikawa M, Iki T, Tsutsui Y, Miyashita K, Poethig RS, Habu Y, Ishikawa M: **3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3.** *Proc Natl Acad Sci U S A* 2013, **110**:4117–22.
54. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MMH, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, et al.: **Empirical analysis of transcriptional activity in the Arabidopsis genome.** *Science (80-)* 2003, **302**:842–6.
55. Zhang X, Xia J, Lii Y, Barrera-Figueroa B, Zhou X, Gao S, Lu L, Niu D, Liang W, Chen Z, Leung C, Wong T, Guo J, Li Y, Liu R, Zhu J-K, Zhang W, Jin H, Zhang H: **Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function.** *Genome Biol* 2012:R20.
56. Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S: **Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis.** *Genes Dev* 2010, **24**:1010–1021.
57. Axtell MJ: **Classification and comparison of small RNAs from plants.** *Annu Rev Plant Biol* 2013, **64**:137–59.
58. Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ: **Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.** *Nature* 2008, **453**:534–538.
59. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H: **Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes.** *Nature* 2008, **453**:539–543.

60. Pumplin N, Voinnet O: **RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence.** *Nat Rev Microbiol* 2013, **11**:745–60.
61. Wilson RC, Doudna J a: **Molecular mechanisms of RNA interference.** *Annu Rev Biophys* 2013, **42**:217–39.
62. Han J, Lee Y, Yeom K-H, Kim Y-K, Jin H, Kim VN: **The Drosha-DGCR8 complex in primary microRNA processing.** *Genes Dev* 2004, **18**:3016–3027.
63. Okamura K: **Diversity of animal small RNA pathways and their biological utility.** *Wiley Interdiscip Rev RNA* 2011, **3**:351–68.
64. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC: **Genetic and functional diversification of small RNA pathways in plants.** *PLoS Biol* 2004, **2**:E104.
65. Ji X: **The mechanism of RNase III action: how dicer dices.** *Curr Top Microbiol Immunol* 2008, **320**:99–116.
66. Dong Z, Han M-H, Fedoroff N: **The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1.** *Proc Natl Acad Sci U S A* 2008, **105**:9970–9975.
67. Liu C, Axtell MJ, Fedoroff N V.: **The Helicase and RNaseIIIa Domains of Arabidopsis Dicer-Like1 Modulate Catalytic Parameters during MicroRNA Biogenesis.** *PLANT Physiol* 2012, **159**:748–758.
68. Fang Y, Spector DL: **Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants.** *Curr Biol* 2007, **17**:818–823.
69. Vazquez F, Blevins T, Ailhas J, Boller T, Meins F: **Evolution of Arabidopsis MIR genes generates novel microRNA classes.** *Nucleic Acids Res* 2008, **36**:6429–6438.
70. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: **A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana.** *Genes Dev* 2006, **20**:3407–3425.
71. Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M: **Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses.** *Genome Res* 2009, **19**:57–69.
72. Voinnet O: **Origin, biogenesis, and activity of plant microRNAs.** *Cell* 2009, **136**:669–687.
73. Fukudome A, Kanaya A, Egami M, Nakazawa Y, Hiraguri A, Moriyama H, Fukuhara T: **Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of Arabidopsis Dicer-like 4.** *RNA* 2011, **17**:750–760.

74. Borsani O, Zhu JJ-K, Verslues PE, Sunkar R: **Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis.** *Cell* 2005, **123**:1279–1291.
75. Ding S-W: **RNA-based antiviral immunity.** *Nat Rev Immunol* 2010, **10**:632–644.
76. Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC, Voinnet O: **Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense.** *Science* 2006, **313**:68–71.
77. Gascioli V, Mallory AC, Bartel DP, Vaucheret HH: **Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs.** *Curr Biol* 2005, **15**:1494–1500.
78. Pontes O, Vitins A, Ream TS, Hong E, Pikaard CS, Costa-Nunes P: **Intersection of small RNA pathways in Arabidopsis thaliana sub-nuclear domains.** *PLoS One* 2013, **8**:e65652.
79. Papp I, Mette MF, Aufsatz W, Daxinger L, Schauer SE, Ray A, van der Winden J, Matzke M, Matzke AJM: **Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors.** *Plant Physiol* 2003, **132**:1382–1390.
80. Yu B, Bi L, Zheng B, Ji L, Chevalier D, Agarwal M, Ramachandran V, Li W, Lagrange T, Walker JC, Chen X: **The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis.** *Proc Natl Acad Sci U S A* 2008, **105**:10073–10078.
81. Dou Y, Zhang S, Ren G, Xie M, Yu B, Zhang C: **Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis.** *Proc Natl Acad Sci U S A* 2012, **109**:12817–12821.
82. Laubinger S, Sachsenberg T, Zeller G, Busch W, Lohmann JU, Ratsch G, Weigel D: **Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana.** *Proc Natl Acad Sci U S A* 2008, **105**:8795–8800.
83. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X: **Methylation as a crucial step in plant microRNA biogenesis.** *Science* 2005, **307**:932–935.
84. Li J, Yang Z, Yu B, Liu J, Chen X: **Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis.** *Curr Biol* 2005, **15**:1501–1507.
85. Ji L, Chen X: **Regulation of small RNA stability: methylation and beyond.** *Cell Res* 2012:624–636.
86. Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, Kiledjian M: **Differential regulation of microRNA stability.** *RNA* 2010, **16**:1032–1039.

87. Kim YK, Yeo J, Ha M, Kim B, Kim VN: **Cell Adhesion-Dependent Control of MicroRNA Decay.** *Mol Cell* 2011, **43**:1005–1014.
88. Tang F, Hajkova P, O'Carroll D, Lee C, Tarakhovsky A, Lao K, Surani MA: **MicroRNAs are tightly associated with RNA-induced gene silencing complexes in vivo.** *Biochem Biophys Res Commun* 2008, **372**:24–29.
89. Winter J, Diederichs S: **Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization.** *RNA Biol* 2011:1149–1157.
90. Vaucheret H, Mallory AC, Bartel DP: **AGO1 Homeostasis Entails Coexpression of MIR168 and AGO1 and Preferential Stabilization of miR168 by AGO1.** *Mol Cell* 2006, **22**:129–136.
91. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS: **Nuclear processing and export of microRNAs in Arabidopsis.** *Proc Natl Acad Sci U S A* 2005, **102**:3691–3696.
92. Tolia NH, Joshua-Tor L: **Slicer and the argonautes.** *Nat Chem Biol* 2007, **3**:36–43.
93. Ma J-B, Ye K, Patel DJ: **Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain.** *Nature* 2004, **429**:318–322.
94. Ma J-B, Yuan Y-R, Meister G, Pei Y, Tuschl T, Patel DJ: **Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein.** *Nature* 2005, **434**:666–670.
95. Song J-J, Smith SK, Hannon GJ, Joshua-Tor L: **Crystal structure of Argonaute and its implications for RISC slicer activity.** *Science* 2004, **305**:1434–1437.
96. Liu J, Carmell MA, Rivas F V, Marsden CG, Thomson JM, Song J-J, Hammond SM, Joshua-Tor L, Hannon GJ: **Argonaute2 is the catalytic engine of mammalian RNAi.** *Science* 2004, **305**:1437–1441.
97. Manavella P, Koenig D, Rubio-Somoza I, Burbano H a, Becker C, Weigel D: **Tissue-specific silencing of Arabidopsis thaliana SUVH8 by miR171a star.** *Plant Physiol* 2012.
98. Eamens AL, Smith NA, Curtin SJ, Wang M-B, Waterhouse PM: **The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes.** *RNA* 2009, **15**:2219–2235.
99. Manavella PA, Hagmann J, Ott F, Laubinger S, Franz M, MacEk B, Weigel D: **Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1.** *Cell* 2012, **151**:859–870.
100. Iki T, Yoshikawa M, Meshi T, Ishikawa M: **Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants.** *EMBO J* 2011:267–278.

