UNIWERSYTET MEDYCZNY W ŁODZI

Regulacja ekspresji genu związanego z biosyntezą prekursora tanszinonów u Szałwii czerwonokorzeniowej (*Salvia miltiorrhiza*)

mgr Małgorzata Majewska

Rozprawa doktorska w oparciu o cykl publikacji naukowych

Promotor: prof. UM dr hab. n. farm. Łukasz Kuźma
Promotor pomocniczy: dr n. biol. Piotr Szymczyk

Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź 2023

Składam najserdeczniejsze podziękowania:

prof. UM dr hab. Łukaszowi Kuźmie *oraz* ***dr Piotrowi Szymczykowi***
za nieocenioną pomoc naukową, zaufanie, wyrozumiałość i życzliwość,
których doświadczałam na każdym etapie powstawania tej pracy

Współautorom *publikacji będących podstawą niniejszej rozprawy doktorskiej*
za zaangażowanie i owocną współpracę

Marcinowi
za wsparcie, cenne rady i motywację, dzięki którym ta droga była łatwiejsza

Spis treści

Wykaz skrótów i akronimów ... **5**
1. Wstęp ... **7**
2. Cel pracy ... **14**
3. Materiały i metody ... **15**
3.1. Założenie hodowli *S. miltiorrhiza* ... 15
3.2. Izolacja sekwencji promotora *HMGR4 S. miltiorrhiza* ... 15
3.3. Analizy *in silico* sekwencji promotora *HMGR4 S. miltiorrhiza* ... 15
3.4. Porównanie TFs wykrytych *in silico* z TFs potwierdzonymi doświadczalnie ... 15
3.5. Porównanie sekwencji promotorów *HMGR S. miltiorrhiza* ... 16
3.6. Ocena konserwatywności roślinnych promotorów *HMGR* ... 16
3.7. Oznaczenie specyficznej dla typu organu roślinnego ekspresji genu *HMGR4 S. miltiorrhiza* ... 16
3.8. Wybór regulatorów ekspresji genu *HMGR4 S. miltiorrhiza* i traktowanie materiału roślinnego ... 17
3.9. Przygotowanie konstruktu nadekspresyjnego pRI201-AN-HMGR4, uzyskanie hodowli transformantów *S. miltiorrhiza* i traktowanie materiału roślinnego ... 17
3.10. Analiza statystyczna wyników ... 18
4. Kluczowe wyniki, dyskusja i wnioski ... **19**
4.1. Analizy *in silico* sekwencji promotora *HMGR4 S. miltiorrhiza* ... 19
4.2. Porównanie TFs wykrytych *in silico* z TFs potwierdzonymi doświadczalnie ... 20
4.3. Porównanie sekwencji promotorów *HMGR S. miltiorrhiza* ... 21
4.4. Ocena konserwatywności roślinnych promotorów *HMGR* ... 21
4.5. Oznaczenie specyficznej dla typu organu roślinnego ekspresji genu *HMGR4 S. miltiorrhiza* ... 22
4.6. Wpływ GA_3, IAA, SA na ekspresję genu *HMGR4 S. miltiorrhiza* ... 22
4.7. Wpływ transformacji konstruktem pRI201-AN-HMGR4 na ekspresję genu *HMGR4 S. miltiorrhiza* ... 23
4.8. Wpływ nadekspresji genu *HMGR4* na biosyntezę tanszinonów w gatunku *S. miltiorrhiza* ... 23
4.9. Wpływ GA_3 oraz IAA na biosyntezę tanszinonów w transformantach *S. miltiorrhiza* ... 23
4.10. Zależna od typu organu roślinnego biosynteza tanszinonów w transformantach *S. miltiorrhiza* 24
4.11. Wpływ środowiska wzrostu na biosyntezę tanszinonów w transformantach *S. miltiorrhiza* ... 25
5. Wnioski końcowe oraz koncepcja dalszych badań ... **26**
6. Piśmiennictwo ... **27**
7. Streszczenie ... **36**
8. Summary ... **38**
„Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview" (**Publikacja I**) ... 40

„Isolation and comprehensive *in silico* characterisation of a new *3-hydroxy-3-methylglutaryl-coenzyme A reductase 4* (*HMGR4*) gene promoter from *Salvia miltiorrhiza*: comparative analyses of plant *HMGR* promoters" (**Publikacja II**) **52**

„The expression profiles of the *Salvia miltiorrhiza 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4* gene and its influence on the biosynthesis of tanshinones" (**Publikacja III**) **87**

Dorobek naukowy **114**

Publikacje będące podstawą rozprawy doktorskiej 114

Pozostały dorobek naukowy 114

Oświadczenie autora rozprawy doktorskiej **117**

Oświadczenia współautorów publikacji wchodzących w skład rozprawy doktorskiej **118**

Wykaz skrótów i akronimów

5′UTR (ang. 5′ untranslated region) – region 5′ niepodlegający translacji

BRE (ang. TFIIB recognition element) – element rozpoznający czynnik TFIIB

CT (ang. cryptotanshinone) – kryptotanszinon

DHTI (ang. dihydrotanshinone I) – dihydrotanszinon I

DMAPP (ang. dimethylallyl pyrophosphate) – pirofosforan dimetyloallilu

DPE (ang. downstream promoter element) – dalszy element promotora

DXS (ang. 1-deoxy-D-xylulose-5-phosphate synthase) – syntaza 5-fosforanu 1-deoksy-D-ksylulozy

EEC (ang. early elongation complex) – wczesny kompleks elongacyjny

GA_3 (ang. gibberellic acid) – kwas giberelinowy

GGPP (ang. geranylgeranyl pyrophosphate) – pirofosforan geranylogeranylu

GGPPS (ang. geranylgeranyl pyrophosphate synthase) – syntaza pirofosforanu geranylogeranylu

GTFs (ang. general transcription factors) – ogólne czynniki transkrypcyjne

HMGR (ang. 3-hydroxy-3-methylglutaryl-coenzyme A reductase) – reduktaza 3-hydroksy-3-metyloglutarylokoenzymu A

HPLC (ang. high performance liquid chromatography) – wysokosprawna chromatografia cieczowa

IAA (ang. indole-3-acetic acid) – kwas indolilo-3-octowy

Inr (ang. initiator element) – element inicjujący

IPP (ang. isopentenyl pyrophosphate) – pirofosforan izopentenylu

IUPAC (ang. International Union of Pure and Applied Chemistry) – Międzynarodowa Unia Chemii Czystej i Stosowanej

MEP (ang. methylerythritol phosphate) – fosforan metyloerytrytolu

MS (ang. Murashige and Skoog)

MVA (ang. mevalonic acid) – kwas mewalonowy

NLS (ang. nuclear localisation signal) – sygnał lokalizacji jądrowej

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

PFM (ang. position frequency matrix) – macierz częstości pozycji

PIC (ang. preinitiation complex) – kompleks reinicjacyjny

PRS (ang. pyrimidine-rich sequence) – sekwencja bogata w pirymidyny

PWM (ang. position weight matrix) – macierz wag pozycji

RT-qPCR (ang. real-time quantitative polymerase chain reaction) – ilościowa reakcja łańcuchowa polimerazy w czasie rzeczywistym

SA (ang. salicylic acid) – kwas salicylowy

sRNA (ang. small RNA) – małe RNA

TI (ang. tanshinone I) – tanszinon I

TIIA (ang. tanshinone IIA) – tanszinon IIA

TAFs (ang. TBP-associated factors) – białka związane z TBP

TF(s) (ang. transcription factor(s)) – czynnik(i) transkrypcyjny(e)

TFBS(s) (ang. transcription factor binding site(s)) – miejsce(a) wiązania czynnika(ów) transkrypcyjnego(ych)

TSS (ang. transcription start site) – miejsce startu transkrypcji

UHPLC (ang. ultra high performance liquid chromatography) – ultrawysokosprawna chromatografia cieczowa

1. Wstęp

Salvia miltiorrhiza Bunge, zwana również Szałwią chińską lub Szałwią czerwoną, to roślina wieloletnia należąca do rodzaju *Salvia* (szałwia) i rodziny *Lamiaceae* (jasnotowate) [1]. Rośnie ona w co najmniej dziewiętnastu prowincjach Chin, a także w Mongolii, Korei, Wietnamie, Japonii i Australii. Preferuje stanowiska słoneczne i wilgotne położone na wysokości od 100 do 1300 m n.p.m. z glebą o pH obojętnym, lekko kwaśnym lub lekko zasadowym. *S. miltiorrhiza* używana jest jako modelowy przedstawiciel roślin leczniczych. Zawdzięcza to ona krótkiemu cyklowi życia, łatwości rozmnażania, niewymagającej uprawie, możliwości hodowli w warunkach *in vitro*, a także stosunkowo małemu rozmiarowi genomu (około 600 Mb). Pierwsze wzmianki na temat tego gatunku pochodzą z najstarszej chińskiej monografii medycznej pt. „Shen Nong's Classic of the Materia Medica" napisanej około II w. p.n.e., a następne z kolejnych tego typu starożytnych dzieł. Obecnie jest jedną z najpopularniejszych roślin tradycyjnej medycyny chińskiej. Jej wysuszony korzeń, zwany Danshen lub Tanshen, stosowany jest samodzielnie lub w połączeniu z innymi ziołami w leczeniu m.in.: chorób układu krążenia, zaburzeń miesiączkowania, bezsenności [2,3]. Właściwości lecznicze tej rośliny wynikają ze zdolności do biosyntezy ponad dwustu substancji bioaktywnych, głównie typu diterpenoidów (w tym tanszinonów) i kwasów fenolowych, a także triterpenoidów, flawonoidów, olejków eterycznych i innych [1,4].

Tanszinony to grupa związków obejmująca abietanowe nor-diterpenoidowe chinony [5]. Dotychczas wyizolowano z *S. miltiorrhiza* kilkadziesiąt tanszinonów [4], spośród których najlepiej przebadano dihydrotanszinon I (ang. dihydrotanshinone I, DHTI), kryptotanszinon (ang. cryptotanshinone, CT), tanszinon I (ang. tanshinone I, TI) i tanszinon IIA (ang. tanshinone IIA, TIIA). Według przeprowadzonych badań tanszinony zapewniają ochronę dla układu sercowo-naczyniowego [6,7], regulują metabolizm [8], zapobiegają chorobom układu nerwowego [9], osteoporozie [10] i nowotworom [11], a także wykazują właściwości przeciwzapalne [12], antyoksydacyjne [13], fitoestrogenne [14] i przeciwbakteryjne [15]. Tanszinony gromadzą się głównie w perydermie korzenia, jednakże nawet tam ich zawartość jest bardzo niska i nie zaspokaja ciągle rosnącego na nie zapotrzebowania [1,16,17]. Stąd konieczne wydaje się opracowanie metod poprawiających wydajność ich biosyntezy.

Analiza porównawcza wyników wysokoprzepustowych badań omicznych (genom, transkryptom, proteom, metabolom) oraz funkcjonalna charakterystyka *in vitro* i *in vivo* uzyskanych kandydatów genów umożliwiły poznanie szlaku biosyntezy tanszinonów w gatunku *S. miltiorrhiza* [5,18]. Szlak ten można podzielić na trzy główne etapy. Punktem wyjścia jest tworzenie prekursorów terpenoidów, czyli pirofosforanu izopentenylu (ang. isopentenyl

pyrophosphate, IPP) i pirofosforanu dimetyloallilu (ang. dimethylallyl pyrophosphate, DMAPP) w szlaku kwasu mewalonowego (ang. mevalonic acid, MVA) w cytoplazmie i w szlaku fosforanu metyloerytrytolu (ang. methylerythritol phosphate, MEP) w plastydach [5,19]. Następnie w wyniku kondensacji DMAPP z IPP powstaje pirofosforan geranylogeranylu (ang. geranylgeranyl pyrophosphate, GGPP), kluczowy prekursor szkieletu węglowego diterpenoidów [5,20]. Ostatni etap to modyfikacje strukturalne. Cyklizacja i defosforylacja GGPP prowadzą do powstania miltiradienu, z którego w wyniku reakcji takich jak utlenianie, hydroksylacja, dekarboksylacja syntetyzowane są biologicznie aktywne tanszinony [5,19]. Reakcje z opisanego powyżej szlaku są katalizowane przez specyficzne enzymy wykazujące w większości niską homologię z enzymami innych gatunków roślin [5]. Wyniki przeprowadzonych badań wskazują, że reduktaza 3-hydroksy-3-metyloglutarylokoenzymu A (ang. 3-hydroxy-3-methylglutaryl-coenzyme A reductase, HMGR) ze szlaku MVA, syntaza 5-fosforanu 1-deoksy-D-ksylulozy (ang. 1-deoxy-D-xylulose-5-phosphate synthase, DXS) ze szlaku MEP oraz syntaza pirofosforanu geranylogeranylu (ang. geranylgeranyl pyrophosphate synthase, GGPPS) to kluczowe enzymy ograniczające tempo biosyntezy tanszinonów [5,21].

HMGR katalizuje konwersję 3-hydroksy-3-metyloglutarylokoenzymu A do MVA, prekursora wielu istotnych związków izoprenoidowych działających w komórkach roślinnych [22,23]. Fundamentalną rolę jaką HMGR pełni w metabolizmie roślin podkreśla precyzyjna regulacja jego funkcji na etapie transkrypcji, translacji, obróbki potranskrypcyjnej i potranslacyjnej [24,25]. Do chwili obecnej zidentyfikowano i zdeponowano w bazie danych GenBank pięć sekwencji genów *HMGR S. miltiorrhiza* (*HMGR–HMGR4*) [26-28]. *HMGR* (EU680958.1) i *HMGR2* (FJ747636.1) przebadano dotąd najszerzej i to zarówno pod kątem strukturalnym, jak i funkcjonalnym [21,26,28-30]. Pozostałe geny *HMGR* nie zostały jak dotąd wystarczająco zbadane, w szczególności dotyczy to *HMGR4* (JN831103.1).

W celu zapewnienia prawidłowego funkcjonowania poszczególnych typów komórek, zgodnie z ich wymaganiami fizjologicznymi czy warunkami środowiska zewnętrznego, konieczna jest regulacja aktywności genów. Regulacja ta odbywa się na poziomie epigenetyki, transkrypcji, obróbki RNA, translacji oraz potranslacyjnej modyfikacji białek. Zrozumienie tych skomplikowanych procesów stanowi obecnie jedno z palących zagadnień biologii molekularnej. U roślin kluczowym etapem tej regulacji jest transkrypcja, czyli przepisywanie informacji zawartej w DNA na RNA [31]. Wskazuje na to chociażby wysoki odsetek genów zaangażowanych w ten proces (15% dla chromosomu czwartego *Arabidopis thaliana*) [32].

Aby transkrypcja mogła zostać zapoczątkowana konieczne są takie zmiany w strukturze chromatyny, które sprawią że promotor i sekwencja kodująca będą bardziej dostępne dla czynników transkrypcyjnych (ang. transcription factors, TFs) i polimerazy RNA. Zalicza się do

nich acetylację histonów, usuwanie grup metylowych, aktywność białek przebudowujących chromatynę czy zaangażowanych w demontaż i ponowny montaż nukleosomów [33]. Podczas inicjacji transkrypcji genów kodujących białko ma miejsce formowanie kompleksu preinicjacyjnego (ang. preinitiation complex, PIC) złożonego z ogólnych czynników transkrypcyjnych (ang. general transcription factors, GTFs) i polimerazy RNA II. Czynnik TFIID wiąże się poprzez podjednostkę białka TBP (ang. TATA box-binding protein) do sekwencji TATA box promotora genu, a pozostałe białka związane z TBP (ang. TBP-associated factors, TAFs) zaczynają oddziaływać z sekwencją DNA położoną poniżej miejsca startu transkrypcji (ang. transcription start site, TSS). Następnie przyłączone zostają czynniki TFIIA i TFIIB, co umożliwia rekrutację kompleksu nieufosforylowanej polimerazy RNA II z TFIIF oraz dołączenie TFIIE i TFIIH. Czynnik TFIIH fosforyluje C-końcową domenę CTD polimerazy RNA II oraz wykazuje aktywność helikazy, prowadząc do powstania aktywnego kompleksu transkrypcyjnego [34]. Opisany powyżej sposób montowania PIC nosi nazwę sekwencyjnego. Po zsyntetyzowaniu RNA o długości około 30 nt następuje pierwsze zatrzymanie polimerazy RNA II oraz odłączenie enzymu od promotora i części elementów maszynerii transkrypcyjnej. Prowadzi to ostatecznie do stabilnego związania się polimerazy RNA II z powstającym pre-mRNA i uformowania wczesnego kompleksu elongacyjnego (ang. early elongation complex, EEC) [33,35]. EEC podlega ciągłym kontrolom zanim stanie się w pełni funkcjonalny. Ma to miejsce szczególnie przed głównym etapem elongacji, kiedy następuje ponowne zatrzymanie polimerazy RNA II i sprawdzenie czy jest ona przygotowana do nadchodzącego etapu [33]. Po odłączeniu się negatywnych czynników elongacji i przyłączeniu pozytywnych enzym odzyskuje aktywność, a łańcuch pre-mRNA ulega dalszemu wydłużaniu [33]. Na tym etapie ma miejsce dodanie czapeczki guanylowej (7-metyloguanozyny) do końca 5′ transkryptu, co zabezpiecza mRNA przed działaniem egzonukleaz i promuje translację [36]. Za sprawą spliceosomu dochodzi do usuwania intronów z pierwotnego transkryptu i łączenia eksonów w procesie splicingu [37]. Stabilność mRNA i wydajność translacji zwiększa dodanie do końca 3′ transkryptu ogonka poli(A) składającego się z nukleotydów adeninowych [38]. Tak utworzony mRNA łączy się z białkami i transportowany jest do cytoplazmy w celu translacji.