101. Carbonell A, Fahlgren N, Garcia-Ruiz H, Gilbert KB, Montgomery TA, Nguyen T, Cuperus JT, Carrington JC: **Functional Analysis of Three Arabidopsis ARGONAUTES Using Slicer-Defective Mutants.** *Plant Cell* 2012;3613–3629.
102. Vaucheret H: **Plant ARGONAUTES.** *Trends Plant Sci* 2008;350–358.
103. Vaucheret H, Vazquez F, Cr  t   P, Bartel DP: **The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development.** *Genes Dev* 2004, **18**:1187–1197.
104. Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC: **The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci.** *Plant Cell* 2010, **22**:321–334.
105. Jeong D-H, Thatcher SR, Brown RSH, Zhai J, Park S, Rymarquis L a, Meyers BC, Green PJ: **Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage.** *Plant Physiol* 2013, **162**:1225–45.
106. Takeda A, Iwasaki S, Watanabe T, Utsumi M, Watanabe Y: **The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins.** *Plant Cell Physiol* 2008, **49**:493–500.
107. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y: **Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide.** *Cell* 2008, **133**:116–127.
108. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangel JL, Carrington JC: **High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes.** *PLoS One* 2007, **2**.
109. Souret FF, Kastenmayer JP, Green PJ: **AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets.** *Mol Cell* 2004, **15**:173–183.
110. Gy I, Gascioli V, Laressergues D, Morel J-B, Gombert J, Proux F, Proux C, Vaucheret H, Mallory AC: **Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors.** *Plant Cell* 2007, **19**:3451–3461.
111. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC: **Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation.** *Cell* 2008, **133**:128–141.
112. Ji L, Liu X, Yan J, Wang W, Yumul RE, Kim YJ, Dinh TT, Liu J, Cui X, Zheng B, Agarwal M, Liu C, Cao X, Tang G, Chen X: **ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis.** *PLoS Genet* 2011, **7**.

113. Maunoury N, Vaucheret H: **AGO1 and AGO2 act redundantly in miR408-mediated Plantacyanin regulation.** *PLoS One* 2011, **6**.
114. Zhu H, Hu F, Wang R, Zhou X, Sze SH, Liou LW, Barefoot A, Dickman M, Zhang X: **Arabidopsis argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development.** *Cell* 2011, **145**:242–256.
115. Iwakawa H, Tomari Y: **Molecular insights into microRNA-mediated translational repression in plants.** *Mol Cell* 2013, **52**:591–601.
116. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O: **Widespread translational inhibition by plant miRNAs and siRNAs.** *Science* 2008, **320**:1185–1190.
117. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D: **Specific effects of microRNAs on the plant transcriptome.** *Dev Cell* 2005, **8**:517–527.
118. Bartel DP: **MicroRNAs: Target Recognition and Regulatory Functions.** *Cell* 2009:215–233.
119. Friedman RC, Farh KK-H, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs.** *Genome Res* 2009, **19**:92–105.
120. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**:495–500.
121. Lai EC, Axtell MJ, Westholm JO: **Vive la différence: biogenesis and evolution of microRNAs in plants and animals.** *Genome Biol* 2011, **12**:221.
122. Brousse C, Liu Q, Beauclair L, Deremetz A, Axtell MJ, Bouché N: **A non-canonical plant microRNA target site.** *Nucleic Acids Res* 2014:1–10.
123. Dalakouras A, Wassenegger M: **Revisiting RNA-directed DNA methylation.** *RNA Biol* 2013, **10**:453–5.
124. Wierzbicki AT: **The role of long non-coding RNA in transcriptional gene silencing.** *Curr Opin Plant Biol* 2012:517–522.
125. Law J a, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE: **Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1.** *Nature* 2013, **498**:385–9.
126. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, Gregory BD, Ecker JR, Tang H, Pikaard CS: **Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome.** *Genes Dev* 2012:1825–1836.

127. Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, Tu A, Jacobsen SE: **DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons.** *Nat Struct Mol Biol* 2012;870–875.
128. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, Patel DJ, Jacobsen SE: **SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation.** *Nature* 2014:1–2.
129. El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, Vega D, Hakimi M-A, Jacobsen SE, Cooke R, Lagrange T: **Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components.** *Genes Dev* 2007, **21**:2539–2544.
130. Gao Z, Liu H-L, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, Miki D, Zhan X, Pontier D, Lagrange T, Jin H, Matzke AJM, Matzke M, Pikaard CS, Zhu J-K: **An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation.** *Nature* 2010, **465**:106–109.
131. Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, Zhou X, Qi Y: **Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes.** *Mol Cell* 2012, **46**:859–870.
132. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW: **Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways.** *Cell* 2004, **117**:69–81.
133. Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM: **The evolution and diversification of Dicers in plants.** *FEBS Lett* 2006, **580**:2442–2450.
134. Qian Y, Cheng Y, Cheng X, Jiang H, Zhu S, Cheng B: **Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize.** *Plant Cell Rep* 2011, **30**:1347–1363.
135. Curtin SJ, Kantar MB, Yoon HW, Whaley AM, Schlueter JA, Stupar RM: **Co-expression of soybean Dicer-like genes in response to stress and development.** *Funct Integr Genomics* 2012.
136. Liu X, Lu T, Dou Y, Yu B, Zhang C: **Identification of RNA silencing components in soybean and sorghum.** *BMC Bioinformatics* 2014, **15**:4.
137. Capitão C, Paiva J a P, Santos DM, Fevereiro P: **In Medicago truncatula, water deficit modulates the transcript accumulation of components of small RNA pathways.** *BMC Plant Biol* 2011, **11**:79.
138. Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KAT, Tang H,

- Rombauts S, Zhao PX, Zhou P, et al.: **The Medicago genome provides insight into the evolution of rhizobial symbioses.** *Nature* 2011.
139. Mukherjee K, Campos H, Kolaczkowski B: **Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants.** *Mol Biol Evol* 2013, **30**:627–41.
140. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna J a: **Structural basis for double-stranded RNA processing by Dicer.** *Science* 2006, **311**:195–8.
141. Lau PW, Guiley KZ, De N, Potter CS, Carragher B, Macrae IJ: **The molecular architecture of human Dicer.** *Nat Struct Mol Biol* 2012, **19**:436–40.
142. Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X: **Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III.** *Cell* 2006, **124**:355–366.
143. Park J-E, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN: **Dicer recognizes the 5' end of RNA for efficient and accurate processing.** *Nature* 2011, **475**:201–205.
144. Ma E, Zhou K, Kidwell MA, Doudna J a.: **Coordinated activities of human dicer domains in regulatory RNA processing.** *J Mol Biol* 2012, **422**:466–476.
145. MacRae IJ, Zhou K, Doudna JA: **Structural determinants of RNA recognition and cleavage by Dicer.** *Nat Struct Mol Biol* 2007, **14**:934–940.
146. Tian Y, Simanshu DK, Ma J, Park J, Heo I, Kim VN, Patel DJ: **A Phosphate-Binding Pocket within the Platform-PAZ-Connector Helix Cassette of Human Dicer.** *Mol Cell* 2014:1–11.
147. Zou J, Chang M, Nie P, Secombes CJ: **Origin and evolution of the RIG-I like RNA helicase gene family.** *BMC Evol Biol* 2009, **9**:85.
148. Taylor DW, Ma E, Shigematsu H, Cianfrocco M a, Noland CL, Nagayama K, Nogales E, Doudna J a, Wang H-W: **Substrate-specific structural rearrangements of human Dicer.** *Nat Struct Mol Biol* 2013, **20**:662–70.
149. Kawamata T, Tomari Y, Tsutsumi A, Izumi N, Seitz H: **Recognition of the pre-miRNA structure by Drosophila Dicer-1.** *Nat Struct Mol Biol* 2011, **18**:1153–1158.
150. Ma E, MacRae IJ, Kirsch JF, Doudna JA: **Autoinhibition of Human Dicer by Its Internal Helicase Domain.** *J Mol Biol* 2008, **380**:237–243.
151. Cenik ES, Fukunaga R, Lu G, Dutcher R, Wang Y, Tanaka Hall TM, Zamore PD: **Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease.** *Mol Cell* 2011, **42**:172–184.

152. Fukunaga R, Colpan C, Han BW, Zamore PD: **Inorganic phosphate blocks binding of pre-miRNA to Dicer-2 via its PAZ domain.** *EMBO J* 2014, **33**:371–84.
153. Gu S, Jin L, Zhang Y, Huang Y, Zhang F, Valdmanis PN, Kay MA: **The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo.** *Cell* 2012, **151**:900–11.
154. Schauer SE, Jacobsen SE, Meinke DW, Ray A: **DICER-LIKE1: blind men and elephants in Arabidopsis development.** *Trends Plant Sci* 2002, **7**:487–491.
155. Liu Q, Yan Q, Liu Y, Hong F, Sun Z, Shi L, Huang Y, Fang Y: **Complementation of HYPONASTIC LEAVES1 by double-strand RNA-binding domains of DICER-LIKE1 in nuclear dicing bodies.** *Plant Physiol* 2013, **163**:108–17.
156. Lau P-W, Potter CS, Carragher B, MacRae IJ: **DOLORS: versatile strategy for internal labeling and domain localization in electron microscopy.** *Structure* 2012, **20**:1995–2002.
157. Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA: **Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction.** *RNA* 2010, **16**:474–481.
158. Nagano H, Fukudome A, Hiraguri A, Moriyama H, Fukuhara T: **Distinct substrate specificities of Arabidopsis DCL3 and DCL4.** *Nucleic Acids Res* 2013, **42**:1845–1856.
159. Mari-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O: **Reconstructing de novo silencing of an active plant retrotransposon.** *Nat Genet* 2013, **45**:1029–39.
160. Mateos JL, Bologna NG, Chorostecki U, Palatnik JF: **Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor.** *Curr Biol* 2010, **20**:49–54.
161. Bologna NG, Mateos JL, Bresso EG, Palatnik JF: **A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159.** *EMBO J* 2009, **28**:3646–56.
162. Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC, Palatnik JF: **Multiple RNA recognition patterns during microRNA biogenesis in plants.** *Genome Res* 2013, **23**:1675–89.
163. Zhu H, Zhou Y, Castillo-González C, Lu A, Ge C, Zhao Y-T, Duan L, Li Z, Axtell MJ, Wang X-J, Zhang X: **Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1.** *Nat Struct Mol Biol* 2013, **20**:1106–15.
164. Nakagawa A, Shi Y, Kage-Nakadai E, Mitani S, Xue D: **Caspase-dependent conversion of Dicer ribonuclease into a death-promoting deoxyribonuclease.** *Science* 2010, **328**:327–334.

165. Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, Gentzbittel L, Childs KL, Yandell M, Gundlach H, Mayer KF, Schwartz DC, Town CD: **An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*.** *BMC Genomics* 2014, **15**:312.
166. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH: **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic Acids Res* 2011, **39**:D225–D229.
167. Skinner MEE, Uzilov AV V, Stein LDD, Mungall CJJ, Holmes IHH: **JBrowse: a next-generation genome browser.** *Genome Res* 2009, **19**:1630–1638.
168. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H: **ExpASY: SIB bioinformatics resource portal.** *Nucleic Acids Res* 2012, **40**.
169. Letunic I, Doerks T, Bork P: **SMART 7: recent updates to the protein domain annotation resource.** *Nucleic Acids Res* 2012, **40**:D302–5.
170. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S: **MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.** *Mol Biol Evol* 2013, **30**:2725–9.
171. Murashige T, Skoog F: **A revised medium for rapid growth and bioassays with tobacco tissue cultures.** *Physiol Plant* 1962, **15**:473–497.
172. Gamborg OL, Miller RA, Ojima K: **Nutrient requirements of suspension cultures of soybean root cells.** *Exp Cell Res* 1968, **50**:151–158.
173. Ehrhardt DW, Atkinson EM, Long SR: **Depolarization of alfalfa root hair membrane potential by *Rhizobium meliloti* Nod factors.** *Science* 1992, **256**:998–1000.
174. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, Vessella RL, Tewari M: **Absolute quantification by droplet digital PCR versus analog real-time PCR.** *Nat Methods* 2013, **10**:1003–5.
175. Xie Z, Kasschau KD, Carrington JC: **Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation.** *Curr Biol* 2003, **13**:784–789.
176. Kozomara A, Griffiths-Jones S: **MiRBase: Annotating high confidence microRNAs using deep sequencing data.** *Nucleic Acids Res* 2014, **42**.
177. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008, **36**.

178. Blaszczyk J, Tropea JE, Bubunenko M, Routzahn KM, Waugh DS, Court DL, Ji X: **Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage.** *Structure* 2001, **9**:1225–1236.
179. Zhang H, Kolb F a, Brondani V, Billy E, Filipowicz W: **Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP.** *EMBO J* 2002, **21**:5875–5885.
180. Saraiya AA, Li W, Wang CC: **A microRNA derived from an apparent canonical biogenesis pathway regulates variant surface protein gene expression in *Giardia lamblia*.** *RNA* 2011:2152–2164.
181. Graham PH, Vance CP: **Legumes: importance and constraints to greater use.** *Plant Physiol* 2003, **131**:872–877.
182. Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward JE, May GD, Gentzbittel L, Ben C, Denny R, Sadowsky MJ, Ronfort J, Bataillon T, Young ND, Tiffin P: **PNAS Plus: Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*.** *Proc Natl Acad Sci* 2011:E864–E870.
183. Bustos-Sanmamed P, Bazin J, Hartmann C, Crespi M, Lelandais-Brière C: **Small RNA pathways and diversity in model legumes: lessons from genomics.** *Front Plant Sci* 2013, **4**:236.
184. Lavin M, Herendeen PS, Wojciechowski MF: **Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary.** *Syst Biol* 2005, **54**:575–594.
185. Liu Q, Feng Y, Zhu Z: **Dicer-like (DCL) proteins in plants.** *Funct Integr Genomics* 2009, **9**:277–86.
186. Bazin J, Bustos-Sanmamed P, Hartmann C, Lelandais-Briere C, Crespi M: **Complexity of miRNA-dependent regulation in root symbiosis.** *Philos Trans R Soc B Biol Sci* 2012:1570–1579.
187. Combier J-P, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernié T, Ott T, Gamas P, Crespi M, Niebel A: **MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*.** *Genes Dev* 2006, **20**:3084–3088.
188. Boualem A, Laporte P, Jovanovic M, Laffont C, Plet J, Combier J-PP, Niebel A, Crespi M, Frugier F: **MicroRNA166 controls root and nodule development in *Medicago truncatula*.** *Plant J* 2008, **54**:876–887.
189. Li H, Deng Y, Wu T, Subramanian S, Yu O: **Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation.** *Plant Physiol* 2010, **153**:1759–1770.

190. Lelandais-Brière C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy J, Crespi M: **Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules.** *Plant Cell* 2009, **21**:2780–2796.
191. Simon S a, Meyers BC, Sherrier DJ: **MicroRNAs in the rhizobia legume symbiosis.** *Plant Physiol* 2009, **151**:1002–8.
192. Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu J-K, Yu O: **Novel and nodulation-regulated microRNAs in soybean roots.** *BMC Genomics* 2008, **9**:160.
193. Denison RF, Kiers ET: **Lifestyle alternatives for rhizobia: Mutualism, parasitism, and forgoing symbiosis.** *FEMS Microbiol Lett* 2004:187–193.
194. Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H: **A novel class of bacteria-induced small RNAs in *Arabidopsis*.** *Genes Dev* 2007, **21**:3123–3134.
195. Song X, Li P, Zhai J, Zhou M, Ma L, Liu B, Jeong D-H, Nakano M, Cao S, Liu C, Chu C, Wang X-J, Green PJ, Meyers BC, Cao X: **Roles of DCL4 and DCL3b in rice phased small RNA biogenesis.** *Plant J* 2012, **69**:462–74.
196. Luo D, Kohlway A, Pyle AM: **Duplex RNA activated ATPases (DRAs): platforms for RNA sensing, signaling and processing.** *RNA Biol* 2013, **10**:111–20.
197. Guo X, Zhang R, Wang J, Ding S-W, Lu R: **Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in *C. elegans* by distinct mechanisms.** *Proc Natl Acad Sci U S A* 2013, **110**:16085–90.
198. Lu R, Yigit E, Li WX, Ding SW: **An RIG-I-like RNA helicase mediates antiviral RNAi downstream of viral siRNA biogenesis in *Caenorhabditis elegans*.** *PLoS Pathog* 2009, **5**.
199. Gu W, Shirayama M, Conte D, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen CCG, Chaves DA, Duan S, Kasschau KD, Fahlgren N, Yates JR, Mitani S, Carrington JC, Mello CC: **Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the *C. elegans* Germline.** *Mol Cell* 2009, **36**:231–244.
200. Fitzgerald ME, Vela A, Pyle AM: **Dicer-related helicase 3 forms an obligate dimer for recognizing 22G-RNA.** *Nucleic Acids Res* 2014:1–12.
201. Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Pang K, Brownell DR, Harding S, Mitani S, Ruvkun G, Yates JR, Mello CC: **Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways.** *Cell* 2006, **124**:343–354.
202. Starega-Roslan J, Krol J, Koscianska E, Kozlowski P, Szlachcic WJ, Sobczak K, Krzyzosiak WJ: **Structural basis of microRNA length variety.** *Nucleic Acids Res* 2011, **39**:257–268.