Jak wynika z przebiegu transkrypcji, w inicjacji i regulacji tego procesu niezwykle istotną rolę pełni sekwencja DNA zlokalizowana powyżej części kodującej genu zwana promotorem. Budowa promotora dostarcza wielu cennych informacji na temat regulacji aktywności kontrolowanego genu. W strukturze promotorów roślinnych wyróżnia się część rdzeniową (tzw. minimalny promotor), bliższą i dalszą [39]. Część rdzeniowa może zawierać sekwencję TATA box, element inicjujący (ang. initiator element, Inr) czy element rozpoznający czynnik TFIIB (ang. TFIIB recognition element, BRE). Każdy z tych motywów pełni określone funkcje związane

z procesem inicjacji transkrypcji. Kaseta TATA była pierwszym rdzeniowym elementem promotora zidentyfikowanym w eukariotycznych genach kodujących białka [40]. Jest ona zlokalizowana w obszarze od –25 do –39 nt od TSS i wiąże białko TBP [41]. Początkowo sądzono, że sekwencja TATA jest niezbędna do inicjacji transkrypcji i co za tym idzie występuje w promotorach wszystkich organizmów eukariotycznych. Jednakże wraz z rozwojem genomiki wykazano, że obecna jest ona tylko u części promotorów. U roślin potwierdzono jej występowanie u 29% promotorów *A. thaliana* [42] i u 19% promotorów ryżu [43]. Inr otacza TSS lub znajduje się w jego pobliżu. Jest on pod względem funkcjonalnym podobny do TATA box i może z nim współdziałać podczas inicjacji transkrypcji. W przypadku promotorów nieposiadających kasety TATA to Inr kontroluje inicjację transkrypcji, a w niektórych przypadkach funkcje pomocnicze pełni również dalszy element promotora (ang. downstream promoter element, DPE) [44]. DPE zlokalizowany jest w obszarze od +28 do +32 nt od TSS i odpowiada za wiązanie czynnika TFIID do promotora [40]. Z kolei BRE położony jest tuż przed TATA box i wiąże czynnik TFIIB [40]. Co ciekawe, promotory pozbawione wszystkich opisanych powyżej motywów również okazują się być funkcjonalne [35]. Na odcinku bliższym promotora, obejmującym kilkaset pz w górę od rdzeniowej części, rozlokowane są kilkunukleotydowe specyficzne motywy niekodującego DNA (ang. transcription factor binding sites, TFBSs) rozpoznawane przez TFs funkcjonujące jako aktywatory lub represory transkrypcji [45]. Z kolei w części dalszej promotora, znajdującej się w odległości setek, tysięcy a nawet milionów pz od części bliższej, umiejscowione są elementy regulatorowe nazywane wzmacniaczami i wyciszaczami. Regulatory te przy udziale wiążących się do nich TFs poprawiają lub hamują tempo transkrypcji poprzez bezpośrednie lub pośrednie (z udziałem kompleksu mediatora) oddziaływanie z aparatem transkrypcyjnym [46,47]. Takie przekazywanie sygnałów między wzmacniaczem lub wyciszaczem a promotorem rdzeniowym jest możliwe dzięki fizycznemu zbliżeniu tych regionów poprzez tworzenie pętli DNA [48].

Do tej pory zidentyfikowano u roślin kilka tysięcy TFs (ponad 2300 u *Arabidopsis* i ponad 5600 u *Glycine max*) [49]. Część TFs jest konstytutywna, jak GTFs oddziałujące z promotorem rdzeniowym, podczas gdy inne są specyficzne dla typu komórki, stadium rozwoju lub ulegają indukcji przez sygnały zewnętrzne i wewnętrzne, np. stres oksydacyjny, hormony, światło, zimno, upał, suszę, warunki beztlenowe, zranienie, patogeny [50]. Aktywowane TFs wiążą się następnie do sekwencji określonych promotorów, regulując w ten sposób ekspresję szeregu genów. Większość TFs jest wyposażona w sygnał lokalizacji jądrowej (ang. nuclear localisation signal, NLS), domenę wiążącą DNA, domenę oligomeryzacji i domenę efektorową. NLS to sekwencja aminokwasowa bogata w dodatnio naładowane lizyny i argininy, która znakuje TF w celu jego importu do jądra komórkowego. U roślin sygnały te różnią się sekwencją, organizacją i liczbą,

a niektóre TFs w ogóle ich nie posiadają [51]. Domena wiążąca DNA zawiera reszty aminokwasowe, które specyficznie kontaktują się z zasadami sekwencji promotora (TFBSs). Wiązanie TF z promotorem wzmacniają niespecyficzne oddziaływania kolejnych reszt aminokwasowych z grupą fosforanową lub deoksyrybozą [51]. Budowa domeny wiążącej DNA decyduje o podziale TFs na rodziny i wskazuje na ich funkcje [52]. Niektóre TFs działają samodzielnie, podczas gdy inne, celem zwiększenia swoistości i powinowactwa wiązania z DNA oraz dla precyzyjniejszej kontroli ekspresji genów, tworzą dimery lub struktury wyższego rzędu (oligomery). Jest to możliwe dzięki domenie oligomeryzacji [51]. Struktury te mogą powstawać z połączenia identycznych białek lub z połączenia różnych białek, zwykle z tej samej rodziny. Warto tutaj dodać, że pojedynczy TF może łączyć się z wieloma różnymi TFs, tworząc w ten sposób kompleksy o odmiennych właściwościach i funkcjach. O tym jaki kompleks zostanie utworzony decyduje stężenie poszczególnych monomerów w komórce, ich modyfikacje potranslacyjne, a także powinowactwo wiązania z innymi monomerami [53]. Część TFs tworzących takie struktury zachowuje zdolność do samodzielnego wiązania się z DNA, podczas gdy inne bezwzględnie wymagają obecności białkowego partnera [45]. W ostatnich latach dowiedziono, że niektóre TFs nie łączą się ze sobą bezpośrednio, a jedynie poprzez resztę białkową w miejscu wiązania się z DNA [54]. Domena efektorowa, a przez to TF który ją posiada, pośredniczy w aktywacji lub represji genu [55]. Aktywatory oddziałując z kompleksem inicjacji transkrypcji ułatwiają jego składanie i poprawiają stabilność, a także rekrutują enzymy rozluźniające strukturę nukleosomów, dzięki czemu staje się ona dostępna dla TFs. Z kolei represory oddziałując z aktywatorami blokują ich domenę efektorową lub tworzą kompleksy białkowe niezdolne do ich związania się z DNA. Represory wiążą się również z motywami rozpoznawanymi przez aktywatory czy z kompleksem inicjacji transkrypcji zmniejszając jego aktywność, a także przyciągają enzymy zwiększające upakowanie nukleosomów [45,56]. Dany TF może funkcjonować jako aktywator lub represor zależnie od dostępnych TFBSs i TFs [56].

mikroRNA (miRNA) to krótkie, jednoniciowe, niekodujące cząsteczki RNA znane głównie jako potranskrypcyjne i translacyjne inhibitory. Przecinają one nici mRNA i destabilizują je poprzez skrócenie ogonka poli(A) oraz redukują wydajność translacji [57,58]. Badania przeprowadzone u ludzi i *A. thaliana* wykazały, że miRNA mogą również rozpoznawać specyficzne motywy w sekwencji promotorów i w ten sposób brać udział w regulacji ekspresji genów na poziomie transkrypcji [59-62].

Oddziaływania typu TF−promotor, miRNA−promotor i TF−TF tworzą na poziomie całego organizmu skomplikowaną sieć zależności. Decydują one o ekspresji konkretnego genu we właściwym miejscu, momencie i na wymaganym poziomie, a co za tym idzie stoją na straży

procesów zachodzących w komórkach [46]. Dlatego też badanie TFs, miRNA i miejsc ich wiązania dostarcza niezwykle cennych informacji dla zrozumienia regulacji ekspresji genów.

Obecnie badanie sekwencji promotorów opiera się w znacznej mierze na analizach *in silico*. Wzrost mocy obliczeniowej komputerów, obniżenie kosztów przechowywania danych i szybki przyrost liczby kompletnych sekwencji genomowych zaowocował w ostatnich latach stworzeniem licznych, często publicznych baz danych i narzędzi opartych na algorytmach znajdujących zastosowanie w takich badaniach (Publikacja I) [63]. Wykorzystanie baz danych i narzędzi pozwala na przyspieszenie tempa badań, obniżenie ich kosztów oraz stanowi doskonały wstęp do prac eksperymentalnych. Do tej pory opracowano wiele programów różniących się założeniami i sposobem działania. Do dyspozycji są takie, które poszukują w badanej sekwencji motywów wcześniej poznanych albo zupełnie nowych [64]. Te pierwsze opierają się na danych doświadczalnych otrzymanych w przeszłości dla różnych organizmów i zgromadzonych w bazach danych. Drugie zaś poszukują motywów nadreprezentowanych [65]. Istnieją narzędzia skanujące zadaną sekwencję na obecność pojedynczych TFBS lub też ich kombinacji o znanej funkcji transkrypcyjnej [66]. Postawą do opracowania niektórych programów jest podział genów na klastry ze względu na wzory ich ekspresji [67]. Zakłada się, że geny należące do tego samego klastra ulegają wspólnej regulacji, a ich promotory posiadają takie same motywy wiązane przez takie same TFs. Programy te typują grupy oligonukleotydów wystarczająco podobnych do siebie aby mogły zostać rozpoznane przez ten sam TF i oceniają je pod względem statystycznym. W użyciu są także narzędzia, które wykorzystują technikę śladu filogenetycznego [66]. Wyodrębniają one motywy konserwatywne poprzez porównanie ze sobą wielu sekwencji ortologicznych pochodzących od różnych gatunków. Takie postępowanie zdecydowanie podnosi wiarygodność uzyskanych danych. Ryzyko pojawienia się fałszywych wyników podczas analiz *in silico* jest również obniżane poprzez dodanie parametrów określających prawdopodobieństwo dla otrzymanych przewidywań [68]. Wyniki analiz *in silico* dotyczących TFBSs mogą być przedstawiane w kilku formach. Jest to związane z tym, że dany TFBS wykazuje różnorodność sekwencyjną bez utraty funkcji. Wyróżnia się tutaj sekwencję zgodności zapisaną kodem IUPAC (ang. International Union of Pure and Applied Chemistry), PFM (ang. position frequency matrix) podającą liczbę poszczególnych nukleotydów w danej pozycji, PWM (ang. position weight matrix) określającą znormalizowaną częstość nukleotydów w danej pozycji lub logo sekwencji pozwalające na szybką wizualną ocenę częstości nukleotydów [69].

Aktywność genów można również regulować dzięki zdobyczom inżynierii genetycznej umożliwiającym m.in. nadekspresję. W jej wyniku powstaje więcej kopii pre-mRNA, co po modyfikacjach potranskrypcyjnych przekłada się na wzrost ilości mRNA i zwykle prowadzi do zwiększenia produkcji białka w porównaniu z niemodyfikowanym organizmem dzikim.

Doświadczenia tego typu można przeprowadzać na komórkach czy organizmie z którego badany gen pierwotnie pochodzi lub stosując obce systemy ekspresyjne, np. ssacze, owadzie, roślinne, drożdżowe, bakteryjne, bezkomórkowe [70]. Wybór odpowiedniego systemu zależy od wielu czynników, takich jak złożoność struktury białka, jego modyfikacje potranslacyjne, rozpuszczalność, lokalizacja komórkowa, względy finansowe. Niezwykle istotnym elementem jest zastosowanie odpowiedniego wektora ekspresyjnego, zwykle plazmidu bakteryjnego lub genomu wirusowego, o budowie odpowiadającej celom przeprowadzanego eksperymentu i jego założeniom. Najważniejszymi składowymi takiego wektora jest konstytutywny lub indukowalny promotor, sekwencja kodująca eksprymowanego genu, miejsce inicjacji translacji, terminator transkrypcji, gen oporności na antybiotyki [71]. Dobór odpowiedniej metody transformacji jest ściśle związany z typem wybranego systemu ekspresyjnego. W przypadku roślin proces ten często oparty jest o komórki *Rhizobium* (dawniej *Agrobacterium*) [72-74]. Wyróżnia się dwa typy nadekspresji, tj. absolutną i względną [75]. Pierwsza polega na wymianie natywnego promotora na taki, który silnie i konstytutywnie pobudzi ekspresję kontrolowanego genu. W drugiej wykorzystuje się plazmid wielokopijny, dzięki któremu następuje wzrost liczby kopii badanego genu. Technika nadekspresji jest wykorzystywana w biotechnologii do wytwarzania produktu genu na masową skalę czy też do badania jego biologicznej funkcji poprzez obserwację zmian fenotypowych, biochemicznych i sygnałowych następujących w jej wyniku [76]. W ostatnich latach przeprowadzono wiele badań, których celem było uzyskanie nadekspresji genów zaangażowanych w biosyntezę metabolitów wtórnych u roślin, np. flawonoidów [77], alkaloidów [78], terpenoidów [79], pochodnych kwasu benzoesowego [80], glikozydów cyjonogennych [81]. Wynikiem tych doświadczeń był wzrost poziomu produktu nadeksprymowanego genu, a co za tym idzie większa produkcja związków kolejno powstających w szlaku (nasilenie przepływu związków przez szlak) i często zwiększenie biosyntezy metabolitów.

2. Cel pracy

Głównym celem niniejszej rozprawy doktorskiej było zbadanie mechanizmów regulacji ekspresji genu *HMGR4* w gatunku *S. miltiorrhiza*. Cele szczegółowe obejmowały:

- analizę w warunkach *in silico* niepoznanej dotąd sekwencji promotora genu *HMGR4 S. miltiorrhiza* pod kątem obecności kluczowych motywów, miejsc wiązania dla TFs i miRNA oraz ocenę interakcji między wykrytymi TFs,
- porównanie TFs wykrytych metodą *in silico* z TFs potwierdzonymi doświadczalnie,
- porównanie dostępnych sekwencji promotorów *HMGR S. miltiorrhiza* (*HMGR1*, *HMGR2*, *HMGR4*) w kontekście wspólnych TFBSs i oddziałujących z nimi TFs,
- ocenę konserwatywności sekwencji promotora *HMGR4 S. miltiorrhiza* względem innych roślinnych promotorów *HMGR*,
- oznaczenie specyficznej dla typu organu roślinnego ekspresji genu *HMGR4 S. miltiorrhiza*,
- analizę wpływu wybranych regulatorów na ekspresję genu *HMGR4 S. miltiorrhiza*,
- zbadanie znaczenia nadekspresji genu *HMGR4* dla biosyntezy tanszinonów w gatunku *S. miltiorrhiza*,
- zbadanie wpływu wybranych regulatorów, typu organu roślinnego i środowiska hodowli na biosyntezę tanszinonów w transformantach *S. miltiorrhiza*.

3. Materiały i metody

3.1. Założenie hodowli *S. miltiorrhiza* (Publikacja II i III)

Materiał wyjściowy do założenia hodowli w warunkach glebowych (Publikacja II) oraz kultur *in vitro* (Publikacja III) *S. miltiorrhiza* stanowiły nasiona otrzymane z Ogrodu Roślin Leczniczych Uniwersytetu Medycznego w Łodzi. Hodowlę tradycyjną prowadzono w kompozytowej glebie przy dostępie naturalnego światła. Hodowle *in vitro* uzyskano poprzez odkażenie nasion, a następnie ich wykiełkowanie i wzrost pędów na podłożu podstawowym Murashige i Skoog (MS) [82] zestalonym agarem (7 g/l). Części nadziemne *S. miltiorrhiza* hodowano na stałym podłożu MS w warunkach fotoperiodu 16/8 h (światło/ciemność), a korzenie w płynnym podłożu B5 typu Gamborg'a [83] w ciemności.

3.2. Izolacja sekwencji promotora *HMGR4 S. miltiorrhiza* (Publikacja II)

Niepoznany dotąd region promotora *HMGR4 S. miltiorrhiza* wyizolowano za pomocą techniki genome walking [84], wykorzystując fragment sekwencji kodującej zdeponowany w GenBank pod numerem JN831103.1 oraz genomowy DNA otrzymany metodą wg Khan i wsp. [85] z młodych części nadziemnych hodowli glebowej *S. miltiorrhiza*. Amplifikowane fragmenty DNA poddano sekwencjonowaniu techniką Sangera.

3.3. Analizy *in silico* sekwencji promotora *HMGR4 S. miltiorrhiza* (Publikacja II)

Położenie promotora, kasety TATA, TSS i 5′UTR (ang. 5′ untranslated region) w badanej sekwencji wyznaczono za pomocą programu TSSP (Softberry Inc.), przewidującego potencjalne motywy za pomocą liniowej funkcji dyskryminacyjnej [86]. Narzędzie PlantPan 2.0 i zintegrowane z nim programy posłużyły do wskazania powtórzeń tandemowych, wysp CpG, TFBSs i oddziałujących TFs [87,88]. Identyfikacja powtórzeń tandemowych nastąpiła przy pomocy zmodyfikowanego programu Tandem repeats finder opartego o metodę słów (k-tuple) [89]. Do wykrycia wysp CpG wykorzystano aplikację CpGProD [90]. Skanowanie sekwencji na obecność TFBSs i TFs miało miejsce dzięki narzędziu Match [91], zawierającemu bibliotekę PWMs pochodzącą z bazy TRANSFAC, a także dzięki zastosowaniu PWMs z baz PLACE, AGRIS i JASPAR. Obecność miRNA i miejsc ich wiązania do sekwencji promotora i 5′UTR weryfikowano za pomocą narzędzia miRBase [92].

3.4. Porównanie TFs wykrytych *in silico* z TFs potwierdzonymi doświadczalnie (Publikacja II)

Na wstępie, używając narzędzi Protein BLAST (NCBI) i MEGA X (Pennsylvania State University) [93], wytypowano gen *HMGR A. thaliana* wykazujący najwyższą homologię wobec

genu *HMGR4 S. miltiorrhiza*. Właściwe analizy objęły znalezienie genów kodujących TFs, które reagują podobnie pod względem poziomów ekspresji jak wytypowany gen *HMGR A. thaliana*. Wykorzystano w nich narzędzia Expression Angler (BAR) [94] i *Arabidopsis* RNA-seq Database [95], posiadające dostęp do wyników ekspresji dla, odpowiednio, 22 000 genów *Arabidopsis* i 28 164 bibliotek RNA-seq *Arabidopsis*. Uzyskane wyniki porównano z danymi otrzymanymi za pomocą PlantPan 2.0, a następnie wyodrębniono pulę wspólnych TFs oraz sprawdzono możliwe oddziaływania zachodzące między nimi, wykorzystując narzędzia Pathway System (Genomatix) i BioGRID [96]. Dane na temat interakcji zawarte w zastosowanych narzędziach pochodziły albo z baz danych i literatury naukowej (Pathway System), albo wyłącznie z recenzowanych eksperymentalnych publikacji naukowych (BioGRID).

3.5. Porównanie sekwencji promotorów *HMGR S. miltiorrhiza* (Publikacja II)

Dostępne sekwencje promotorów *HMGR S. miltiorrhiza*, tj.: *HMGR1* (GU367911.1), *HMGR2* (KF297286.1) i *HMGR4* (KT921337.1), przeanalizowano za pomocą narzędzi Common TFs, FrameWorker i DiAlign TF (Genomatix). Common TFs użyto do wstępnej analizy wspólnych TFBSs i oddziałujących z nimi TFs zlokalizowanych w dowolnym miejscu badanych sekwencji. Narzędzie FrameWorker umożliwiło wyodrębnienie tylko takich TFBSs, które ułożone były w takiej samej kolejności i w określonym miejscu we wszystkich sekwencjach. Z kolei dzięki DiAlign TF znaleziono regiony konserwatywne i położone w ich obrębie TFBSs.

3.6. Ocena konserwatywności roślinnych promotorów *HMGR* (Publikacja II)

Konserwatywność 36 sekwencji obejmujących bliższe części promotorów i 5´UTR genów *HMGR* pochodzących od takich roślin jak: *A. thaliana*, *Arabidopsis lyrata*, *Glycine max*, *Gossypium hirsutum*, *Oryza sativa*, *Solanum lycopersicum*, *Zea mays* i *S. miltiorrhiza* sprawdzono poprzez ich nałożenie na siebie za pomocą algorytmu MUSCLE z narzędzia MEGA X [93]. Narzędzie DiAlign TF (Genomatix) posłużyło do rozpoznania TFBSs umiejscowionych w konserwatywnych regionach.

3.7. Oznaczenie specyficznej dla typu organu roślinnego ekspresji genu *HMGR4 S. miltiorrhiza* (Publikacja III)

Do badań wykorzystano liście, łodygi i korzenie pochodzące z założonych kultur *in vitro S. miltiorrhiza* oraz ilościową reakcję łańcuchową polimerazy w czasie rzeczywistym (ang. real-time quantitative polymerase chain reaction, RT-qPCR) ze starterami zaprojektowanymi wobec sekwencji genów *HMGR4* (JN831103.1) i *ACT7* (HM051058.1, gen referencyjny).

3.8. Wybór regulatorów ekspresji genu *HMGR4 S. miltiorrhiza* i traktowanie materiału roślinnego (Publikacja III)

Potencjalne regulatory ekspresji genu *HMGR4 S. miltiorrhiza* wytypowano na postawie analiz *in silico* sekwencji jego promotora pod kątem TFBSs i oddziałujących TFs, z wykorzystaniem narzędzia PlantPan 2.0 [87,88].

Wpływ roztworu 1 mg/l (2,89 μM) kwasu giberelinowego (ang. gibberellic acid, GA_3) lub 0,5 mg/l (2,85 μM) kwasu indolilo-3-octowego (ang. indole-3-acetic acid, IAA) bądź 20 mg/l (144,80 μM) kwasu salicylowego (ang. salicylic acid, SA) (Sigma-Aldrich) na aktywność *HMGR4* oceniono poprzez inkubację wyhodowanych roślin *S. miltiorrhiza z* ww. hormonami, a następnie przeprowadzono analizę materiału genetycznego pochodzącego z liści, używając metody RT-qPCR ze starterami specyficznymi dla genów *HMGR4* i *ACT7*.

3.9. Przygotowanie konstruktu nadekspresyjnego pRI201-AN-HMGR4, uzyskanie hodowli transformantów *S. miltiorrhiza* i traktowanie materiału roślinnego (Publikacja III)

Sekwencję kodującą genu *HMGR4 S. miltiorrhiza* zsyntetyzowano na podstawie sekwencji JN831103.1 i wstawiono do wektora pUC57 (Gene Universal Inc.) celem namnożenia, a następnie przeniesiono do binarnego wektora ekspresyjnego pRI201-AN (Takara Bio Inc.) w miejsce NdeI/SalI (Eurofins Genomics). Poprawność wstawki *HMGR4* i regionów flankujących sprawdzono poprzez sekwencjonowanie techniką Sangera.

Komórki kompetentne *Rhizobium radiobacter* (wcześniej *Agrobacterium tumefaciens*) GV2260 ($C58C1Rif^R$ z pGV2260) transformowano przygotowanym konstruktem pRI201-AN-HMGR4 lub pustym wektorem pRI201-AN, stosując metodę zamrażania i rozmrażania bakterii [97]. Hodowlę prowadzono na podłożach selekcyjnych. W celu potwierdzenia transformacji wyizolowano plazmidowy DNA metodą lizy alkalicznej [98] wraz z jego oczyszczaniem z użyciem mieszaniny fenol/chloroform/alkohol izoamylowy [99], a następnie wykorzystano go do amplifikacji fragmentu sekwencji genu oporności na kanamycynę metodą reakcji łańcuchowej polimerazy (ang. polymerase chain reaction, PCR). Otrzymane produkty rozdzielono za pomocą elektroforezy w żelu agarozowym.

Wirulencję transformowanych bakterii *R. radiobacter* indukowano w płynnym podłożu MS z dodatkiem acetosyringonu (Sigma-Aldrich) [100]. Liście *S. miltiorrhiza* z hodowli glebowej zakażono transformowanymi bakteriami i prowadzono kokultywację zgodnie ze zmodyfikowaną procedurą opisaną przez Dandekar i Fisk [100]. Zakażone liście przenoszono na świeże podłoże zabijające bakterie i stymulujące najpierw powstawanie kallusa, a następnie rozwój pędów. W kolejnych pasażach zastosowano antybiotyk selekcyjny, umożliwiający eliminację nietransformowanych roślin. Transformację potwierdzono poprzez izolację genomowego DNA z roślin, jego amplifikację metodą PCR i rozdział elektroforetyczny uzyskanych produktów.

Hodowle transformantów *S. miltiorrhiza* prowadzono zarówno w warunkach *in vitro* (części nadziemne na stałym podłożu MS w warunkach fotoperiodu 16/8 h i korzenie w płynnym podłożu B5 wg Gamborg'a w ciemności), jak i w kompozytowej glebie przy dostępie do naturalnego światła.

Wpływ transformacji konstruktem pRI201-AN-HMGR4 lub pustym wektorem pRI201-AN na ekspresję genu *HMGR4* zbadano w transformowanych liściach, łodygach i korzeniach *S. miltiorrhiza* hodowanych w warunkach *in vitro*, wykorzystując metodę RT-qPCR wraz ze starterami specyficznymi wobec genów *HMGR4* i *ACT7*.

Znaczenie nadekspresji genu *HMGR4*, typu organu roślinnego i środowiska hodowli dla biosyntezy tanszinonów oceniono w transformowanych liściach, łodygach i korzeniach *S. miltiorrhiza* rosnących w glebie oraz w warunkach *in vitro*.

Wpływ GA_3 (1 mg/l, 2,89 µM) lub IAA (0,5 mg/l, 2,85 µM) (Sigma-Aldrich) na produkcję tanszinonów oszacowano w liściach, łodygach i korzeniach *S. miltiorrhiza* z nadekspresją *HMGR4*, hodowanych *in vitro* i traktowanych hormonami.

Wszystkie oznaczenia zawartości tanszinonów w metanolowych ekstraktach roślinnych *S. miltiorrhiza* wykonano za pomocą opracowanej i zoptymalizowanej metody UHPLC (ang. ultra high performance liquid chromatography) z wykorzystaniem standardów DHTI, CT, TI i TIIA o czystości HPLC (ang. high performance liquid chromatography).

3.10. Analiza statystyczna wyników (Publikacja III)

Opracowaniu statystycznemu poddano wyniki badania ekspresji oraz zawartości tanszinonów, używając oprogramowania Statistica 13.3 (TIBCO Software Inc.). W pierwszej kolejności przeprowadzono analizę zgodności badanych cech z rozkładem normalnym, a następnie dokonano oceny istotności różnic między badanymi grupami za pomocą odpowiednich testów statystycznych. Wartości z $p < 0,05$ uznano za statystycznie istotne. Wyniki ekspresji przedstawiono w postaci średnich wraz z odchyleniem standardowym, a zawartości tanszinonów jako mediany z pierwszym i trzecim kwartylem.

Szczegółowy opis materiałów roślinnych, odczynników, aparatury, zastosowanych metod badawczych i testów statystycznych znajduje się w odpowiednich publikacjach będących podstawą niniejszej rozprawy doktorskiej.

4. Kluczowe wyniki, dyskusja i wnioski

4.1. Analizy *in silico* sekwencji promotora *HMGR4 S. miltiorrhiza* (Publikacja II)

Wyizolowaną sekwencję promotora *HMGR4 S. miltiorrhiza* o długości 1646 pz, zdeponowaną w bazie GenBank pod numerem KT921337.1 poddano analizom strukturalnym mogącym dostarczyć wielu cennych informacji na temat regulacji ekspresji kontrolowanego genu. W wyniku przeprowadzonych analiz zlokalizowano: TATA box (od −28 do −33 nt od TSS), powtórzenie tandemowe (od −1296 do −1353 nt od TSS), sekwencję bogatą w pirymidyny (ang. pyrimidine-rich sequence, PRS) w 5′UTR, 5369 potencjalnych TFBSs i 365 oddziałujących z nimi TFs opisanych wcześniej w gatunku *A. thaliana* oraz 12 miejsc wiązania dla dojrzałych miRNA zidentyfikowanych uprzednio u roślin; nie potwierdzono natomiast obecności wysp CpG.

Badania przeprowadzone na modelach ludzkich i drożdżowych dowiodły, że występowanie kasety TATA jest charakterystyczne dla genów stymulowanych czynnikami stresowymi i zewnątrzkomórkowymi, poddawanych ścisłej regulacji [101-105], a także może wskazywać na większą konserwatywność sekwencji promotora [106]. Z kolei geny pozbawione TATA box zwykle ulegają stałej ekspresji i powiązane są z kluczowymi procesami takimi jak wzrost komórek.

Powtórzenia tandemowe występują u 25% zbadanych promotorów [107]. Są one bardziej podatne na mutacje, co wpływa na długość powtórzenia i lokalne pozycjonowanie nukleosomów, prowadząc do wyższej zmienności transkrypcji.

Obecność PRS w 5′UTR jest rzadkim zjawiskiem i wiąże się często z wysokimi poziomami transkrypcji posiadającego ją genu [108].

Brak wysp CpG, a tym samym brak możliwości metylacji cytozyn w dinukleotydach CG wysp, może obniżać ryzyko hamowania ekspresji genu *HMGR4* [109,110]. Ponadto wyspy CpG są bardziej charakterystyczne dla genów ulegających konstytutywnej ekspresji [111].

Analizowana sekwencja okazała się być niezwykle bogata w TFBSs, co według przeprowadzonych wcześniej badań koreluje dodatnio z wyższą zmiennością ekspresji genów [112]. Na szczególną uwagę zasługują TFBSs zlokalizowane w części bliższej promotora, jako że to właśnie one najczęściej okazują się być funkcjonalne *in vivo* [113,114]. Analiza danych uzyskanych dla 666 takich TFBSs wykrytych w sekwencji promotora *HMGR4 S. miltiorrhiza* wykazała, że oddziałują one z TFs aktywowanymi głównie przez: światło (GATA), SA (WRKY), zakażenie bakteryjne (WRKY), auksyny (bZIP), kwas abscysynowy (MYB-related, WRKY i C2H2) i gibereliny (MYB-related i MADS box). Wydaje się wysoce prawdopodobne, że wymienione czynniki mogą być zaangażowane w regulację ekspresji badanego genu *HMGR4*, co jednak nie zostało dotąd potwierdzone doświadczalnie. Warto tutaj dodać, że takie oznaczenia

wykonano w przeszłości dla szerzej przebadanych genów *S. miltiorrhiza*, tj.: *HMGR*, *HMGR1* i *HMGR2* [115-121]. Dodatkowo przeprowadzone analizy wykazały obecność w sekwencji promotora *HMGR4 S. miltiorrhiza* miejsc wiązania dla następujących TFs o potwierdzonym pozytywnym wpływie na biosyntezę tanszinonów: BHLH6 (MYC2) (14 TFBSs), BHLH74 (2 TFBSs), BZIP20 (32 TFBSs), WRKY2 (8 TFBSs) i WRKY61 (9 TFBSs) [122-126]. Uzyskane wyniki mogą wskazywać na zaangażowanie, a nawet istotne znaczenie badanego genu *HMGR4* podczas biosyntezy tych metabolitów wtórnych.

Wykrycie miejsc wiązania dla dojrzałych miRNA w regionie badanego promotora i 5′UTR może świadczyć o udziale tych cząsteczek w regulacji ekspresji genu *HMGR4 S. miltiorrhiza* na poziomie transkrypcji. Warto tutaj dodać, że obecność dwóch spośród wykrytych w warunkach *in silico* miRNA, tj. miR1128 i miR1436, potwierdziło głębokie sekwencjonowanie małych RNA (ang. small RNA, sRNA) w gatunku *S. miltiorrhiza* [127].