203. Fukunaga R, Han BW, Hung JH, Xu J, Weng Z, Zamore PD: **Dicer partner proteins tune the length of mature miRNAs in flies and mammals.** *Cell* 2012, **151**:533–546.
204. Lee HY, Doudna JA: **TRBP alters human precursor microRNA processing in vitro.** *RNA* 2012:2012–2019.
205. Gräslund S, Nordlund P, Consortium SG, Weigelt J, Center BSG, Hallberg BM, Consortium CSG, Bray J, Innovation F, Center ISP, Gileadi O, Knapp S, Structural N, Consortium G, Oppermann U, Arrowsmith C, Facility OPP, Hui R, Production PS, Initiative P, Ming J, Dhe-Paganon S, Park H, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim S-H, et al.: **Protein production and purification.** *Nat Methods* 2008, **5**:135–46.
206. Burdisso P, Suarez IP, Bologna NG, Palatnik JF, Bersch B, Rasia RM: **Second double-stranded RNA binding domain of dicer-like ribonuclease 1: structural and biochemical characterization.** *Biochemistry* 2012, **51**:10159–66.
207. Kurihara Y, Takashi Y, Watanabe Y: **The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis.** *RNA* 2006, **12**:206–212.
208. Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, Murai Y, Koiwa H, Seki M, Shinozaki K, Fukuhara T: **Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana.** *Plant Mol Biol* 2005, **57**:173–188.
209. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E: **The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools.** *Nucleic Acids Res* 2012, **40**.
210. Li J, Dai X, Liu T, Zhao PX: **LegumeIP: An integrative database for comparative genomics and transcriptomics of model legumes.** *Nucleic Acids Res* 2012, **40**.
211. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell RC: **The TIGR Rice Genome Annotation Resource: Improvements and new features.** *Nucleic Acids Res* 2007, **35**.

X. ZAŁĄCZNIKI

Załącznik 1. Wykaz sekwencji kodujących białka DCL z *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Lotus japonicus* (Lj) i *Oryza sativa* (Os) wykorzystywanych w pracy. W tabeli wskazano oznaczenia sekwencji aminokwasowych, genów kodujących poszczególne białka DCL i nazwy baz danych, z których pochodzą. Wszystkie sekwencje dostępne są także za pośrednictwem bazy danych NCBI GenBank.

Nazwa białka	Oznaczenie genu	Oznaczenie sekwencji aminokwasowej	Źródło
AtDCL1	At1g01040	At1g01040.2	<i>The Arabidopsis Information Resource</i> (TAIR) arabidopsis.org [209]
AtDCL2	At3g03300	At3g03300.1	
AtDCL3	At3g43920	At3g43920.2	
AtDCL4	At5g20320	At5g20320.1	
GmDCL1a	Glyma03g42290	Glyma03g42290.1	<i>LegumeIP</i> plantgrn.noble.org/LegumeIP [210]
GmDCL1b	Glyma19g45060	Glyma19g45060.1	
GmDCL2a	Glyma09g02930	Glyma09g02930.1	
GmDCL2b	Glyma09g02920	Glyma09g02920.1	
GmDCL3	Glyma04g06060	Glyma04g06060.1	
GmDCL4a	Glyma17g11240	Glyma17g11240.1	
GmDCL4b	Glyma13g22450	Glyma13g22450.1	
LjDCL2a	chr6.CM0437.480.r2.m	chr6.CM0437.480.r2.m	
LjDCL2b	chr6.CM0437.490.r2.m	chr6.CM0437.490.r2.m	
OsDCL1	LOC_Os03g02970	LOC_Os03g02970.1	<i>Rice Genome Annotation Project</i> rice.plantbiology.msu.edu [211]
OsDCL2a	LOC_Os03g38740	LOC_Os03g38740.1	
OsDCL2b	LOC_Os09g14610	LOC_Os09g14610.1	
OsDCL3a	LOC_Os01g68120	LOC_Os01g68120.1	
OsDCL3b	LOC_Os10g34430	LOC_Os10g34430.1	
OsDCL4	LOC_Os04g43050	LOC_Os04g43050.1	

Załącznik 2. Wyniki uzyskane z pomiarów akumulacji mRNA w poszczególnych częściach rośliny metodą ddPCR. Matryce cDNA odpowiadające całkowitej puli mRNA w danej tkance oznaczono następującymi symbolami: RT01 – 3-dniowa siewka (liścień), RT02 – 3-dniowa siewka (korzeń), RT03 – 10-dniowa siewka (liścień), RT04 – 10-dniowa siewka (korzeń), RT05 – stożek wzrostu łodygi, RT06 – łodyga, RT07 – brodawka, RT08 – korzeń, RT09 – nasiono, RT10 – liść.

Badany transkrypt	cDNA	Powtórzenie biologiczne	Stężenie transkryptu (liczba kopii/ μ l)	Maksimum przedziału ufności*	Minimum przedziału ufności*	Liczba analizowanych kropli	Liczba kropli pozytywnych	Liczba kropli negatywnych	Średnia wartość fluorescencji kropli pozytywnych	Średnia wartość fluorescencji kropli negatywnych	Wartość fluorescencji stanowiąca punkt odcięcia kropli negatywnych
DCL1	RT01	Rep1	27,9	31	24,8	13257	311	12946	27856	11208	14883
DCL1	RT01	Rep2	28,8	31,7	26	15979	387	15592	26952	11594	14953
DCL1	RT02	Rep1	87	93	82	14594	1043	13551	28364	11200	15161
DCL1	RT02	Rep2	87	93	82	15532	1112	14420	27494	11542	15106
DCL1	RT03	Rep1	75,1	79,9	70,3	15243	943	14300	28650	11080	14952
DCL1	RT03	Rep2	73,7	78,3	69,1	16252	987	15265	27507	11426	15036
DCL1	RT04	Rep1	146	152	139	15523	1807	13716	28838	11092	15075
DCL1	RT04	Rep2	128	134	121	16264	1671	14593	27600	11318	15193
DCL1	RT05	Rep1	155	162	148	13721	1694	12027	28904	11204	15102
DCL1	RT05	Rep2	123	129	117	16291	1615	14676	27939	11411	15131
DCL1	RT06	Rep1	161	168	154	15970	2043	13927	28769	11030	14812
DCL1	RT06	Rep2	137	144	131	15532	1713	13819	27631	11180	14944
DCL1	RT07	Rep1	188	196	180	13373	1975	11398	29088	11385	15633
DCL1	RT07	Rep2	182	189	174	15936	2283	13653	28767	12676	16504
DCL1	RT08	Rep1	116	122	110	15184	1425	13759	28850	10910	14510
DCL1	RT08	Rep2	127	133	121	16233	1658	14575	29534	12390	15936
DCL1	RT09	Rep1	72,1	77	67,3	14412	857	13555	29110	10725	14382
DCL1	RT09	Rep2	74,7	79,4	70	15525	955	14570	28363	10993	14372
DCL1	RT10	Rep1	297	307	287	15718	3505	12213	29307	10934	14801
DCL1	RT10	Rep2	264	273	255	16343	3282	13061	27408	10807	14277
DCL1-AS	RT01	Rep1	6,8	8,5	5,4	13330	77	13253	27619	7447,2	11527
DCL1-AS	RT01	Rep2	5,4	6,8	4,3	15882	73	15809	24857	7552,4	8256
DCL1-AS	RT02	Rep1	17,5	19,9	15,2	14612	216	14396	28568	7405,8	11417
DCL1-AS	RT02	Rep2	14,9	17	12,8	15488	195	15293	27721	7517,1	10155
DCL1-AS	RT03	Rep1	12,2	14,1	10,3	15407	159	15248	28058	7265,7	10038

DCL1-AS	RT03	Rep2	10,1	11,8	8,4	16218	139	16079	27705	7512,1	10925
DCL1-AS	RT04	Rep1	26,3	29,1	23,5	15429	341	15088	28256	7205,5	10112
DCL1-AS	RT04	Rep2	26,6	29,4	23,8	15878	355	15523	27643	7329,8	10698
DCL1-AS	RT05	Rep1	22,1	24,9	19,4	13631	254	13377	28990	7450,2	11200
DCL1-AS	RT05	Rep2	20,8	23,2	18,4	16260	285	15975	27425	7360,4	9381
DCL1-AS	RT06	Rep1	30,1	33,1	27,2	15865	401	15464	28092	7038,4	11067
DCL1-AS	RT06	Rep2	25,1	27,7	22,5	16756	354	16402	27680	7140,8	10964
DCL1-AS	RT07	Rep1	26,8	29,8	23,8	13823	311	13512	28045	7307,3	11247
DCL1-AS	RT07	Rep2	30,1	33	27,3	16682	422	16260	28304	9558,1	12975
DCL1-AS	RT08	Rep1	17,6	20	15,3	14543	216	14327	28644	7197,9	8429
DCL1-AS	RT08	Rep2	18,8	21,1	16,5	16304	259	16045	26585	8613,6	11130
DCL1-AS	RT09	Rep1	7,1	8,7	5,7	14337	86	14251	29253	7198,9	9829
DCL1-AS	RT09	Rep2	5,8	7,1	4,6	16571	81	16490	27315	6994,4	10676
DCL1-AS	RT10	Rep1	51,6	55,5	47,8	16101	691	15410	29294	7292,7	11510
DCL1-AS	RT10	Rep2	46,3	49,9	42,7	16349	631	15718	26958	6999,1	9375
DCL2	RT01	Rep1	18,5	21	16,1	13690	214	13476	26495	5520,3	8808
DCL2	RT01	Rep2	18,3	20,7	16	14801	229	14572	25692	5330,6	8625
DCL2	RT02	Rep1	43,3	46,9	39,7	15327	554	14773	26174	5449,4	9557
DCL2	RT02	Rep2	42,1	45,8	38,5	14578	513	14065	25886	5352,8	8936
DCL2	RT03	Rep1	21,3	23,8	18,8	15417	277	15140	26641	5450,7	8090
DCL2	RT03	Rep2	19,7	22,1	17,3	15245	253	14992	26128	5335,5	9036
DCL2	RT04	Rep1	39,4	43	35,9	14202	468	13734	26062	5382,3	9301
DCL2	RT04	Rep2	36,2	39,5	32,9	15178	460	14718	25567	5320,6	9334
DCL2	RT05	Rep1	72	76,8	67,1	14358	852	13506	27247	5583,1	9644
DCL2	RT05	Rep2	61,3	65,6	56,9	15037	763	14274	26767	5350,8	8960
DCL2	RT06	Rep1	57,6	61,7	53,5	15705	750	14955	26987	5466,9	6934
DCL2	RT06	Rep2	48,3	52,2	44,5	15352	618	14734	26509	5348,4	9036
DCL2	RT07	Rep1	79	84	74	14208	927	13281	26862	5713,3	10191
DCL2	RT08	Rep1	40,2	44	36,4	12913	434	12479	27175	5594,2	8199
DCL2	RT09	Rep1	47,7	51,9	43,6	13075	520	12555	27294	5377,6	9061
DCL2	RT09	Rep2	43,1	47	39,2	13031	469	12562	25848	5073,5	8972
DCL2	RT10	Rep1	23,2	25,8	20,6	16125	315	15810	26119	5490,2	6766
DCL2	RT10	Rep2	20,4	22,8	18,1	16734	288	16446	25545	5173,8	6878
DCL2A	RT07	Rep2	82	86,9	77,2	16482	1110	15372	25949	6766,7	10935
DCL2A	RT08	Rep2	41,5	44,9	38	16258	563	15695	27166	6857,9	10432
DCL2B	RT01	Rep1	2,7	3,6	1,9	16634	38	16596	14375	11282	13092
DCL2B	RT01	Rep2	1	1,7	0,6	15795	14	15781	16169	11093	13139
DCL2B	RT02	Rep1	2,7	3,6	1,9	16395	37	16358	17547	10983	13146

DCL2B	RT02	Rep2	1,9	2,8	1,3	14681	24	14657	18448	10885	13038
DCL2B	RT03	Rep1	2,6	3,5	1,8	16060	35	16025	14114	10898	12686
DCL2B	RT03	Rep2	1,8	2,6	1,2	16539	25	16514	14474	10462	12111
DCL2B	RT04	Rep1	8,6	10,1	7	16255	118	16137	27532	10125	12874
DCL2B	RT04	Rep2	11,3	13,1	9,6	16897	162	16735	28749	10305	14043
DCL2B	RT05	Rep1	4,6	5,8	3,6	16945	66	16879	28100	10357	13917
DCL2B	RT05	Rep2	4,9	6,2	3,8	16312	68	16244	28963	10248	13992
DCL2B	RT06	Rep1	2,6	3,5	1,9	17528	39	17489	23517	10166	11456
DCL2B	RT06	Rep2	2,2	3,2	1,5	15278	29	15249	28018	10014	12407
DCL2B	RT07	Rep1	27,2	29,9	24,5	16663	381	16282	29472	10123	14120
DCL2B	RT07	Rep2	28,3	31,1	25,5	16473	391	16082	28951	11743	14452
DCL2B	RT08	Rep1	1	1,7	0,6	15974	14	15960	15101	9971,8	11565
DCL2B	RT08	Rep2	1,5	2,2	0,9	16865	21	16844	15437	10945	12744
DCL2B	RT09	Rep1	2,8	3,8	2	15809	38	15771	20673	9584,6	11227
DCL2B	RT09	Rep2	2,4	3,4	1,7	15600	32	15568	19417	9503,3	11528
DCL2B	RT10	Rep1	1,9	2,7	1,2	14516	23	14493	12912	9566,6	11441
DCL2B	RT10	Rep2	1,3	2	0,8	15433	17	15416	13954	9240,8	10625
DCL2C	RT01	Rep1	3,2	4,3	2,4	16011	44	15967	16150	13457	15151
DCL2C	RT01	Rep2	2,7	3,7	1,9	14073	32	14041	18086	13683	15338
DCL2C	RT02	Rep1	4,2	5,3	3,2	16734	59	16675	25765	13088	15277
DCL2C	RT02	Rep2	4,1	5,4	3,1	14608	51	14557	24989	13003	15066
DCL2C	RT03	Rep1	1,9	2,8	1,3	16459	27	16432	18755	12670	14394
DCL2C	RT03	Rep2	2,2	3,1	1,5	14637	27	14610	18982	12819	14637
DCL2C	RT04	Rep1	13,7	15,7	11,7	16328	189	16139	27629	12345	14731
DCL2C	RT04	Rep2	10,7	12,5	8,9	15016	136	14880	28248	12349	15738
DCL2C	RT05	Rep1	3,5	4,6	2,6	15701	47	15654	27021	12045	15313
DCL2C	RT05	Rep2	2,7	3,7	1,9	14983	34	14949	27307	12364	15188
DCL2C	RT06	Rep1	4,7	6	3,7	16431	66	16365	25006	11626	13738
DCL2C	RT06	Rep2	4,9	6,2	3,7	14819	61	14758	25266	11888	13448
DCL2C	RT07	Rep1	35,1	38,3	32	16048	472	15576	27702	11652	14990
DCL2C	RT07	Rep2	29,5	32,4	26,7	16550	410	16140	26941	13971	16555
DCL2C	RT08	Rep1	6,4	7,7	5,2	17586	95	17491	26964	11348	13749
DCL2C	RT08	Rep2	6,7	8,2	5,3	15423	87	15336	26710	14407	16854
DCL2C	RT09	Rep1	5,1	6,5	4	14674	64	14610	27130	11027	14279
DCL2C	RT09	Rep2	7,5	9,2	6,1	14565	93	14472	25815	11042	14145
DCL2C	RT10	Rep1	1,9	2,7	1,2	16507	26	16481	14420	10913	12352
DCL2C	RT10	Rep2	1,7	2,5	1,1	15670	23	15647	15048	10763	12291
DCL3	RT01	Rep1	7,5	9,1	6	14340	91	14249	25935	3692,3	7818

DCL3	RT01	Rep2	9,5	11,2	7,9	15835	128	15707	24517	3777,1	4871
DCL3	RT02	Rep1	11,9	13,7	10	15628	157	15471	25303	3751,1	6498
DCL3	RT02	Rep2	11,8	13,6	9,9	15878	158	15720	26193	3785,3	6968
DCL3	RT03	Rep1	9,2	10,8	7,6	15776	123	15653	25023	3723,5	4861
DCL3	RT03	Rep2	8,9	10,5	7,4	16698	126	16572	26827	3813,8	7162
DCL3	RT04	Rep1	25,4	28,2	22,6	14613	312	14301	26164	3683,5	7884
DCL3	RT04	Rep2	27,1	29,8	24,3	16301	371	15930	26225	3755,8	8011
DCL3	RT05	Rep1	41,4	45,1	37,8	14559	504	14055	26934	3876,7	4776
DCL3	RT05	Rep2	37,2	40,4	34,1	16918	527	16391	26893	3743,6	6670
DCL3	RT06	Rep1	43,9	47,5	40,3	15267	559	14708	27448	3729,1	6744
DCL3	RT06	Rep2	47,9	51,6	44,3	16739	668	16071	26802	3790,8	6801
DCL3	RT07	Rep1	44,2	48,7	39,8	10332	381	9951	25065	3814	8275
DCL3	RT07	Rep2	54,7	58,7	50,7	16006	727	15279	25772	5491,3	9499
DCL3	RT08	Rep1	22,7	25,3	20,1	15111	289	14822	27147	3919,5	4940
DCL3	RT08	Rep2	25,5	28,2	22,8	16231	348	15883	26400	4978,2	8035
DCL3	RT09	Rep1	26,5	29,4	23,5	14120	314	13806	26722	3544,3	6430
DCL3	RT09	Rep2	27,7	30,8	24,7	13523	315	13208	25911	3564,8	7053
DCL3	RT10	Rep1	22,7	25,2	20,1	16025	306	15719	27741	3868,2	6133
DCL3	RT10	Rep2	22,6	25,6	19,7	12013	229	11784	25055	3630,4	6628
DCL4	RT01	Rep1	2,6	3,6	1,8	14412	32	14380	25974	7430,4	11273
DCL4	RT01	Rep2	3,5	4,7	2,5	12634	37	12597	26607	7535,3	11013
DCL4	RT02	Rep1	18,3	20,7	15,9	14650	226	14424	27105	7625,7	11618
DCL4	RT02	Rep2	17,3	19,7	14,9	13890	203	13687	27274	7448,7	11231
DCL4	RT03	Rep1	9	10,7	7,4	15672	120	15552	28015	7722,5	10792
DCL4	RT03	Rep2	9,2	11	7,5	14190	111	14079	27115	7499	11329
DCL4	RT04	Rep1	37	40,5	33,5	13808	427	13381	29003	7760,5	11036
DCL4	RT04	Rep2	39,4	43	35,8	14000	461	13539	27518	7450,2	11427
DCL4	RT05	Rep1	38,9	42,6	35,3	13638	444	13194	29052	7945,9	10408
DCL4	RT05	Rep2	33,8	37,2	30,4	13586	385	13201	27828	7629,5	11195
DCL4	RT06	Rep1	65,4	69,8	61,1	15989	865	15124	29398	7903,6	11786
DCL4	RT06	Rep2	66,2	70,9	61,5	14127	773	13354	27750	7529,9	11313
DCL4	RT07	Rep1	122	129	115	12720	1254	11466	29621	8294,4	12380
DCL4	RT07	Rep2	116	121	110	17195	1610	15585	26453	9072,4	12018
DCL4	RT08	Rep1	21,7	24,3	19,1	14797	270	14527	29617	8091,3	11594
DCL4	RT08	Rep2	22,2	24,7	19,7	16367	306	16061	27217	9095,4	12175
DCL4	RT09	Rep1	26	28,8	23,2	14912	326	14586	29547	7880,5	10294
DCL4	RT09	Rep2	21	23,4	18,6	16378	290	16088	27171	7272,5	8589
DCL4	RT10	Rep1	29,7	32,7	26,7	15254	380	14874	29880	8140,1	10043