4.2. Porównanie TFs wykrytych *in silico* z TFs potwierdzonymi doświadczalnie (Publikacja II)

Analiza wyników mikromacierzy wykazała istnienie 166 genów kodujących TFs, które reagują podobnie pod względem poziomów ekspresji jak wytypowany gen *HMGR1 A. thaliana* w zakresie współczynnika korelacji *r* 0,5−1,0. Porównanie wyników uzyskanych podczas analizy *in silico* za pomocą PlantPan 2.0 z wynikami z mikromacierzy ujawniło 32 wspólne TFs, głównie z rodzin HD-ZIP i WRKY, o zdolności wiązania się do promotora *HMGR4 S. miltiorrhiza*.

Analiza przeprowadzona za pomocą narzędzia Pathway System wskazała na istnienie potencjalnych interakcji między niektórymi ze wspólnych TFs, tj.: SVP−AGL18−SPL3, PDF2−ATML1−ANL2, EIL3−EIL1−EBP, ATHB13−HB-1, NAC3 z RD26 i ZF2. W przeważającej mierze wykryte interakcje dotyczyły występowania miejsc wiązania dla danego TF w promotorze innego TF.

Dodatkowo analiza wspólnych TFs wiążących się do bliższej części badanego promotora, wykonana przy użyciu narzędzia BioGRID, wykazała możliwość tworzenia dimerów. Struktury takie zwiększają swoistość i powinowactwo wiązania z DNA, co skutkuje precyzyjniejszą kontrolą ekspresji genu. Dotyczyło to TFs z następujących rodzin: HD-ZIP (ATML1, PDF2, HDG1), WRKY (WRKY2, WRKY14, WRKY45, WRKY57, WRKY69) i DOF (DOF5.4). Pewnym uwiarygodnieniem uzyskanych w warunkach *in silico* wyników są znalezione dane piśmiennicze. Według nich TFs należące do HD-ZIP nie są zdolne do wiązania się z DNA jako monomery [128], lecz dzięki motywowi zamka leucynowego formują homo- i heterodimery [53]. Dla przykładu jeden z członków tej rodziny, ATML1, był w stanie tworzyć homodimery w warunkach *in vitro* [129,130] oraz heterodimery z PDF2 u *Nicotiana benthamiana* i *A. thaliana* [129,131]. Z kolei przedstawiciele rodziny WRKY wchodzą w interakcje z DNA w postaci monomerów, homo- i heterodimerów [132-134]. Jak dotąd wykazano zdolność do tworzenia

homodimerów przez WRKY2 u *Hordeum vulgare* [135] oraz przez WRKY45 u *Oryza sativa* w warunkach *in vitro* [132]. Przeprowadzone badania wskazują również, że czynniki DOF dzięki posiadaniu wielofunkcyjnej domeny mogą wiązać się z DNA, wchodzić w interakcje z innymi białkami oraz formować homo- i heterodimery [136].

4.3. Porównanie sekwencji promotorów *HMGR S. miltiorrhiza* (Publikacja II)

Analiza przeprowadzona za pomocą narzędzia Common TFs wykazała występowanie w badanych promotorach wielu wspólnych TFBSs reprezentujących 22 rodziny macierzy zlokalizowanych w dowolnym miejscu sekwencji oraz uderzające podobieństwo ich rozkładu w obrębie *HMGR1* i *HMGR2*. Ocena TFs oddziałujących z tak wykrytymi TFBSs wskazała, że są to białka odpowiadające głównie na czynniki abiotyczne (auksyny, gibereliny, kwas abscysynowy, SA, kwas jasmonowy, brassinosteroidy, światło, brak wody, stres solny, chłód, niedobór fosforanów) i czynniki biotyczne (bakterie, grzyby, wirusy) oraz związane z organogenezą korzeni, łodyg, liści i kwiatów. Uzyskane wyniki mogą świadczyć o koregulacji genów *HMGR1*, *HMGR2* i *HMGR4 S. miltiorrhiza*.

Za pomocą narzędzia FrameWorker wykryto w obrębie badanych promotorów 10 000 10-elementowych układów TFBSs, w których poszczególne motywy ułożone były w takiej samej kolejności i w określonym miejscu we wszystkich badanych sekwencjach. Z danych literaturowych wynika, że struktury tego typu są zwykle powiązane z konkretną funkcją biologiczną lub określonym wzorem ekspresji genów i często wykazują konserwatywność [137].

Analiza z wykorzystaniem narzędzia DiAlign TF ujawniła, że największe podobieństwo wynoszące 97% wykazują sekwencje promotorów *HMGR1* i *HMGR2*. Podobieństwo między promotorami *HMGR2* i *HMGR4* osiągnęło 17%, a między *HMGR1* i *HMGR4* 14%. W miejscach gdzie wszystkie trzy badane sekwencje wykazywały duże lokalne podobieństwo, czyli we fragmentach bliższych oraz na początku i w środku dalszych regionów promotorów, wykryto siedem konserwatywnych TFBSs.

4.4. Ocena konserwatywności roślinnych promotorów *HMGR* (Publikacja II)

Przeprowadzona analiza ujawniła, że badane roślinne promotory *HMGR* i sekwencje 5′UTR, w tym *HMGR4 S. miltiorrhiza*, nie są wysoce konserwatywne, a jedynie zawierają pewne konserwatywne regiony. We fragmentach konserwatywnych zidentyfikowano: PRS, TATA box oraz 11 TFBSs. Najczęściej rozpoznawanym konserwatywnym motywem był TATA box występujący u 41,7% analizowanych sekwencji, podczas gdy pozostałe wykryte TFBSs okazały się wspólne dla co najwyżej 27,8% z nich. Opublikowane dotąd badania wskazują, że regiony niekodujące generalnie nie są silnie konserwatywne, a znalezione w ich obrębie konserwatywne motywy odznaczają się często funkcjonalną istotnością [138].

4.5. Oznaczenie specyficznej dla typu organu roślinnego ekspresji genu *HMGR4 S. miltiorrhiza* (Publikacja III)

Wyniki RT-qPCR wykazały, że gen *HMGR4* uległ ekspresji we wszystkich analizowanych organach *S. miltiorrhiza*, ale z różną intensywnością. Liście i łodygi odznaczały się wyższymi poziomami transkryptu *HMGR4* niż kalibrator, odpowiednio $1{,}14 \pm 0{,}08$ i $1{,}05 \pm 0{,}01$; w korzeniach poziom ten był niższy i wynosił $0{,}95 \pm 0{,}07$.

4.6. Wpływ GA_3, IAA, SA na ekspresję genu *HMGR4 S. miltiorrhiza* (Publikacja III)

Gen *HMGR4 S. miltiorrhiza* wykazał dwufazową odpowiedź na GA_3. W 12, 24 i 48 h miała miejsce stymulacja jego ekspresji względem nietraktowanej kontroli ($R > 1$), następnie jej obniżenie w 72 h ($R < 1$) i ponowny wzrost w 96 h ($R > 1$). W innych badaniach 2,89 µM GA_3 miał podobny wpływ na gen *HMGR2 S. miltiorrhiza* [118]. Jednak w tym przypadku ekspresja badanego genu wzrosła w porównaniu z kontrolą w 12 h, po czym nastąpił jej spadek i ponowny wzrost w 72 i 96 h. Wobec otrzymanych wyników można postawić hipotezę, że egzogenny GA_3 mógł stymulować ekspresję genu *HMGR4 S. miltiorrhiza* i produkcję kodowanego przez niego enzymu, wpływając na kolejne etapy szlaku MVA i produkcję mediatorów niezbędnych do biosyntezy endogennych giberelin, takich jak ent-kauren [139]. Nowo wyprodukowany endogenny GA_3 mógł następnie stymulować transkrypcję *HMGR4*, która uległa osłabieniu w wyniku zużycia egzogennego hormonu.

Wpływ IAA na ekspresję *HMGR4 S. miltiorrhiza* miał również dwufazowy charakter i podobny przebieg jak w doświadczeniach z GA_3. Pewną analogię do uzyskanych wyników prezentują badania Lv i wsp., według których 100 µM IAA najpierw podniósł poziom transkryptu *HMGR4 Malus domestica* w stosunku do kontroli, a następnie spowodował jego obniżenie [140]. Zaobserwowany dwufazowy efekt może wynikać z oddziaływania różnych TFs, z których niektóre stymulują ekspresję genu, a inne ją hamują.

Zastosowanie SA spowodowało wzrost ekspresji *HMGR4 S. miltiorrhiza* w 12, 24 i 72 h ($R > 1$) oraz spadek w 48 i 96 h ($R < 1$) w stosunku do materiału nietraktowanego. Podobny wpływ wywołał 10 mM SA wobec *HMGR3 Ginkgo biloba* [141]. Jednak w końcowej fazie tego badania (96 i 120 h) poziom mRNA *HMGR3* wzrósł w stosunku do kontroli. Z kolei w innych eksperymentach SA powodował ciągły wzrost poziomu transkryptu *HMGR* w korzeniach włośnikowatych *S. miltiorrhiza* [117] oraz *Salvia przewalskii* [142] względem kontroli.

Podsumowując, zastosowane hormony wywierały wpływ na ekspresję genu *HMGR4 S. miltiorrhiza*, potwierdzając tym samym wartość przeprowadzonych analiz *in silico*.

4.7. Wpływ transformacji konstruktem pRI201-AN-HMGR4 na ekspresję genu *HMGR4 S. miltiorrhiza* (Publikacja III)

We wszystkich organach roślin *S. miltiorrhiza* transformowanych konstruktem nadekspresyjnym pRI201-AN-HMGR4 zaobserwowano wyższe poziomy transkryptu *HMGR4* w porównaniu do kontroli (R > 1). Wartości te wynosiły dla łodyg, korzeni i liści, odpowiednio, $1{,}28 \pm 0{,}19$, $1{,}25 \pm 0{,}14$ i $1{,}10 \pm 0{,}14$. Otrzymane wyniki wskazują na prawidłowy przebieg transformacji i funkcjonowanie wprowadzonego konstruktu.

4.8. Wpływ nadekspresji genu *HMGR4* na biosyntezę tanszinonów w gatunku *S. miltiorrhiza* (Publikacja III)

Nadekspresja genu *HMGR4 S. miltiorrhiza* spowodowała istotny wzrost zawartości wszystkich oznaczanych tanszinonów w korzeniach względem kontroli ($p < 0{,}01$), tj. (pierwsze dwie wartości w nawiasach dotyczą hodowli w glebie; kolejne dwie wartości warunków *in vitro*): CT (2,43-krotny/5,39 mg/g s.m. lub 3,62-krotny/2,40 mg/g s.m.), DHTI (2,19-krotny/0,59 mg/g s.m. lub 2,47-krotny/0,44 mg/g s.m.), TI (1,86-krotny/0,71 mg/g s.m. lub 2,21-krotny/0,65 mg/g s.m.) i TIIA (1,51-krotny/1,88 mg/g s.m. lub 1,82-krotny/0,55 mg/g s.m.). Podobne wyniki uzyskano wcześniej w badaniach nad nadekspresją *HMGR* [21] i *HMGR2* [26] *S. miltiorrhiza*.

Ponadto nadekspresja *HMGR4 S. miltiorrhiza* indukowała zawartość TIIA do około 50 µg/g s.m. w łodygach i liściach rosnących w glebie i *in vitro*.

Uzyskane wyniki po raz pierwszy potwierdzają nie tylko udział, ale i istotne znaczenie badanego genu *HMGR4* w procesie biosyntezy oznaczanych tanszinonów. Wobec ciągle rosnącego zapotrzebowania, nadekspresja *HMGR4* wydaje się stanowić jedną z możliwych ścieżek prowadzących do podniesienia poziomu tych cennych leczniczo metabolitów.

4.9. Wpływ GA_3 oraz IAA na biosyntezę tanszinonów w transformantach *S. miltiorrhiza* (Publikacja III)

Dodatek GA_3 do hodowli *in vitro* korzeni *S. miltiorrhiza* wykazujących nadekspresję *HMGR4* podniósł istotnie poziomy CT (1,24-krotnie/0,79 mg/g s.m., $p = 0{,}0000$) i TIIA (1,07-krotnie/88,1 µg/g s.m., $p = 0{,}0404$), nie miał wpływu na DHTI i obniżył istotnie poziom TI (1,29-krotnie/0,27 mg/g s.m., $p = 0{,}0000$) w porównaniu z materiałem nietraktowanym. Wobec powyższych wyników można postawić hipotezę, że GA_3, poza genem *HMGR4*, może szczególnie silnie indukować ekspresję kluczowego enzymu(ów) zaangażowanych w końcowy etap biosyntezy CT, co przekłada się na wzrost zawartości tego tanszinonu i powstającego z niego TIIA. Hormon ten może nie wykazywać podobnego wpływu na produkcję DHTI oraz syntetyzowanego z niego TI [5].

Drugi zastosowany hormon, IAA, obniżył istotnie zawartość wszystkich badanych tanszinonów, tj.: CT (34,06-krotnie/3,21 mg/g s.m.), TIIA (11,49-krotnie/1,11 mg/g s.m.), DHTI (8,84-krotnie/0,66 mg/g s.m.) i TI (5,05-krotnie/0,96 mg/g s.m.) w hodowli *in vitro* korzeni *S. miltiorrhiza* z nadekspresją *HMGR4* względem nietraktowanej kontroli (p = 0,0000). Dla porównania spadki poziomów CT, TI i TIIA obserwowano również w hodowli korzeni włośnikowatych *S. miltiorrhiza* traktowanych 5,71 µM IAA [143].

Wpływ GA_3 lub IAA na poziom oznaczanych tanszinonów w łodygach i liściach *S. miltiorrhiza* rosnących w warunkach *in vitro* i wykazujących nadekspresję *HMGR4* był nieznaczny, z wahaniami zawartości wynoszącymi od 0,7 do 9,9 µg/g s.m. wobec nietraktowanej kontroli.

4.10. Zależna od typu organu roślinnego biosynteza tanszinonów w transformantach *S. miltiorrhiza* (Publikacja III)

Głównym miejscem wykrywania oznaczanych tanszinonów w gatunku *S. miltiorrhiza* okazały się korzenie z zawartością zależną od statusu nadekspresji *HMGR4* i środowiska wzrostu, tj.: 0,91–9,17 mg/g s.m. CT, 0,67–5,61 mg/g s.m. TIIA, 0,54–1,53 mg/g s.m. TI, 0,30–1,08 mg/g s.m. DHTI. Li i wsp. również wskazali korzenie, a dokładniej warstwę perydermy, jako podstawowe miejsce gromadzenia się wszystkich badanych tanszinonów; wewnętrzna warstwa korzeni i zewnętrzna część łodyg zawierała znacznie mniejsze ich ilości [16]. Dodatkowo analizy transkryptomiczne genów szlaków MVA i MEP oraz innych enzymów zaangażowanych w biosyntezę tanszinonów wykazały, że najsilniejsza ekspresja większości badanych genów (*AACT1−AACT6*, *HMGS2*, *HMGR1*, *HMGR2*, *MK*, *PMK*, *MDC1*, *MDC2*, *IPI1*, *GGPPS3*, *DXS2*, *DXS4*, *DXR*, *MCT*, *CMK*, *MDS*, *HDS*, *HDR1−HDR3*, *CPS1*, *CPS5*, *KSL1*, *KSL7*, *KSL8*, *CYP76AH1*) miała miejsce w perydermie korzeni *S. miltiorrhiza* [17]. Wydaje się zatem, że warstwa perydermy korzeni jest nie tylko głównym magazynem, ale także miejscem głównej biosyntezy tanszinonów.

Niewielką ilość oznaczanych tanszinonów wykryto również w łodygach i liściach *S. miltiorrhiza* z medianą wynoszącą od 50 do 73,5 µg/g s.m. Nadekspresja *HMGR4* spowodowała pojawienie się TIIA w łodygach i liściach, którego brak obserwowano w materiałach kontrolnych. Mimo to zawartość tego tanszinonu w częściach zielonych w stosunku do korzeni była około 104,5-krotnie niższa (5,55 mg/g s.m.) dla upraw glebowych i około 23,4-krotnie niższa (1,16 mg/g s.m.) dla warunków *in vitro*. Tanszinonów nie wykryto w kwiatach.