DCL4	RT10	Rep2	26,4	29,1	23,7	16592	368	16224	27383	7484,5	9082
REF-ACT	RT01	Rep1	49,6	53,6	45,7	14548	601	13947	24283	3555,6	6925
REF-ACT	RT01	Rep2	34,2	37,4	31	15408	441	14967	29538	3662,9	8285
REF-ACT	RT02	Rep1	111	116	105	16071	1442	14629	25245	3577	7523
REF-ACT	RT02	Rep2	99	104	93	15727	1264	14463	29749	3612,4	8416
REF-ACT	RT03	Rep1	52,9	56,8	49	16336	718	15618	24473	3405,8	7360
REF-ACT	RT03	Rep2	36,2	39,5	32,9	15327	464	14863	30535	3644,2	7717
REF-ACT	RT04	Rep1	68,2	72,6	63,9	16828	948	15880	25338	3442,6	7176
REF-ACT	RT04	Rep2	51,5	55,5	47,5	14938	640	14298	29482	3570,6	7868
REF-ACT	RT05	Rep1	81,8	86,6	77	16592	1114	15478	24962	3341,9	6564
REF-ACT	RT05	Rep2	57	61,1	52,8	15379	727	14652	29897	3548,9	8018
REF-ACT	RT06	Rep1	133	139	127	17064	1823	15241	25118	3246,5	7242
REF-ACT	RT06	Rep2	109	115	103	15876	1405	14471	30340	3568,1	8306
REF-ACT	RT07	Rep1	36,2	39,3	33,1	17118	519	16599	25749	3278,4	6352
REF-ACT	RT07	Rep2	26,8	29,6	24	15577	351	15226	30535	3584,1	6536
REF-ACT	RT08	Rep1	52,3	56,1	48,6	17201	748	16453	24613	3081	6829
REF-ACT	RT08	Rep2	41,1	44,7	37,6	14813	509	14304	29295	3437,3	7937
REF-ACT	RT09	Rep1	80	85	75,1	15251	1003	14248	25945	3163	7111
REF-ACT	RT09	Rep2	60,9	65,1	56,7	15922	803	15119	30353	3542,4	7865
REF-ACT	RT10	Rep1	94	100	89	15557	1199	14358	26054	3136,3	7307
REF-ACT	RT10	Rep2	65,1	69,4	60,9	16774	903	15871	30475	3526	8289

* Przedstawione wartości wyznaczają granice 95% przedziału ufności dla każdego oszacowania stężenia transkryptu opartego o rozkład Poissona.

Załącznik 3. Sklonowane sekwencje kodujące białka MtDCL1 i MtDCL2a.

>MtDCL1-ORF

ATGGAGGATGATGCCGATGCCGCTGCCGCTGCCGCGAATAACAACAATAGGGTTTCTTCGTCAA
 ATCCTTACTGGCTGGATGCCTGTGAAGATATTTCTATATCTTGTGATGATTTTCATTGATTTTGA
 TGTTTCTAATGATTCTGACCAACAACCAACAACAACAACCAAGACTTTTTTGGTGGTATTGAT
 CGAATCTTTGACAGCATCAAAAACGGCGCTGGTCTCCCTGATCATCCTCCTGCTTCTGCTGCTG
 ATGATGGGAATCAACTACCTCCTACTCATCTTCAACTGGAAGAACAACACCAACAGGTAGAAGA
 AGAAAAACAAGAACAAGAAAGGTTCAACAAAAGGGCTCGCCTCACCGAAGGAGAAAGGAGAATC
 AATCATCAACCAACAACAACAACAACAACAACAACAACAACAACAAGGACAGGTATG
 AGAGGTGTTTCAACAATAATAGGAAGAGACCAAGGAATAATAATAATCATAATCATTCTTTAA
 AAGAGATGTCAAGGGTTATTGGGAGAGGGATAAGTCTTCTAATGACTTGGTCTTTCGCTCCGGA
 ACTTGGGAACCGAATCTTCATACTTCCCAACACAAAATGGACATAGACAACCCCATCAAACAGG
 AAAACCCAAGGAGACCGTTTCTGTTTCTGAGGAAAAAGCCAGACAGTATCAGTTGGATGTTCT
 TGAGCAGGCAAGGACCAGAAATACAATAGCTTTTCTAGAACTGGAGCTGGCAAAACCCATCATT
 GCCGTTCTTCTCATCAAGTCTATACATGAGACATTACATCTCCAAAATAAGAAAATGCTTGCTG
 TATTTTTGGTTCCAAAAGTTCCACTTGTACCAGCAAGCTGAGGTAATCCGTGAGCGCACTGG
 ATATCAAGTGGGGCACTATTGTGGAGAAATGGGGCAGGATTTCTGGGATGCTCGAAGGTGGCAG
 CGTGAATTCGACACCAAGCATGTTTTAGTCATGACCGCTCAAATACTTTTGAACATTTTGGAGC
 ATAGCATCATAAAAATGGAAGCAATCAATCTCCTGATTCTTGATGAGTGCCATCATGCTGTGAA
 GAAACACCCATATTCGCTGGTGTGATGTCTGAGTTCTACCATAACAACGCCCAAAGAAAAGAGACCA
 TCTGTATTTGGAATGACTGCTTCTCCTGTAACTTAAAAGGTGTTTCCAGCCAAGTAGACTGTG
 CAATTAATAATTCGTAATCTTGAAAGTAAGCTAGATTCAATAGTTTGTACCATTAAAGATCGTAA
 AGAGCTAGAGAAACACGTGCCTATGCCCTCTGAGATTGTGGTGAAGTATGATAAAGCGGCCAGT
 CTATGTTATTTACATGAGCAAATAAAGCAGATGGAGACTGAAGTTGAGGAAGCTGCAAAATCTA
 GTTCAAGAAGAAGCAAATGGCAGTTTATGGGAGCTAGAGATGCTGGAGCTAAGGAAGAGCTGCG
 CCAAGTATATGGTGTCTTCTGAAAGAACAGAAAGTGATGGAGCTGCAAACCTAATTCAGAAGTTG
 AGAGCTGTTAATTATGCACTTGGTGAATTGGGACAGTGGTGTGCATATAAGGTTGCACAATCTT
 TTCTTGCAGCTCTGCAAAATGATGAAAGGGCAAACCTATCAACTTGATGTCAAGTTTCAGGAATC
 TTATCTGAGTAAAGTTGTGTCTCTTCTGAAATGTCAGCTGTGAGAGGGTGCCGTTTCTGACAAA
 AATGATGGAGTAGATGACTCACAAAATGGTGCAGCTCACAGTGGGTCTGATGAACATGAAGAGA
 TTGAAGAAGGAGAGCTTCCAGACAGTCACGTTGTCTCTGGTGGGAGCATGTTGATGTCATAAT
 AGGTGCTGCTGTGGCTGATGGCAAGGTGACCCCAAGGTGCAGGCTTTAATCAAATACTCCTC
 AAGTATCAGAACACAGACGACTTCCGTGCAATCATCTTTGTTGAACGTGTTGTGTCTGCATTAG
 TTCTACCCAAGGTTTTTGGCGAGCTTCCATCTCTTAGTTTTGTCAAGTGTGCAAGCTTGATTGG
 TCACAACAATAGTCAGGAAATGCGGACCCACCAAATGCATGATACAATTGCCAAGTTCCGTGAT
 GGGCGGGTCACGTTGTTAGTTGCCACTAGTGTGCGGAAGGAGGTCTTGATATCCGGCAGTGTA
 ACGTTGTTATTCGCTTTGATCTTGCGAAGACTGTCTGGCATAACATTCAGTCTAGGGGACGTGC
 TAGAAAGCCTGGGTCTGATTACATCCTAATGGTTGAGAGGGGAAAATTTGTACATGAAGCATTC
 CTAAAGAATGCCAGGAACAGTGAGGAACTTTGCGGAGAGAAGCAATAGAGAGAACAGACCTTA
 GTCATCTGAAAGACAGTTCAAGGTTGATTTCTGTAGATACCCATCCAGAAACAATGTACCAGGT
 GAAATCTACCGGTGCAGTAGTTAGCTTAAATTCGCTGTTAGGTCTTGTCCACTTCTATTGCTCC
 CAGCTACCTAGTGACAGATATTCAATTCTTCGTCCCGAGTTTATTATGGAGAAGCATGAAAACCT
 CCGGAGGTCCCACTGAATATTCTTGCAGGCTTCAACTGCCATGTAATGCACCATTTGAAAATCT
 GGAGGGTCCATCTGCAGTTCAATCCGCTTGGCACAACAGGCTGTTTGTCTAGCTGCTTGCAGA
 AAATTTGCACGAAATGGGAGCATTCCTGACATGCTATTGCCGGATAAAGGAAGTGGGGAAGAAA
 AAGAAAAGGCTGAGCAAAATGATGAAGGGGATGCACCTCCAGGGACTGCTAGACATAGGGAGTT
 CTATCCTGAAGGTGTGGCTGACATACTGAAGGGAGAATGGATCTTATCTGGAAAAGATGCTTGC
 AACAACCTCAAATTTGCTTCATCTTTATATGTATAACCATAAAGTGCGAGAATGTAGGCCATTCAA
 AGGATCCGTTCTTACTCAGGTTTTCAGATTTTGCAGTACTTTTTGGCACTGAACTGGATGCAGA
 GGTCTTATCGATGTGATGGATCTATTTATCGCTCGAACCCTGACTACGAAGACGTCTCTTGTG
 TTTAGAGAGTCAATAGATATCACTGAGAGTCAGTTGACATCCCTGAAAAGCTTTCATGTAAGGT
 TAATGAGCATTGTCTTGGATGTGGATGTTGAACCATCTACCACTCCTTGGGATCCTGCAAAAGC
 ATATCTGTTTGTCCCAATGGTTGGCGATAAGTCTTCAGATCCTATGAACCAAATGACTGGCAT
 CTGGTTGAGACAATAATGGAGCCGATGCATGGAAAAATCCCTCCAGAAGGCTCGACCAGATG