Uzyskane wyniki wskazują, że nadekspresja genu *HMGR4 S. miltiorrhiza* nie zmieniła zależnego od typu organu roślinnego wzoru biosyntezy tanszinonów. Korzenie nadal były ich głównym źródłem, a łodygi i liście zawierały śladowe ilości.

4.11. Wpływ środowiska wzrostu na biosyntezę tanszinonów w transformantach *S. miltiorrhiza* (Publikacja III)

Analiza statystyczna wskazała, że środowisko glebowe sprzyjało uzyskiwaniu istotnie wyższych poziomów wszystkich oznaczanych tanszinonów w porównaniu z warunkami *in vitro*, tj. (pierwsze dwie wartości w nawiasach dotyczą roślin z nadekspresją *HMGR4*; kolejne dwie wartości jej braku): TIIA (4,62-krotnie/4,39 mg/g s.m. lub 5,58-krotnie/3,06 mg/g s.m.), CT (2,77-krotnie/5,86 mg/g s.m. lub 4,14-krotnie/2,87 mg/g s.m.), DHTI (1,45-krotnie/0,34 mg/g s.m. lub 1,63-krotnie/0,19 mg/g s.m.), TI (1,28-krotnie/0,34 mg/g s.m. lub 1,52-krotnie/0,28 mg/g s.m.) w korzeniach ($p < 0,01$) oraz TIIA (1,12-krotnie/6,2 μg/g s.m.) w liściach z nadekspresją *HMGR4* ($p = 0,0000$). Takie wyniki mogą być spowodowane obecnością w środowisku naturalnym (ryzosfera, filosfera, endosfera) mikroorganizmów mających potencjał wpływania na biosyntezę metabolitów wtórnych [144,145]. Według Yan i wsp. endofityczne bakterie *Pseudomonas brassicacearum* subsp. *neoauraniaca* podniosły aktywność enzymów HMGR i DXS w korzeniach włośnikowatych *S. miltiorrhiza*, co ostatecznie spowodowało znaczny wzrost zawartości wszystkich oznaczanych tanszinonów, a szczególnie DHTI (19,2-krotny), CT (11,3-krotny) i całkowitej zawartości tanszinonów (3,7-krotny) w porównaniu z grupą kontrolną [119]. W kolejnym badaniu frakcja polisacharydowa otrzymana z ryzobakterii *Bacillus cereus* stymulowała (7-krotnie/1,4 mg/g s.m.) akumulację tanszinonów w korzeniach *S. miltiorrhiza* względem kontroli [146]. Innym potencjalnym powodem uzyskania niższych wydajności tanszinonów w warunkach *in vitro* mogą być zmiany zachodzące w morfologii, anatomii i fizjologii roślin podczas tego typu hodowli względem hodowli tradycyjnej [147,148].

Warto tutaj dodać, że korzenie rosnące w warunkach *in vitro* i wykazujące nadekspresję *HMGR4* odznaczały się wyższą zawartością DHTI (1,51-krotnie/0,25 mg/g s.m.) oraz TI (1,45-krotnie/0,37 mg/g s.m.) niż korzenie hodowane w glebie i niewykazujące tej nadekspresji. Zatem w przypadku tych dwóch tanszinonów nadekspresja genu *HMGR4* wydaje się kompensować mniej korzystne warunki *in vitro*.

Podsumowując doświadczenia z nadekspresją, najwyższą biosyntezę DHTI, CT, TI i TIIA uzyskano w korzeniach roślin *S. miltiorrhiza* z nadekspresją *HMGR4* hodowanych w warunkach glebowych.

5. Wnioski końcowe oraz koncepcja dalszych badań

W prezentowanej pracy przeprowadzono zaawansowane badania bioinformatyczne regionu promotora genu *HMGR4 S. miltiorrhiza*, co doprowadziło do uzyskania wielu nowych i cennych informacji na temat jego budowy oraz potencjalnych TFs i miRNA regulujących aktywność kontrolowanego przez niego genu. Część otrzymanych wyników została potwierdzona eksperymentalnie. Ponadto zaobserwowano korzystny wpływ nadekspresji sekwencji kodującej genu *HMGR4 S. miltiorrhiza* na zawartość oznaczanych tanszinonów w transformowanym materiale roślinnym, a szczególnie w korzeniach hodowanych w glebie oraz w warunkach *in vitro*. Uzyskana wiedza stanowi wypełnienie istniejącej dotąd luki dotyczącej mechanizmów regulacji ekspresji genów *HMGR S. miltiorrhiza* oraz ich udziału w procesie biosyntezy cennych leczniczo metabolitów wtórnych.

Jednym z najbardziej interesujących kierunków dalszych badań byłoby klonowanie, a następnie nadekspresja sekwencji kodującej wybranych TFs związanych z regulacją ekspresji genu *HMGR4 S. miltiorrhiza*. Umożliwiłoby to ocenę wpływu tych czynników na biosyntezę tanszinonów, a nawet na cały transkryptom i proteom w gatunku *S. miltiorrhiza*. Dodatkowo mutageneza wybranych TFBSs zlokalizowanych w sekwencji promotora *HMGR4 S. miltiorrhiza* mogłaby posłużyć do eksperymentalnej weryfikacji ich znaczenia dla odpowiedzi genu na czynniki biotyczne lub abiotyczne.

Z biotechnologicznego punktu widzenia w badaniach prowadzonych na kulturach *in vitro* roślin leczniczych istotnym celem jest nie tylko dążenie do otrzymania jak najwyższych zawartości metabolitów wtórnych, ale również jak najwyższej wydajności ich biosyntezy. Cel ten można osiągnąć poprzez uzyskanie korzeni transformowanych z użyciem bakterii glebowej *Rhizobium rhizogenes*, bazując na wyselekcjonowanych i opisanych w niniejszej rozprawie pędach *S. miltiorrhiza*, a następnie dokonując optymalizacji warunków prowadzenia hodowli oraz stopniowo zwiększając jej skalę.

6. Piśmiennictwo

1. Lu, S. Salvia miltiorrhiza: an economically and academically important medicinal plant. In *The Salvia miltiorrhiza genome*; Lu, S., Ed.; Springer: Switzerland, 2019; pp. 1-15.
2. Xu, Y.; Wan, R.Z.; Lin, Y.; Yang, L.; Chen, Y.; Liu, C. Recent advance on research and application of Salvia miltiorrhiza. *Asian J. Pharmacodyn. Pharmacokinet.* 2007, 7, 99-130.
3. Cheng, T.O. Cardiovascular effects of Danshen. *Int. J. Cardiol.* 2007, 121, 9-22.
4. Mei, X.D.; Cao, Y.F.; Che, Y.Y.; Li, J.; Shang, Z.P.; Zhao, W.J.; Qiao, Y.J.; Zhang, J.Y. Danshen: a phytochemical and pharmacological overview. *Chin. J. Nat. Med.* 2019, 17, 59-80.
5. Guo, J.; Ma, Y. Biosynthetic pathway of tanshinones in Salvia miltiorrhiza. In *The Salvia miltiorrhiza genome*; Lu, S., Ed.; Springer: Switzerland, 2019; pp. 129-139.
6. Zhang, H.; Long, M.; Wu, Z.; Han, X.; Yu, Y. Sodium tanshinone IIA silate as an add-on therapy in patients with unstable angina pectoris. *J. Thorac. Dis.* 2014, 6, 1794-1799.
7. Lee, A.R.; Wu, W.L.; Chang, W.L.; Lin, H.C.; King, M.L. Isolation and bioactivity of new tanshinones. *J. Nat. Prod.* 1987, 50, 157-160.
8. Jia, L.Q.; Zhang, N.; Xu, Y.; Chen, W.N.; Zhu, M.L.; Song, N.; Ren, L.; Cao, H.M.; Wang, J.Y.; Yang, G.L. Tanshinone IIA affects the HDL subfractions distribution not serum lipid levels: Involving in intake and efflux of cholesterol. *Arch. Biochem. Biophys.* 2016, 592, 50-59.
9. Wang, S.; Jing, H.; Yang, H.; Liu, Z.; Guo, H.; Chai, L.; Hu, L. Tanshinone I selectively suppresses pro-inflammatory genes expression in activated microglia and prevents nigrostriatal dopaminergic neurodegeneration in a mouse model of Parkinson's disease. *J. Ethnopharmacol.* 2015, 164, 247-255.
10. Wang, L.; Cheng, L.; Zhang, B.; Wang, N.; Wang, F. Tanshinone prevents alveolar bone loss in ovariectomized osteoporosis rats by up-regulating phosphoglycerate dehydrogenase. *Toxicol. Appl. Pharmacol.* 2019, 376, 9-16.
11. Chiu, S.C.; Huang, S.Y.; Chen, S.P.; Su, C.C.; Chiu, T.L.; Pang, C.Y. Tanshinone IIA inhibits human prostate cancer cells growth by induction of endoplasmic reticulum stress in vitro and in vivo. *Prostate Cancer Prostatic Dis.* 2013, 16, 315-322.
12. Kim, S.Y.; Moon, T.C.; Chang, H.W.; Son, K.H.; Kang, S.S.; Kim, H.P. Effects of tanshinone I isolated from Salvia miltiorrhiza bunge on arachidonic acid metabolism and in vivo inflammatory responses. *Phytotherapy research: PTR* 2002, 16, 616-620.
13. Tang, F.; Wu, X.; Wang, T.; Wang, P.; Li, R.; Zhang, H.; Gao, J.; Chen, S.; Bao, L.; Huang, H.; et al. Tanshinone II A attenuates atherosclerotic calcification in rat model by inhibition of oxidative stress. *Vascul. Pharmacol.* 2007, 46, 427-438.
14. Wang, C.; Du, X.; Yang, R.; Liu, J.; Xu, D.; Shi, J.; Chen, L.; Shao, R.; Fan, G.; Gao, X.; et al. The prevention and treatment effects of tanshinone IIA on oestrogen/androgen-induced benign prostatic hyperplasia in rats. *J. Steroid Biochem. Mol. Biol.* 2015, 145, 28-37.
15. Honda, G.; Koezuka, Y.; Tabata, M. Isolation of an antidermatophytic substance from the root of Salvia miltiorrhiza. *Chem. Pharm. Bull.* 1988, 36, 408-411.
16. Li, S.; Zhu, N.; Tang, C.; Duan, H.; Wang, Y.; Zhao, G.; Liu, J.; Ye, Y. Differential distribution of characteristic constituents in root, stem and leaf tissues of Salvia miltiorrhiza using MALDI mass spectrometry imaging. *Fitoterapia* 2020, 146.
17. Xu, Z.; Peters, R.J.; Weirather, J.; Luo, H.; Liao, B.; Zhang, X.; Zhu, Y.; Ji, A.; Zhang, B.; Hu, S.; et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. *The Plant journal: for cell and molecular biology* 2015, 82, 951-961.

18. Xu, H.; Song, J.; Luo, H.; Zhang, Y.; Li, Q.; Zhu, Y.; Xu, J.; Li, Y.; Song, C.; Wang, B.; et al. Analysis of the genome sequence of the medicinal plant Salvia miltiorrhiza. *Mol. Plant* 2016, 9, 949-952.

19. Jiang, Z.; Gao, W.; Huang, L. Tanshinones, critical pharmacological components in Salvia miltiorrhiza. *Front. Pharmacol.* 2019, 10, 202.

20. Ma, X.H.; Ma, Y.; Tang, J.F.; He, Y.L.; Liu, Y.C.; Ma, X.J.; Shen, Y.; Cui, G.H.; Lin, H.X.; Rong, Q.X.; et al. The biosynthetic pathways of tanshinones and phenolic acids in Salvia miltiorrhiza. *Molecules (Basel, Switzerland)* 2015, 20, 16235-16254.

21. Kai, G.; Xu, H.; Zhou, C.; Liao, P.; Xiao, J.; Luo, X.; You, L.; Zhang, L. Metabolic engineering tanshinone biosynthetic pathway in Salvia miltiorrhiza hairy root cultures. *Metab. Eng.* 2011, 13, 319-327.

22. Liao, P.; Hemmerlin, A.; Bach, T.J.; Chye, M.L. The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnol. Adv.* 2016, 34, 697-713.

23. Bach, T.J. Hydroxymethylglutaryl-CoA reductase, a key enzyme in phytosterol synthesis? *Lipids* 1986, 21, 82-88.

24. Hemmerlin, A. Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant science: an international journal of experimental plant biology* 2013, 203-204, 41-54.

25. Leivar, P.; Antolín-Llovera, M.; Ferrero, S.; Closa, M.; Arró, M.; Ferrer, A.; Boronat, A.; Campos, N. Multilevel control of Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase by protein phosphatase 2A. *The Plant cell* 2011, 23, 1494-1511.

26. Dai, Z.; Cui, G.; Zhou, S.F.; Zhang, X.; Huang, L. Cloning and characterization of a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from Salvia miltiorrhiza involved in diterpenoid tanshinone accumulation. *J. Plant Physiol.* 2011, 168, 148-157.

27. Ma, Y.; Yuan, L.; Wu, B.; Li, X.; Chen, S.; Lu, S. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in Salvia miltiorrhiza. *J. Exp. Bot.* 2012, 63, 2809-2823.

28. Liao, P.; Zhou, W.; Zhang, L.; Wang, J.; Yan, X.; Zhang, Y.; Zhang, R.; Li, L.; Zhou, G.; Kai, G. Molecular cloning, characterization and expression analysis of a new gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from Salvia miltiorrhiza. *Acta Physiol. Plant.* 2009, 31, 565-572.

29. Shi, M.; Luo, X.; Ju, G.; Yu, X.; Hao, X.; Huang, Q.; Xiao, J.; Cui, L.; Kai, G. Increased accumulation of the cardio-cerebrovascular disease treatment drug tanshinone in Salvia miltiorrhiza hairy roots by the enzymes 3-hydroxy-3-methylglutaryl CoA reductase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase. *Funct. Integr. Genom.* 2014, 14, 603-615.

30. Kai, G.; Liao, P.; Xu, H.; Wang, J.; Zhou, C.; Zhou, W.; Qi, Y.; Xiao, J.; Wang, Y.; Zhang, L. Molecular mechanism of elicitor-induced tanshinone accumulation in Salvia miltiorrhiza hairy root cultures. *Acta Physiol. Plant.* 2012, 34, 1421–1433.

31. Singh, K.B. Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol.* 1998, 118, 1111-1120.

32. Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.; Stiekema, W.; et al. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. *Nature* 1998, 391, 485-488.

33. Saunders, A.; Core, L.J.; Lis, J.T. Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* 2006, 7, 557-567.

34. Kwak, H.; Lis, J.T. Control of transcriptional elongation. *Annu. Rev. Genet.* 2013, 47, 483-508.

35. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* 2004, 11, 394-403.

36. Shuman, S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol.* 2001, 66, 1-40.

37. Lorković, Z.J.; Wieczorek Kirk, D.A.; Lambermon, M.H.; Filipowicz, W. Pre-mRNA splicing in higher plants. *Trends Plant Sci.* 2000, 5, 160-167.

38. Li, Q.; Hunt, A.G. The polyadenylation of RNA in plants. *Plant Physiol.* 1997, 115, 321-325.

39. Yasmeen, E.; Wang, J.; Riaz, M.; Zhang, L.; Zuo, K. Designing artificial synthetic promoters for accurate, smart, and versatile gene expression in plants. *Plant Commun.* 2023, 4, 100558.

40. Smale, S.T.; Kadonaga, J.T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 2003, 72, 449-479.

41. Joshi, C.P. An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 1987, 15, 6643-6653.

42. Molina, C.; Grotewold, E. Genome wide analysis of Arabidopsis core promoters. *BMC Genom.* 2005, 6.

43. Civán, P.; Svec, M. Genome-wide analysis of rice (Oryza sativa L. subsp. japonica) TATA box and Y Patch promoter elements. *Genome* 2009, 52, 294-297.

44. Smale, S.T. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* 1997, 1351, 73-88.

45. Gonzalez, D.H. Introduction to transcription factor structure and function. In *Plant transcription factors. Evolutionary, structural and functional aspects*; Gonzalez, D.H., Ed.; Academic Press: 2016; pp. 3-11.

46. Peremarti, A.; Twyman, R.M.; Gómez-Galera, S.; Naqvi, S.; Farré, G.; Sabalza, M.; Miralpeix, B.; Dashevskaya, S.; Yuan, D.; Ramessar, K.; et al. Promoter diversity in multigene transformation. *Plant Mol. Biol.* 2010, 73, 363-378.

47. Yang, Y.; Li, L.; Qu, L.J. Plant Mediator complex and its critical functions in transcription regulation. *J. Integr. Plant Biol.* 2016, 58, 106-118.