TTTACCTTGGCACTAATGAGAGAACTTTAGGTGGTGACAGAAGGGAATACGGATTTGGGAAATT
 GCGTCATGGCTTGGCTTTTGGGCAAAAATCACATCCAACCTATGGAATCAGAGGAGCTGTGGCC
 CAGTTTGATGTAGTGAAAGCTTCAGGATTGGTTCCCAACAGAGATTCATGGAAACACAAAATC
 ATATCAATATGACTACCAAAGGAAAATTGATGATGGCAGATACATGTACCAGTCCAGATGATCT
 AGTAGGGAGAATTGTAACGCTGCTCATTACAGGGAAGAGGTTTTATGTTGATTCTATACGCTAT
 GAAATGACAGCAGAGAACTCTTTCCCTAGGAAAGAAGGTTATCTTGGTCCTCTGGAATACAGTT
 CATATGCTGATTACTACAAGCAAAAATATGGTGTGATTTGGCTTACAAGCAACAACCTCTTAT
 AAGAGGGCGTGGTGTACCATACTGCAAGAATCTTTTGTCCCCTCGATTTGAACACAGTGAAGGA
 CATGAAGATGAAACTGAAGAGACCCACGACAAACTTACTATGTTTTCTTCCCTCCAGAGCTAT
 GCCTAGTACCCCCACTGCCTGGATCACTCGTTCGTGGTGGTGCAGAGGCTGCCTTCAATAATGAG
 GAGGGTTGAAAGCATGCTGCTTGCAGTTCAGCTTAAGAATATGATAAATTATCCTGTCCAGGCT
 TCAAAGATCTTGAAGCCTTGACCGCTGCCTCTTGCCAGGAGACTTTCTGCTATGAAAGGGCAG
 AGCTTCTAGGAGATGCTTACCTGAAATGGGTTGTTAGTCGTTTTCTTTTCTTAAACATCCACA
 GAAACATGAAGGCCAGCTTACTAGGATGAGACAACAAATGGTCAGTAACATGGTCTTATATCGG
 TACGCATTAAGTAAAGGTCTTCAATCATATATCCTAGCAGATCGTTTTGCCCCATCCAGATGGG
 CAGCTCCTGGGGTGTACCTGTCTTTGATGAGGATACAAAGGATGAGGAGTCATCTTTATTTGA
 CCAGGAGCGTTCATTTTAAAGGCCGAGAGGATGGACAACACTGATGAATTTGAGGATGAAATG
 GAAGATGGTGAGCTTGAGAGTGATTCAAGTCTTACCGAGTTCGTCCAGCAAGACACTTGCAG
 ATGTTGTGGAAGCACTGATTGGGGTATATTATGTTGAAGGTGGTAAGAATGCTGCTAACCACCT
 TATGAAATGGATTGGTATTCACATAGAGATCGATCCTGATGAGATGGAGTGTATAACGAGGCCA
 TCCAATGTTCCAGACAGCATACTTAGGAGTGGTGGATTTTGGATGCTTTAGAAGGTGCTCTAAATA
 TTAAGTTTAAAGACAAGGGGCTGCTGATAGAATCCATAACTCATGCCTCCAGACCATCTTCAGG
 AGTATCCTGTTATCAGCGGTTGGAGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
 CACTTATTTTTCTCGTACACAGATCTGCCACCAGGACGTCTGACTGACCTTCGAGCTGCTGCTG
 TAAACAATGAAAATTTTGCACGTGTTACAGTTAAACACAATCTTCACCTGCATCTGCGGCATGG
 GTCCAGTGCCTTAGAAAAACAGATTAAGATTTTGTGAGGGAAGTCAAGATGAATTGTCAAAA
 CCAGGGTTTAACTCTTTTGGTCTTGGAGACTGCAAAGCTCCCAAGGTCTTGGTGACATCCTTG
 AATCTATGCTGGTGCATTTTTCTTGACAGTGGGAGGAATACTGCTGTTGTTTGAAGGTTTTT
 TCAGCCTCTTTTGCATCCTATGGTTACACCAGAGACTTTGCCAATGCATCCTGTGCGTGAGCTT
 CAAGAGCGATGCCAGCAGCAGGCTGAAGGCTTGGAGTATAGAGCAAGTCGTGCTGGAATTTAG
 CCACTGTTGAAGTATTCATTGATGGGGTACAAGTAGGGGCTGCTCAAATCCACAGAAGAAAAT
 GGCACAGAAATTAGCTGCCCCGAAATGCCCTTGCTGCTTTGAAAGAGAAGGAAGAGAGTAAAATT
 CAGGAGAAAAATGATGAAAAGGAGACAAAGAGTGGGAACCAGACATTTACGAGACAAACATTA
 ATGATATATGTTTGCGAAGAACTGGCCTATGCCATTTCTACCGGTGTGTGAGTGAGGGTGGTCC
 TGCACATGCAAAGAAGTTTACATTTGCAGTGCGAGTGAATACTACTGATAAAGGTTGGACGGAT
 GAATGTGTTGGAGAGCCAATGCCAAGCGTCAAGAAGGCCAAAGACTCAGCTGCTGTTCTTCTCT
 TGGAACTAATAAACAAATTATACTCGTCATAG