48. Heinz, S.; Romanoski, C.E.; Benner, C.; Glass, C.K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* 2015, 16, 144-154.

49. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* 2010, 463, 178-183.

50. Komarnytsky, S.; Borisjuk, N. Functional analysis of promoter elements in plants. In *Genetic engineering*; Setlow, J.K., Ed.; Kluwer Academic/Plenum Publishers: 2003; Volume 25, pp. 113-141.

51. Liu, L.; White, M.J.; MacRae, T.H. Transcription factors and their genes in higher plants functional domains, evolution and regulation. *European journal of pharmaceutical sciences: official journal of the European Federation for Pharmaceutical Sciences* 1999, 262, 247-257.

52. Luscombe, N.M.; Austin, S.E.; Berman, H.M.; Thornton, J.M. An overview of the structures of protein-DNA complexes. *Genome Biol.* 2000, 1, Reviews001.

53. Amoutzias, G.D.; Robertson, D.L.; Van de Peer, Y.; Oliver, S.G. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* 2008, 33, 220-229.

54. Ibarra, I.L.; Hollmann, N.M.; Klaus, B.; Augsten, S.; Velten, B.; Hennig, J.; Zaugg, J.B. Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nat. Commun.* 2020, 11, 124.

55. Frietze, S.; Farnham, P.J. Transcription factor effector domains. In *A handbook of transcription factors* Hughes, T.R., Ed.; Springer Science+Business Media B.V.: 2011; Volume 52, pp. 261-277.

56. Latchman, D.S. Inhibitory transcription factors. *Int. J. Biochem. Cell Biol.* 1996, 28, 965-974.

57. Engels, B.M.; Hutvagner, G. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* 2006, 25, 6163-6169.

58. Pillai, R.S.; Bhattacharyya, S.N.; Filipowicz, W. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.* 2007, 17, 118-126.

59. Yang, G.; Li, Y.; Wu, B.; Zhang, K.; Gao, L.; Zheng, C. MicroRNAs transcriptionally regulate promoter activity in Arabidopsis thaliana. *J. Integr. Plant Biol.* 2019, 61, 1128-1133.

60. 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, 859-770.

61. Place, R.F.; Li, L.C.; Pookot, D.; Noonan, E.J.; Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 1608-1613.

62. Kim, D.H.; Saetrom, P.; Snøve, O., Jr.; Rossi, J.J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 16230-16235.

63. Majewska, M.; Wysokińska, H.; Kuźma, Ł.; Szymczyk, P. Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview. *Gene* 2018, 644, 38-48.

64. Nain, V.; Sahi, S.; Kumar, P.A. In silico identification of regulatory elements in promoters. In *Computational biology and applied bioinformatics*; Lopes, H.S., Cruz, L.M., Eds.; InTech: Croatia, 2011; pp. 47-66.

65. Sandve, G.K.; Drabløs, F. A survey of motif discovery methods in an integrated framework. *Biol. Direct* 2006, 1, 11.

66. Vedel, V.; Scotti, I. Promoting the promoter. *Plant science: an international journal of experimental plant biology* 2011, 180, 182-189.

67. Qiu, P. Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. *Biochem. Biophys. Res. Commun.* 2003, 309, 495-501.

68. Rombauts, S.; Florquin, K.; Lescot, M.; Marchal, K.; Rouzé, P.; van de Peer, Y. Computational approaches to identify promoters and cis-regulatory elements in plant genomes. *Plant Physiol.* 2003, 132, 1162-1176.

69. Wasserman, W.W.; Sandelin, A. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* 2004, 5, 276-287.

70. Sodoyer, R. Expression systems for the production of recombinant pharmaceuticals. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* 2004, 18, 51-62.

71. Gill, D.R.; Pringle, I.A.; Hyde, S.C. Progress and prospects: the design and production of plasmid vectors. *Gene Ther.* 2009, 16, 165-171.

72. Bahramnejad, B.; Naji, M.; Bose, R.; Jha, S. A critical review on use of Agrobacterium rhizogenes and their associated binary vectors for plant transformation. *Biotechnol. Adv.* 2019, 37, 107405.

73. Krenek, P.; Samajova, O.; Luptovciak, I.; Doskocilova, A.; Komis, G.; Samaj, J. Transient plant transformation mediated by Agrobacterium tumefaciens: Principles, methods and applications. *Biotechnol. Adv.* 2015, 33, 1024-1042.

74. Barampuram, S.; Zhang, Z.J. Recent advances in plant transformation. In *Plant chromosome engineering. Methods and protocols*; Birchler, J., Ed.; Humana Press: The United States of America, 2011; Volume 701, pp. 1-35.

75. Moriya, H. Quantitative nature of overexpression experiments. *Mol. Biol. Cell* 2015, 26, 3932-3939.

76. Lloyd, A. Vector construction for gene overexpression as a tool to elucidate gene function. In *Plant functional genomics*; Grotewold, E., Ed.; Humana Press: The United States of America, 2003; Volume 236, pp. 329-344.

77. Muir, S.R.; Collins, G.J.; Robinson, S.; Hughes, S.; Bovy, A.; Ric De Vos, C.H.; van Tunen, A.J.; Verhoeyen, M.E. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat. Biotechnol.* 2001, 19, 470-474.

78. Canel, C.; Lopes-Cardoso, M.I.; Whitmer, S.; van der Fits, L.; Pasquali, G.; van der Heijden, R.; Hoge, J.H.; Verpoorte, R. Effects of over-expression of strictosidine synthase and tryptophan decarboxylase on alkaloid production by cell cultures of Catharanthus roseus. *Planta* 1998, 205, 414-419.

79. Mahmoud, S.S.; Croteau, R.B. Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proc. Natl. Acad. Sci. U. S. A.* 2001, 98, 8915-8920.

80. Verberne, M.C.; Verpoorte, R.; Bol, J.F.; Mercado-Blanco, J.; Linthorst, H.J. Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nat. Biotechnol.* 2000, 18, 779-783.

81. Bak, S.; Olsen, C.E.; Petersen, B.L.; Møller, B.L.; Halkier, B.A. Metabolic engineering of p-hydroxybenzylglucosinolate in Arabidopsis by expression of the cyanogenic CYP79A1 from Sorghum bicolor. *The Plant journal: for cell and molecular biology* 1999, 20, 663-671.

82. Murashige, T.; Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 1962, 15, 473-497.

83. Gamborg, O.L.; Miller, R.A.; Ojima, K. Nutrient requirements of suspension cultures of Soybean root cells. *Experimental cell research* 1968, 50, 151-158.

84. Shyamala, V.; Ames, G.F. Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene* 1989, 84, 1-8.

85. Khan, S.; Qureshi, M.I.; Kamaluddin; Alam, T.; Abdin, M.Z. Protocol for isolation of genomic DNA from dry and fresh roots of medicinal plants suitable for RAPD and restriction digestion. *Afr. J. Biotechnol.* 2007, 6, 175-178.

86. Shahmuradov, I.A.; Solovyev, V.V.; Gammerman, A.J. Plant promoter prediction with confidence estimation. *Nucleic Acids Res.* 2005, 33, 1069-1076.

87. Chow, C.N.; Zheng, H.Q.; Wu, N.Y.; Chien, C.H.; Huang, H.D.; Lee, T.Y.; Chiang-Hsieh, Y.F.; Hou, P.F.; Yang, T.Y.; Chang, W.C. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* 2016, 44, D1154-D1160.

88. Chang, W.C.; Lee, T.Y.; Huang, H.D.; Huang, H.Y.; Pan, R.L. PlantPAN: Plant promoter analysis navigator, for identifying combinatorial cis-regulatory elements with distance constraint in plant gene groups. *BMC genomics* 2008, 9, 561.

89. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999, 27, 573-580.

90. Ponger, L.; Mouchiroud, D. CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. *Bioinformatics (Oxford, England)* 2002, 18, 631-633.

91. Kel, A.E.; Gössling, E.; Reuter, I.; Cheremushkin, E.; Kel-Margoulis, O.V.; Wingender, E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* 2003, 31, 3576-3579.

92. Griffiths-Jones, S.; Grocock, R.J.; van Dongen, S.; Bateman, A.; Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 2006, 34, D140-D144.

93. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 2018, 35, 1547-1549.

94. Toufighi, K.; Brady, S.M.; Austin, R.; Ly, E.; Provart, N.J. The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. *The Plant journal: for cell and molecular biology* 2005, 43, 153-163.

95. Zhang, H.; Zhang, F.; Yu, Y.; Feng, L.; Jia, J.; Liu, B.; Li, B.; Guo, H.; Zhai, J. A comprehensive online database for exploring ~20,000 public Arabidopsis RNA-seq libraries. *Mol. Plant* 2020, 13, 1231-1233.

96. Oughtred, R.; Rust, J.; Chang, C.; Breitkreutz, B.J.; Stark, C.; Willems, A.; Boucher, L.; Leung, G.; Kolas, N.; Zhang, F.; et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein science* 2021, 30, 187-200.

97. Wise, A.A.; Liu, Z.; Binns, A.N. Three methods for the introduction of foreign DNA into Agrobacterium. In *Agrobacterium protocols*; Wang, K., Ed.; Humana Press: The United States of America, 2006; Volume 343, pp. 43-53.

98. Engebrecht, J.; Brent, R.; Kaderbhai, M.A. Minipreps of plasmid DNA. In *Current protocols in molecular biology*; Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., Eds.; John Wiley & Sons: The United States of America, 2003.

99. Moore, D.; Dowhan, D. Purification and concentration of DNA from aqueous solutions. In *Current protocols in molecular biology*; Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., Eds.; John Wiley & Sons: The United States of America, 2003.

100. Dandekar, A.M.; Fisk, H.J. Plant transformation. Agrobacterium-mediated gene transfer. In *Transgenic plants. Methods and protocols*; Peña, L., Ed.; Humana Press: The United States of America, 2005; pp. 35-46.

101. Tirosh, I.; Weinberger, A.; Carmi, M.; Barkai, N. A genetic signature of interspecies variations in gene expression. *Nat. Genet.* 2006, 38, 830-834.

102. Yang, C.; Bolotin, E.; Jiang, T.; Sladek, F.M.; Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 2007, 389, 52-65.

103. Basehoar, A.D.; Zanton, S.J.; Pugh, B.F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 2004, 116, 699-709.

104. Bae, S.H.; Han, H.W.; Moon, J. Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PloS one* 2015, 10, e0120848.

105. de Jonge, W.J.; O'Duibhir, E.; Lijnzaad, P.; van Leenen, D.; Groot Koerkamp, M.J.; Kemmeren, P.; Holstege, F.C. Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *EMBO J.* 2017, 36, 274-290.

106. Cliften, P.; Sudarsanam, P.; Desikan, A.; Fulton, L.; Fulton, B.; Majors, J.; Waterston, R.; Cohen, B.A.; Johnston, M. Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science (New York, N.Y.)* 2003, 301, 71-76.

107. Vinces, M.D.; Legendre, M.; Caldara, M.; Hagihara, M.; Verstrepen, K.J. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science (New York, N.Y.)* 2009, 324, 1213-1216.

108. Biłas, R.; Szafran, K.; Hnatuszko-Konka, K.; Kononowicz, A.K. Cis-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue and Organ Culture (PCTOC)* 2016, 127, 269 - 287.

109. Davey, C.; Pennings, S.; Allan, J. CpG methylation remodels chromatin structure in vitro. *J. Mol. Biol.* 1997, 267, 276-288.

110. Bell, A.C.; Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 2000, 405, 482-485.

111. Qiu, P. Computational approaches for deciphering the transcriptional regulatory network by promoter analysis. *BIOSILICO* 2003, 1, 125-133.

112. Tirosh, I.; Barkai, N. Evolution of gene sequence and gene expression are not correlated in yeast. *Trends in genetics: TIG* 2008, 24, 109-113.

113. Hernandez-Garcia, C.M.; Finer, J.J. Identification and validation of promoters and cis-acting regulatory elements. *Plant science: an international journal of experimental plant biology* 2014, 217-218, 109-119.

114. Yu, C.P.; Lin, J.J.; Li, W.H. Positional distribution of transcription factor binding sites in Arabidopsis thaliana. *Sci. Rep.* 2016, 6, 25164.

115. Chen, I.J.; Lee, M.S.; Lin, M.K.; Ko, C.Y.; Chang, W.T. Blue light decreases tanshinone IIA content in Salvia miltiorrhiza hairy roots via genes regulation. *J. Photochem. Photobiol. B* 2018, 183, 164-171.

116. Wang, C.H.; Zheng, L.P.; Tian, H.; Wang, J.W. Synergistic effects of ultraviolet-B and methyl jasmonate on tanshinone biosynthesis in Salvia miltiorrhiza hairy roots. *J. Photochem. Photobiol. B* 2016, 159, 93-100.

117. Hao, X.; Shi, M.; Cui, L.; Xu, C.; Zhang, Y.; Kai, G. Effects of methyl jasmonate and salicylic acid on tanshinone production and biosynthetic gene expression in transgenic Salvia miltiorrhiza hairy roots. *Biotechnol. Appl. Biochem.* 2015, 62, 24-31.

118. Szymczyk, P.; Grąbkowska, R.; Skała, E.; Żebrowska, M.; Balcerczak, E.; Jeleń, A. Isolation and characterization of a 3-hydroxy-3-methylglutaryl coenzyme A reductase 2 promoter from Salvia miltiorrhiza. *Journal of Plant Biochemistry and Biotechnology* 2018, 27, 223-236.

119. Yan, Y.; Zhang, S.; Zhang, J.; Ma, P.; Duan, J.; Liang, Z. Effect and mechanism of endophytic bacteria on growth and secondary metabolite synthesis in Salvia miltiorrhiza hairy roots. *Acta Physiol. Plant.* 2014, 36, 1095-1105.

120. Yan, Y.; Zhang, S.; Yang, D.; Zhang, J.; Liang, Z. Effects of Streptomyces pactum Act12 on Salvia miltiorrhiza hairy root growth and tanshinone synthesis and its mechanisms. *Appl. Biochem. Biotechnol.* 2014, 173, 883-893.

121. Yang, D.; Ma, P.; Liang, X.; Wei, Z.; Liang, Z.; Liu, Y.; Liu, F. PEG and ABA trigger methyl jasmonate accumulation to induce the MEP pathway and increase tanshinone production in Salvia miltiorrhiza hairy roots. *Physiol. Plant.* 2012, 146, 173-183.

122. Shi, M.; Wang, Y.; Wang, X.; Deng, C.; Cao, W.; Hua, Q.; Kai, G. Simultaneous promotion of tanshinone and phenolic acid biosynthesis in Salvia miltiorrhiza hairy roots by overexpressing Arabidopsis MYC2. *Industrial crops and products* 2020, 155, 112826.

123. Deng, C.; Hao, X.; Shi, M.; Fu, R.; Wang, Y.; Zhang, Y.; Zhou, W.; Feng, Y.; Makunga, N.P.; Kai, G. Tanshinone production could be increased by the expression of SmWRKY2 in Salvia miltiorrhiza hairy roots. *Plant science: an international journal of experimental plant biology* 2019, 284, 1-8.

124. Zhang, Y.; Xu, Z.; Ji, A.; Luo, H.; Song, J. Genomic survey of bZIP transcription factor genes related to tanshinone biosynthesis in Salvia miltiorrhiza. *Acta Pharm. Sin. B.* 2018, 8, 295-305.

125. Zhang, X.; Luo, H.; Xu, Z.; Zhu, Y.; Ji, A.; Song, J.; Chen, S. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in Salvia miltiorrhiza. *Sci. Rep.* 2015, 5, 11244.

126. Chen, Y.; Wang, Y.; Guo, J.; Yang, J.; Zhang, X.; Wang, Z.; Cheng, Y.; Du, Z.; Qi, Z.; Huang, Y.; et al. Integrated Transcriptomics and Proteomics to Reveal Regulation Mechanism and Evolution of SmWRKY61 on Tanshinone Biosynthesis in Salvia miltiorrhiza and Salvia castanea. *Front. Plant Sci.* 2021, 12, 820582.

127. Xu, X.; Jiang, Q.; Ma, X.; Ying, Q.; Shen, B.; Qian, Y.; Song, H.; Wang, H. Deep sequencing identifies tissue-specific microRNAs and their target genes involving in the biosynthesis of tanshinones in Salvia miltiorrhiza. *PloS one* 2014, 9, e111679.

128. Ariel, F.D.; Manavella, P.A.; Dezar, C.A.; Chan, R.L. The true story of the HD-Zip family. *Trends Plant Sci.* 2007, 12, 419-426.