>MtDCL2a-ORF

ATGGAATGGATGACTACGATGAGCAGCAGATCAGGCCTAATGCACTTCCATTTGCGAGAAGCT
 ATCAGCTTGAAGCATTGAAAAAGCGATCCGTGAAAATACCATAGTGTACTTGGAGACTGGTTC
 TGGAAAGACTTTGATAGCCATCATGCTTCTTCAAGCTACGCTTATCACCTCCGAAAGCCTTCT
 CTTACATTGCTGTGTTTTTGGTTCCCAAAGTTGTTTTGGTTTTCTCAACAAGCTGAAGCTTTGA
 GAAATCACACTGATTTGAAAATTGGAATGTATTGGGGAGATATGGGTGTTGACTACTGGGATGG
 AGATGTATGGAAGGAGAAAATGGAGAAACATGAGGTTCTTGTGATGACTCCTGCTATACTGCTT
 TCTTGCTTGAGGCACAGCTTTATCAAACGTCAATGATAAAGGTTTTAATTATGGATGAATGCC
 ACCATGCTGCGGGTAGACACCCATATGCTTGTATAATGACTGAGTTTTATCATCATCAATTGAG
 ATCTGGTATCACTGAGCTTCCCTCGGATTTTTGGGATGACTGCATCCCCAATTAAGTCGAAAGCT
 GCGAATTCTGAATCGACCTTGTCCAAATCTATTCGGGAACATAATGACTCTAATGCATTCAAAGG
 TTTATACTGTGTATGTGATGCTGTCTCACAGTTCATACCAACATCTACCCCAAAATTCGG
 GTTTTACAGGGACAATGGGATTTTCATATACGCTATTGCAAGAATTAGCCAAAAGGCTTAACGAG
 TTAACAAACAGCAGCAGGCGTTGCATCACAAGTTCAGATTTAACTAAATCAGCTGTCCAATCTG
 CTTACAAAAAATAGCAAGGATTTTTGATGCCTCAATTTTTTGGCTTAGAAGAGCTTGGTATTTG
 GTTGGCTCTGAAGGCTGCGGAGTCCCTTATCTTCCATTGAAATTGAAACATTTCTATGGGGTATC
 TCTGGGGATCAGATTGTTAAAAAATTTAGTTTCACTACTATGATGACACTGCAAAGTTGCATAC
 CATCTGATCCTCGGTGGACTATTGCTGACAATATGAATTTCCAATTTGGAAATGGGTCTATTGAC

CTCCAAAGTATGCTGTCTTATTGAATGTCTACTTGAGTACAGGGGTTTAACTGAAATGAGATGC
ATAGTTTTTGTGGAAAGGGTCATTGCAGCTATTGTCCTTGAGGTTCTATTGAACGCCTTGCTTC
CAAATACAATAGCTGGAGGGCTAAATTCATTGCAGGAACCGGTAGTAACTGCAAACCAAAC
TAGAAAAATTCAAAATGAGATTGTGGAAGAATTTCTGTAAGGGATTGGTCAACATCATTGTTGCA
ACATCAATTCTTGAAGAGGGTTTAGATGTCCAAAGTTGCAATCTGGTTATTAGATTTGACCCAT
CTCCCACGGTGTGCAGTTTCGTGCAGTCCCAGGGCCGTGCCAGAATGCGAAATTCAGATTACAT
CCTGATGGTAAAAAGTGGCGATGCAGTTACACGTTCTCGATTAGAGAAATACCTTGATGGTGGG
CAAATGATGAGGAAGGAGTCCTTGCGTTATTCTTCCCTTCCATGTGAGAGTCTTGAAAGTGATC
GATTTAACGAGCAAGCTTACCGAGTTGCAAGCACTGAAGCTGTTGTTAATCTTAGTTCCAGCAT
TACTTTGATATACTTATACTGCTCACGTCTCCCATCAGATGGGTACTTTAAACCGACTCCGATG
TGGGACAAACATAAAGGAATATTGTATCTTCCCTAAGAGTTGTCCCATACATGCTATTCATGTAC
AAGGTGACACAAAGTTCTTAAAGAATATTGCATGCCTTGAAGCATGCAAACAACACTACATAAGAT
TGGAGCGTTGACAGATAATCTCGTTCCTTCTATTGTCTGTTGAAGAAGCAGAGGTGGAGGAATTC
GGGAATGAACCATATTCCGAAGAACAACCAAGTTATGTACCATTTGAATTGGTGAATCAGATGT
CGAACAATAGTAGCACAATCTACTATTGTTATTTAATAGAGCTAAAACAAAATTTTAGCTATGA
CATCACTGTGCAAGACATTTTTCTCGCAACTAGAGTCGAGCTTGATCCAGAAATGGATGCATG
CAATTTGACATGTGTTTCGACAGAGGTAGTTTGGCTGTAAACTTGAGATACAGAGGAAACATTA
ATCTCTCACCGGATCAGGTTCTTTTGTGTAAAAGATTCCAAGTTAGTATCCTTGGAATCTTAT
GAATCTAAAATGAATAAAGAAAGTTTTTCTGAAAAATACTCTCTGGAAGATGATCTTGAGATA
GATTATCTTCTGCTGCCATCTATTGCTATAGAGGAAACACCATCTGTTGATTGGCTAACAAATCA
GTTTCGATACATCCATCTAAAATTCAGTGTCTGTTATCATGCTGCCAATATATTGACTGAGAAGGG
TTTCGCTTGTGGTTGCATATTGCGAAATGCTTTGATTTGTACTCCTCACAATGGTCGTACTIONTAC
ATCACCGCTGGTATGATGGAATTGGACGGAAACTCACCTCTGGAAGTAGGGGGATGGTGAAGTTA
CTACGTACAAGAAGTACTTTGGACAGAAGCATGGCATTCAATTGCGTTTTTGAACATCAACGCCCT
ACTTAAGGCAAGACACGTTTTTCCAGTAAAAAATATTGTTCATGGATATAGTCAATCGAAAGAC
AGAGATGTGAGCAAGACCTTTGTTGAATTACCTCCTGAACTCTGCTCTATAATCATGTCACCAG
TACCCTGAGCACACTTTATACATTTTTCAATTTATTCATCGATCATGCATAGACTTGAGTCGTT
GCTTGGTGTCTATAACTTCAAAAAGATGCATTTGGATTATTGCCACAAAACAAGATTCAAACA
TTCAAGGTCTTGGAAGCAATGACCACAAAGTCATGCAAGGAGACCTTTCAATTATGAGTCACTAG
AGACACTTGGAGATTCTTTTTTAAAATATGCTGTTAGCCAACAGCTTTTTAATATGTATCAAAA
TCACCACGAGGGTCTTCTTAGTGTGAAGAGGGAGAAGATCATTGTAATGCTGCGCTCTCTAAG
TTAGGTTGCGATTTTAGACTTCCGGGCTTCATACGGAATGACCCTTTTGATCCAAAGACCTGGA
TCATTCTGGAGCTAAATCAGGAAGTTTTAAATTAGAAGAGACAGTTTTCAATGGGAGAAAAAT
ATATACCCGCGGAGATAGAAAATTGAAACGGAAAGTTGTTGCTGATGTTGTGCGAAGCACTAATT
GGTGCATTCCTTAGCACTGGTGGTGGATGGCTGCGTTGTTGTTTATGGATTGGATTGGGATCA
AAGTCAATTTTAAATATTATACCCTATGAGAGGCAATTCATTGCTTGTCCCACAACTCTGGTTAA
TGTCAGATTTCTAGAATCCCTATTGAAGTATTTCGTTCTGGGATCGTTCTCTATTAGTAGAAGCT
TTGACACATGGCTCTTACATGCTTCCCTGACGTGCCTAGATGCTACCAGCGACTTGAATACTTGG
GAGACTCGGTGCTGGACTATCTCATTACCATGCATTTGTATCGAGTATATCCTGGCATGTCACC
TGGGCAGTTAACAGACATGAGGTCTGCATCGGTGAATAATGACTGTTATGCAATGTCTGCCATC
AAGGTTCAATTGCATAAGCATGTCTTCATGCCTCTCAAGAGCTACATAAACATATTGTAGCTA
CGCTTGATAAATTTGACGATCAGTCTTCATCAAGTTTTGGATGGGAATCCGAGGCTACATTCCC
CAAGGTACTTGGAGACATCATAGAGTCTTAGCTGGTGCTATATTTGTTGATTGAGGATACGAT
AAGGAGGCTGTGTGGAAAAGCATAAGGCCCTTTTTGGAGCCCCCTGTGACACCTGATACATTAA
CGATCCATCCGATTAGAGAATTAATGAACTGTGCCAAAAAATGAATTATACTAAGGAGAAGAC
CTTGTCCAGGAACGACGGCGTTACTAGCTGTAAAATCGAGGTAATTGCGGATGGAGTCATTCAT
CAGTTTCGAGTACAAAGTTCTACCGACAAGAAAATGCCACAAGACTAGCCTGCAAGGGAGTTT
TGCATTCTTGAATGGAAGTAA