129. San-Bento, R.; Farcot, E.; Galletti, R.; Creff, A.; Ingram, G. Epidermal identity is maintained by cell-cell communication via a universally active feedback loop in Arabidopsis thaliana. *The Plant journal: for cell and molecular biology* 2014, 77, 46-58.

130. Abe, M.; Takahashi, T.; Komeda, Y. Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. *The Plant journal: for cell and molecular biology* 2001, 26, 487-494.

131. Rombolá-Caldentey, B.; Rueda-Romero, P.; Iglesias-Fernández, R.; Carbonero, P.; Oñate-Sánchez, L. Arabidopsis DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box cis-element. *The Plant cell* 2014, 26, 2905-2919.

132. Cheng, X.; Zhao, Y.; Jiang, Q.; Yang, J.; Zhao, W.; Taylor, I.A.; Peng, Y.L.; Wang, D.; Liu, J. Structural basis of dimerization and dual W-box DNA recognition by rice WRKY domain. *Nucleic Acids Res.* 2019, 47, 4308-4318.

133. Eulgem, T.; Rushton, P.J.; Robatzek, S.; Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 2000, 5, 199-206.

134. Ciolkowski, I.; Wanke, D.; Birkenbihl, R.P.; Somssich, I.E. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol. Biol.* 2008, 68, 81-92.

135. Shen, Q.H.; Saijo, Y.; Mauch, S.; Biskup, C.; Bieri, S.; Keller, B.; Seki, H.; Ulker, B.; Somssich, I.E.; Schulze-Lefert, P. Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science (New York, N.Y.)* 2007, 315, 1098-1103.

136. Yanagisawa, S. Dof DNA-binding domains of plant transcription factors contribute to multiple protein-protein interactions. *European journal of pharmaceutical sciences: official journal of the European Federation for Pharmaceutical Sciences* 1997, 250, 403-410.

137. Werner, T.; Fessele, S.; Maier, H.; Nelson, P.J. Computer modeling of promoter organization as a tool to study transcriptional coregulation. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 2003, 17, 1228-1237.

138. Guilfoyle, T.J. The structure of plant gene promoters. In *Genetic engineering. Principles and methods*; Setlow, J.K., Ed.; Springer Science Business Media LLC: The United States of America, 1997; Volume 19, pp. 15-47.

139. Kasahara, H.; Hanada, A.; Kuzuyama, T.; Takagi, M.; Kamiya, Y.; Yamaguchi, S. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in Arabidopsis. *J. Biol. Chem.* 2002, 277, 45188-45194.

140. Lv, D.; Zhang, Y. Isolation and functional analysis of apple MdHMGR1 and MdHMGR4 gene promoters in transgenic Arabidopsis thaliana. *Plant Cell, Tissue and Organ Culture (PCTOC)* 2017, 129, 133-143.

141. Rao, S.; Meng, X.; Liao, Y.; Yu, T.; Cao, J.; Tan, J.; Xu, F.; Cheng, S. Characterization and functional analysis of two novel 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes (GbHMGR2 and GbHMGR3) from Ginkgo biloba. *Sci. Rep.* 2019, 9, 14109.

142. Li, J.; Li, B.; Luo, L.; Cao, F.; Yang, B.; Gao, J.; Yan, Y.; Zhang, G.; Peng, L.; Hu, B. Increased phenolic acid and tanshinone production and transcriptional responses of biosynthetic genes in hairy root cultures of Salvia przewalskii Maxim. treated with methyl jasmonate and salicylic acid. *Mol. Biol. Rep.* 2020, 47, 8565-8578.

143. Gupta, S.K.; Liu, R.; Liaw, S.; Chan, H.; Tsay, H. Enhanced tanshinone production in hairy roots of 'Salvia miltiorrhiza Bunge' under the influence of plant growth regulators in liquid culture. *Bot. Stud.* 2011, 52, 435-443.

144. Huang, W.; Long, C.; Lam, E. Roles of plant-associated microbiota in traditional herbal medicine. *Trends Plant Sci.* 2018, 23, 559-562.

145. Dojima, T.; Craker, L.E. Potential benefits of soil microorganisms on medicinal and aromatic plants. In *Medicinal and aromatic crops: production, phytochemistry, and utilization*; Jeliazkov (Zheljazkov), V.D., Cantrell, C.L., Eds.; Oxford University Press: The United States of America, 2016; pp. 75-90.

146. Zhao, J.; Zhou, L.; Wu, J. Promotion of Salvia miltiorrhiza hairy root growth and tanshinone production by polysaccharide–protein fractions of plant growth-promoting rhizobacterium Bacillus cereus. *Process Biochem.* 2010, 45, 1517-1522.

147. George, E.F.; Hall, M.A.; De Klerk, G. The anatomy and morphology of tissue cultured plants. In *Plant propagation by tissue culture 3rd edition*; George, E.F., Hall, M.A., De Klerk, G., Eds.; Springer Dordrecht: The Netherlands, 2008; pp. 465-477.

148. Preece, J.E.; Sutter, E.G. Acclimatization of micropropagated plants to the greenhouse and field. In *Micropropagation. Technology and application*; Debergh, P.C., Zimmerman, R.H., Eds.; Springer Dordrecht: The Netherlands, 1991; pp. 71-93.

7. Streszczenie

Salvia miltiorrhiza Bunge to jedna z najpopularniejszych roślin tradycyjnej medycyny chińskiej. Właściwości lecznicze zawdzięcza zdolności do biosyntezy ponad dwustu substancji bioaktywnych, w tym wykazujących wielokierunkowe działanie tanszinonów. Jak dowiedziono, tempo biosyntezy tanszinonów zależy w dużej mierze od reduktazy 3-hydroksy-3-metyloglutarylokoenzymu A (HMGR) ze szklaku kwasu mewalonowego. Do chwili obecnej zidentyfikowano pięć sekwencji genów *HMGR* w gatunku *S. miltiorrhiza* (*HMGR–HMGR4*), z których gen *HMGR4* nie został jeszcze przebadany.

Jednym z istotniejszych zagadnień biologii molekularnej jest poznanie mechanizmów regulacji aktywności genów od transkrypcji do potranslacyjnej modyfikacji białek i na poziomie epigenetyki, a także umiejętność wpływania na te procesy. U roślin kluczowy etap tej regulacji ma miejsce podczas transkrypcji, w której istotną rolę odgrywa promotor. W jego obrębie zachodzi formowanie kompleksu preinicjacyjnego niezbędnego do zapoczątkowania transkrypcji kontrolowanego genu. Umiejscowione są tu również motywy (TFBSs) rozpoznawane przez czynniki transkrypcyjne (TFs) funkcjonujące jako aktywatory lub represory transkrypcji oraz miejsca wiązania dla mikroRNA (miRNA). Stąd analiza budowy promotora oraz oddziałujących z nim białek i RNA ma fundamentalne znaczenie dla poznania mechanizmów regulacji ekspresji badanego genu.

W prezentowanych badaniach niepoznaną dotąd sekwencję promotora genu *HMGR4 S. miltiorrhiza* poddano szeregowi analiz w warunkach *in silico*. Uzyskane wyniki wskazały na obecność: kasety TATA, powtórzenia tandemowego, sekwencji bogatej w pirymidyny w 5′UTR, 5369 potencjalnych TFBSs i 365 wiążących się z nimi TFs oraz 12 miejsc wiązania dla dojrzałych miRNA; nie potwierdzono natomiast występowania wysp CpG. Dane literaturowe wskazują, że geny których promotory wykazują podobną budowę stymulowane są czynnikami stresowymi i zewnątrzkomórkowymi, a ich aktywność odznacza się wysoką zmiennością.

Ocena TFBSs zlokalizowanych w części bliższej badanego promotora *HMGR4 S. miltiorrhiza* i oddziałujących z nimi TFs wskazała, że w regulację ekspresji kontrolowanego genu mogą być głównie zaangażowane takie czynniki jak: światło, kwas salicylowy (SA), zakażenie bakteryjne, auksyny, kwas abscysynowy i gibereliny. Przeprowadzone doświadczenia potwierdziły wpływ kwasu giberelinowego (GA_3), kwasu indolilo-3-octowego (IAA) i SA na ekspresję badanego genu *HMGR4*.

Analiza przeprowadzona za pomocą narzędzia Pathway System wskazała na istnienie interakcji między niektórymi z TFs potencjalnie wiążących się do promotora genu *HMGR4 S. miltiorrhiza*, tj.: SVP−AGL18−SPL3, PDF2−ATML1−ANL2, EIL3−EIL1−EBP,

ATHB13−HB-1, NAC3 z RD26 i ZF2. W przeważającej mierze wykryte interakcje dotyczyły występowania miejsc wiązania dla danego TF w promotorze innego TF. Dodatkowo wykazano, że niektóre TFs mogą tworzyć dimery, co przyczynia się do zwiększenia swoistości i powinowactwa wiązania z promotorem oraz pozwala na bardziej precyzyjną kontrolę ekspresji genów. Dotyczyło to czynników oddziałujących z bliższą częścią badanego promotora *HMGR4* reprezentujących rodziny: HD-ZIP (ATML1, PDF2, HDG1), WRKY (WRKY2, WRKY14, WRKY45, WRKY57, WRKY69) i DOF (DOF5.4).

Analizy wykonane za pomocą narzędzi Common TFs i FrameWorker wykazały występowanie w sekwencjach promotorów *HMGR1*, *HMGR2* i *HMGR4 S. miltiorrhiza* wielu wspólnych TFBSs oraz ich układów. Ocena TFs wiążących się z wykrytymi TFBSs wskazała na możliwość koregulacji genów kontrolowanych przez te promotory w wyniku odpowiedzi na czynniki abiotyczne (auksyny, gibereliny, kwas abscysynowy, SA, kwas jasmonowy, brassinosteroidy, światło, brak wody, stres solny, chłód, niedobór fosforanów) i czynniki biotyczne (bakterie, grzyby, wirusy) oraz w czasie organogenezy korzeni, łodyg, liści i kwiatów.

W sekwencji promotora genu *HMGR4 S. miltiorrhiza* wykryto obecność 65 miejsc wiązania dla TFs o potwierdzonym pozytywnym wpływie na biosyntezę tanszinonów, tj.: BHLH6, BHLH74, BZIP20, WRKY2 i WRKY61. W badaniu przeprowadzonym z udziałem roślin *S. miltiorrhiza* transformowanych konstruktem nadekspresyjnym pRI201-AN-HMGR4 uzyskano istotny wzrost zawartości wszystkich oznaczanych tanszinonów w korzeniach (od 0,44 do 5,39 mg/g s.m.) oraz tanszinonu IIA (TIIA) w łodygach i liściach względem kontroli. Otrzymane wyniki po raz pierwszy potwierdzają nie tylko udział, ale również istotne znaczenie badanego genu *HMGR4 S. miltiorrhiza* w procesie biosyntezy tych metabolitów wtórnych. Nadekspresja *HMGR4* nie zmieniła jednak zależnego od typu organu roślinnego wzoru biosyntezy oznaczanych tanszinonów. Głównym ich źródłem pozostały korzenie, natomiast w łodygach i liściach stwierdzono śladową zawartość. Ponadto dodatek GA_3 do hodowli *in vitro* transformowanych korzeni *S. miltiorrhiza* podniósł istotnie poziomy kryptotanszinonu (o 0,79 mg/g s.m.) i TIIA (o 88,1 µg/g s.m.) w porównaniu z nietraktowaną kontrolą. Nastąpiło to przypuszczalnie w wyniku silnej indukcji ekspresji kluczowego enzymu(ów) zaangażowanych w końcowy etap biosyntezy tych tanszinonów.

Podsumowując, zaprezentowana budowa promotora *HMGR4* oraz wykryte potencjalnie oddziaływania typu TF−promotor, miRNA−promotor i TF−TF odsłaniają wycinek skomplikowanej sieci zależności decydującej o ekspresji genu *HMGR4* i innych genów *S. miltiorrhiza*.

8. Summary

Salvia miltiorrhiza Bunge is one of the most popular plants of traditional Chinese medicine. Its therapeutic properties are due to the ability to biosynthesise more than two hundred bioactive compounds, including tanshinones with multidirectional effects. It has been proven that the rate of tansinone biosynthesis depends largely on a 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) from the mevalonic acid pathway. To date, five *S. miltiorrhiza HMGR* gene sequences (*HMGR–HMGR4*) have been identified, of which the *HMGR4* gene has not yet been studied.

One of the most important issues of molecular biology is to understand the mechanisms of regulation of gene activity from transcription to post-translational modification of proteins and at the level of epigenetics, as well as the ability to influence these processes. In plants, a key step in this regulation occurs during transcription, where the promoter plays an important role. Within the promoter, the assembly of the preinitiation complex necessary for the initiation of gene transcription takes place. Moreover, motifs (TFBSs) recognised by transcription factors (TFs) (activators or repressors) and binding sites for microRNAs (miRNAs) are located here. Hence, the analysis of the promoter structure, the proteins and RNA interacting with it is fundamental to understanding the mechanisms regulating the expression of the studied gene.

In the present study, a new *S. miltiorrhiza HMGR4* promoter sequence was subjected to a series of *in silico* analyses. The results indicated the presence of a TATA box, tandem repeat, pyrimidine-rich sequence in the 5′UTR, 5369 potential TFBSs and 365 interacting TFs, 12 binding sites for mature miRNAs, and the absence of CpG islands. According to literature data, genes with promoters of similar structure are stimulated by stress and extracellular factors, and their activity is highly variable.

The evaluation of TFBSs from the proximal *HMGR4* promoter and interacting TFs indicated that light, salicylic acid (SA), bacterial infection, auxins, abscisic acid, and gibberellins may be mainly involved in the regulation of *S. miltiorrhiza HMGR4* gene expression. The conducted experiments confirmed the effects of gibberellic acid (GA_3), indole-3-acetic acid (IAA), and SA on its activity.

Analysis performed with the Pathway System tool revealed the existence of interactions between some of the TFs potentially binding to the *S. miltiorrhiza HMGR4* promoter, i.e. SVP−AGL18−SPL3, PDF2−ATML1−ANL2, EIL3−EIL1−EBP, ATHB13−HB-1, NAC3 with RD26 and ZF2. These interactions were predominantly related to the presence of binding sites for a given TF in the promoter of another TF. In addition, it has been shown that some TFs can form dimers, which increases the specificity and affinity of binding to the promoter and allows for more precise control of gene expression. This included factors interacting with the proximal part of the

studied *HMGR4* promoter representing the following families: HD-ZIP (ATML1, PDF2, HDG1), WRKY (WRKY2, WRKY14, WRKY45, WRKY57, WRKY69) and DOF (DOF5.4).

Analyses performed with the Common TFs and FrameWorker tools revealed the presence of multiple common TFBSs and their frameworks within the *S. miltiorrhiza HMGR1*, *HMGR2* and *HMGR4* promoters. The evaluation of TFs binding to these TFBSs indicated the possibility of co-regulation of controlled genes in response to abiotic factors (auxins, gibberellins, abscisic acid, SA, jasmonic acid, brassinosteroids, light, water deprivation, salt stress, cold, phosphate deficiency) and biotic factors (bacteria, fungi, viruses) and during root, stem, leaf and flower organogenesis.

In the *S. miltiorrhiza HMGR4* promoter sequence, 65 sites for TFs with a confirmed positive effect on the biosynthesis of tanshinones were detected, i.e. BHLH6, BHLH74, BZIP20, WRKY2 and WRKY61. In a study with *S. miltiorrhiza* plants transformed with a pRI201-AN-HMGR4 overexpression construct, there was a significant increase in the content of all tested tanshinones in roots (from 0.44 to 5.39 mg/g DW) and tanshinone IIA (TIIA) in stems and leaves relative to control. The results confirm for the first time not only the involvement, but also the important role of the *S. miltiorrhiza HMGR4* gene in the biosynthesis of these secondary metabolites. However, *HMGR4* overexpression did not change the characteristic organ-dependent pattern of tanshinone biosynthesis. Roots remained the main source, while trace amounts were found in stems and leaves. Furthermore, the addition of GA_3 to transformed *S. miltiorrhiza in vitro* root culture significantly increased cryptotanshinone (by 0.79 mg/g DW) and TIIA (by 88.1 µg/g DW) levels in comparison to untreated roots. This presumably occurred as a result of strong induction of the expression of some key enzyme(s) involved in the terminal stage of CT and TIIA biosynthesis.

In conclusion, the presented *HMGR4* promoter structure and the detected potential TF−promoter, miRNA−promoter and TF−TF interactions reveal a fragment of the complex network of relationships determining the expression of the *HMGR4* gene and other *S. miltiorrhiza* genes.

„Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview”
(Publikacja I)

Gene 644 (2018) 38–48

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Review

Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview

Małgorzata Majewska[a,*], Halina Wysokińska[a], Łukasz Kuźma[a], Piotr Szymczyk[b]

[a] Department of Biology and Pharmaceutical Botany, Medical University of Lodz, 90-151 Lodz, Poland
[b] Department of Pharmaceutical Biotechnology, Medical University of Lodz, 90-151 Lodz, Poland

ARTICLE INFO

Keywords:
Database
Promoter
Transcription factor
miRNA
Binding site
Gene regulation

ABSTRACT

The complete exploration of the regulation of gene expression remains one of the top-priority goals for researchers. As the regulation is mainly controlled at the level of transcription by promoters, study on promoters and findings are of great importance. This review summarizes forty selected databases that centralize experimental and theoretical knowledge regarding the organization of promoters, interacting transcription factors (TFs) and microRNAs (miRNAs) in many eukaryotic and prokaryotic species. The presented databases offer researchers valuable support in elucidating the regulation of gene transcription.

1. Introduction

Nowadays, one of the most important challenges in Biology is to gain a complete understanding of the regulation of gene expression. At the level of transcription, this regulation is largely dependent on promoters. Promoters are able to bind proteins called transcription factors (TFs), recruited RNA polymerase and thereby regulate gene transcription (Zawel and Reinberg, 1995; Latchman, 1997; Browning and Busby, 2004). Additionally, accumulating evidence suggests that also microRNAs (miRNAs) are involved in transcriptional gene regulation by targeting promoter elements (Huang et al., 2012; Younger and Corey, 2011; Bartel, 2004). However, the mechanism is not clear yet. The acquired knowledge about investigated promoters is often integrated into databases. By providing comprehensive information, the databases facilitate and even guide future research. Moreover, they can also act as a basis for the development of tools that enable in silico analysis of new promoter sequences with regard to gene transcription regulation (Ohler and Niemann, 2001; Qiu, 2003; Wasserman and Sandelin, 2004). Therefore, the databases have become a cornerstone of modern Biology.

The aim of the present work is to summarize forty selected, functioning publicly-available web-based databases (excluding tools used only for in silico analysis) that provide experimental and theoretical knowledge concerning promoters as regulators of transcription of genes in eukaryotes and prokaryotes. Despite our efforts, we were not able to incorporate all existing resources into the review. The data describe the organization of promoters and their interaction with TFs and miRNAs. The selected repositories were classified according to organism and the type of data they present. To provide the fullest possible analysis, several similar databases were chosen from each given area, if available, and examined together. This is a comprehensive and up-to-date review on this important topic. We believe that our manuscript can successfully assist researchers exploring the regulation of gene

Abbreviations: AGRIS, *Arabidopsis* Gene Regulatory Information Server; AtcisDB, *Arabidopsis thaliana cis*-regulatory element database; AtPAN, *Arabidopsis thaliana* Promoter Analysis Net; AtRegNet, *Arabidopsis thaliana* regulatory network; AtTFDB, *Arabidopsis thaliana* transcription factor database; BBLS, Bayesian Branch Length Score; BLAST, Basic Local Alignment Search Tool; B1H, bacterial one-hybrid system; ChIP-chip, chromatin immunoprecipitation-on-chip; ChIP-seq, chromatin immunoprecipitation-sequencing; DBTBS, The Database of Transcriptional Regulation in *Bacillus subtilis*; DCPD, *Drosophila* Core Promoter Database; DNase-seq, DNase sequencing; DoOP, Database of Orthologous Promoters; DPE, downstream promoter element; EPD, Eukaryotic Promoter Database; ESTs, expressed sequence tags; FDR, false discovery rate; FFLs, feed-forward regulatory loops; GRN, gene regulatory network; GTRD, Gene Transcription Regulation Database; hg18, human genome 18; hg19, human genome 19; HOCOMOCO, *Homo sapiens* Comprehensive Model Collection; HT-SELEX, high-throughput systematic evolution of ligands by exponential enrichment; Inr, initiator element; LDSS, local distribution of short sequences; miRNA, microRNA; MFE, minimum free energy; NGS, next-generation sequencing; NLOD, normalized log-odds; PBM, protein binding microarray; PFMs, position frequency matrices; PlantPAN, Plant Promoter Analysis Navigator; PlantProm DB, Plant Promoter Database; ppdb, plant promoter database; PPIs, protein-protein interactions; PRODORIC, Prokaryotic Database of Gene Regulation; Pro54DB, σ^{54} promoter database; PWMs, position weight matrices; RACE, 5′ rapid amplification of cDNA ends; REGs, regulatory element groups; ReIN, Regulatory Networks Interaction Module; RNA-seq, RNA sequencing; SCPD, *Saccharomyces cerevisiae* Promoter Database; SELEX, systematic evolution of ligands by exponential enrichment; SNP, single nucleotide polymorphism; sRNA, small RNA; STIFDB, Stress-Responsive Transcription Factor Database; TF, transcription factor; TFBSs, transcription factor binding sites; TFFMs, transcription factor flexible models; TRED, Transcriptional Regulatory Element Database; TSSs, transcription start sites; UniPROBE, Universal PBM Resource for Oligonucleotide Binding Evaluation; UTR, untranslated region; YCRD, Yeast Combinatorial Regulation Database; YEASTRACT, Yeast Search for Transcriptional Regulators and Consensus Tracking; YPA, Yeast Promoter Atlas

* Corresponding author at: Department of Biology and Pharmaceutical Botany, Medical University of Lodz, Muszyńskiego 1, 90-151 Lodz, Poland.
E-mail address: malgorzata.majewska1@stud.umed.lodz.pl (M. Majewska).

http://dx.doi.org/10.1016/j.gene.2017.10.079
Received 12 May 2017; Received in revised form 26 July 2017; Accepted 27 October 2017
Available online 02 November 2017

expression to find a suitable database by markedly reducing the time needed to identify one.

2. Eukaryotic databases

2.1. Multispecies eukaryotic databases

One of the oldest promoter databases is the Eukaryotic Promoter Database (EPD) (Cavin Perier et al., 1998). It is a collection of 4806 RNA polymerase II promoters with transcription start sites (TSSs). The data presented in the EPD are based on TSS mapping experiments published in journal articles or on expressed sequence tags (ESTs) of full-length cDNA clones used for in silico primer extension. The user is able to download selected species-specific promoter sequences (−499: +100 bp relative to TSS). In 2011, a new section called the EPDnew was added (Dreos et al., 2013). In contrast to the EPD, the EPDnew contains promoters automatically assembled from next-generation sequencing (NGS) and from high-throughput promoter mapping experiments. The EPDnew currently supports 10 species: six animal species (*Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Apis mellifera*, *Caenorhabditis elegans*, *Danio rerio*), two plant species (*Arabidopsis thaliana*, *Zea mays*), and two fungus species (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*). The EPDnew provides over 105,000 promoter sequences and information on their motifs, i.e., TATA box, GC box, CCAAT box, initiator element (Inr), CpG island. The user is free to use various analysis tools to examine promoters (frequency and occurrence of motifs) and selection tools to select subsets of promoters. The EPD and EPDnew are regularly updated, with the most recent update on May 2017.

ECRbase provides a collection of vertebrate promoter sequences with transcription factor binding sites (TFBSs); it contains *H. sapiens*, *Macaca mulatta*, *Pan troglodytes*, *Bos taurus*, *Canis familiaris*, *Monodelphis domestica*, *Rattus norvegicus*, *M. musculus*, *Gallus gallus*, *Xenopus tropicalis*, *D. rerio* and *Takifugu rubripes* annotations (Loots and Ovcharenko, 2007). TFBSs are recognized based on position weight matrices (PWMs) extracted from the TRANSFAC database (Matys et al., 2006) and the tfSearch algorithm (Loots and Ovcharenko, 2004). To mitigate the problem of false positive results, the authors independently optimized thresholds for different TFBSs. The database gives the possibility to download the data. It was recently updated in 2009.

The Transcriptional Regulatory Element Database (TRED) is a repository of human, rat, and mouse *cis*- and *trans*-regulatory elements (Zhao et al., 2005; Jiang et al., 2007). The user is able to retrieve 139,379 promoter sequences with TSSs. The promoters are extracted from experimental literature and are derived by computational predictions. It is also possible to search for TFBSs and obtain information on their sequence, localization, and corresponding TFs. Furthermore, the database provides access to orthologous genes and gene regulatory networks (GRNs) constructed by 34 TF families implicated in cancer pathways and their target genes. The TRED is simple in construction and comprehensive database. The last update of the database could not be determined.

The mechanisms of regulation of genes encoding miRNAs are depicted by DIANA miRGen (Georgakilas et al., 2016). It is a unique repository of 276 precise TSSs of miRNA genes, and over 19,000 binding sites for 202 TFs obtained from nine cell lines and six tissues of *H. sapiens* and *M. musculus*. The user is able to receive information on miRNAs, the genomic position of TSSs and binding sites, binding motif logos and expression of TFs. The promoters incorporated into the database were obtained from experimental and computational studies. The TSSs were predicted by using the microTSS algorithm (Georgakilas et al., 2014) on chromatin immunoprecipitation-sequencing (ChIP-seq), RNA sequencing (RNA-seq), and DNase sequencing (DNase-seq) data. The algorithm is able to identify miRNA TSSs to a resolution of one nucleotide, with 93,6% sensitivity and 100% precision. Binding sites for TFs were estimated by scanning promoter sequences based on position frequency matrices (PFMs). The last update of the database could not be determined.

The CircuitsDB connects transcriptional and post-transcriptional interactions of TFs, miRNAs and genes to establish feed-forward regulatory loops (FFLs) (Friard et al., 2010). A master TF regulates miRNA in the loop, and both regulate target protein-coding genes. The circuits were created here based on ab-initio sequence analysis of *M. musculus* and *H. sapiens* genomes. The CircuitsDB provides detailed information on target genes and miRNAs (including disease involvement), TFs, binding sites for miRNAs in the 3´untranslated region (3′UTR), putative TFBSs in promoters, and tissue expression of loop components. Conserved overrepresented oligos (putative TFBSs) were identified by promoter scanning based on consensus sequences from the TRANSFAC database (Matys et al., 2006). In addition, the CircuitsDB includes 21,446 human (21,316 protein-coding and 130 for pre-miRNA) and 21,944 mouse (21,814 protein-coding and 130 pre-miRNA) promoter regions. In total, the transcriptional network consists of 43,453 genes and 4062 binding sites for 230 TFs, while the post-transcriptional network involves 33,021 genes and 360 binding sites for 283 mature miRNAs. The user can explore transcriptional and post-transcriptional networks separately. The database is flexible, it is possible to query by TF, by miRNA or by target gene and download the data. The last update of the database could not be determined.

A comparable database is RegNetwork (Liu et al., 2015). It is a comprehensive collection of experimentally confirmed and predicted transcriptional and post-transcriptional regulatory networks extracted from 25 databases, including JASPAR (Bryne et al., 2008) and the TRED (Jiang et al., 2007) described here. TF-gene, TF-miRNA, TF-TF, miRNA-TF, miRNA-gene interactions are considered. Each of the interactions is labeled with a degree of confidence (high, medium or low). The human part of RegNetwork consists of 1456 TFs, 1904 miRNAs and 19,719 target genes. The mouse unit comprises 1328 TFs, 1290 miRNAs and 18,120 target genes. The obtained results can be downloaded. The database was recently updated on 1st July 2017.

The microPIR2 database provides a number of possible miRNA binding sites in promoters: about 80 million human and 40 million mouse (Piriyapongsa et al., 2014). The sites were predicted using the RNAhybrid program (Kruger and Rehmsmeier, 2006). The user can also search for conserved miRNA binding sites that can be powerful candidates for subsequent experimental analysis. As a comprehensive database, microPIR2 supplies detailed knowledge about miRNAs, target genes and binding sites (length, position, overlapping with other important sequences, minimum free energy (MFE), *p*-value, conserved score), together with links to other databases. The user can query by gene, miRNA, disease and binding site characteristics. The last update of the database could not be determined.

A comparable repository is miRWalk2.0 (Dweep et al., 2011). It comprises miRNA binding sites in human, mouse, and rat promoters. The sites were identified with the miRWalk algorithm and an additional nine programs (DianamT, miRanda, miRDB, Pictar4, Pictar5, PITA, RNA22, RNAhybrid, Targetscan), to improve prediction. Apart from miRNA binding sites, the database provides information on miRNAs and target genes. The user can search by target gene, miRNA, pathway, class of gene and protein, gene ontology, disorders, diseases and human phenotype ontology. miRWalk2.0 is regularly updated. The last update was on June 2017.

The Database of Orthologous Promoters (DoOP) (Barta et al., 2005) and its tool DoOPSearch (Sebestyen et al., 2009) were created to identify evolutionarily-conserved promoter sequences (possible TFBSs) of a given gene. This method is called phylogenetic footprinting. The conserved sequences were found by multiple alignments using the DIALIGN2 program (Morgenstern, 1999). The DoOP comprises over one million such orthologous sequences (6–50 bp long) from chordate and plant promoters. The database offers the possibility to search for specific promoter clusters (500, 1000 or 3000 bp long) or sequence motifs using the DoOPSearch tool, giving a list of characterized genes and a

graphical map of motifs along promoter and promoter cluster sequences. The gene list can be subsequently examined with the GeneMerge program to identify statistically overrepresented Gene Ontology terms and hence the putative function of the motifs and genes. It is worth noting that it is possible to obtain false positive results using this approach, as the sensitivity and specificity of the DoOPSearch and GeneMerge analyses depend, respectively, upon the chosen parameters and the number of genes launched. The last update of the databases could not be determined.

One of the eukaryotic databases that characterizes TF binding profiles is JASPAR (Mathelier et al., 2016). It consists of several collections, but the best known is JASPAR CORE. The CORE collection contains 1082 TF binding profiles for many species in six taxonomic groups. It includes 519 profiles for vertebrates, 227 for plants, 133 for insects, 176 for fungi, 26 for nematodes and 1 for urochordata. The profiles are described as PFMs and transcription factor flexible models (TFFMs), and are derived from protein binding microarray (PBM), high-throughput systematic evolution of ligands by exponential enrichment (HT-SELEX), and ChIP-seq experiments. JASPAR provides information not only on TF binding profiles, but also supplies general knowledge about TFs with links to external databases. The database is supported by various open-source software tools. Ruby Gem uses TF binding profiles for predicting TFBSs within DNA sequences. The Bioconductor TFBSTools package is used for the analysis and manipulation of TFBSs and their associated TF profile matrices. JASPAR is a regularly-updated intuitive database. It was recently updated in 2016.

Homo sapiens Comprehensive Model Collection (HOCOMOCO) represents a complement to the JASPAR database (Kulakovskiy et al., 2016). It comprises DNA patterns (with quality score) and corresponding TF characteristics. The database involves PWMs for 601 human and 396 mouse TFs. It also contains completely new complementary dinucleotide PWM models for 86 human and 52 mouse TFs. This dinucleotide PWMs are based on ChIP-seq data extracted from the GTRD database (Yevshin et al., 2017). Part of the assembled DNA patterns come from HT-SELEX experiments and the rest from well-known databases, e.g., JASPAR (Mathelier et al., 2014) and SwissRegulon (Pachkov et al., 2013). The last update of the database could not be determined.

A comparable repository of TF binding motifs is MotifMap (Daily et al., 2011). It combines data from JASPAR (Portales-Casamar et al., 2010), TRANSFAC (Matys et al., 2003) and scientific publications with a comparative genomic statistical approach. The database currently supports *H. sapiens* (530 motifs for hg19 (human genome 19) and 570 motifs for hg18 (human genome 18)), *M. musculus* (575 motifs), *S. cerevisiae* (147 motifs), *D. melanogaster* (66 motifs), and *C. elegans* motifs (6 motifs). MotifMap consists of three components: Motif Search, Gene Search and SNPer. Motif Search allows TFs and their binding sequence logos to be retrieved. Using Gene Search, the user is able to select a gene of interest and search for TFBSs in upstream and downstream regions from TSS. SNPer enables the identification of TF binding motifs that overlap with single nucleotide polymorphism (SNP) sites. The results obtained with Gene Search and SNPer are provided with a normalized log-odds (NLOD) score achieved from PWMs, with a Bayesian Branch Length Score (BBLS) to measure the degree of evolutionary conservation, and with a false discovery rate (FDR) score. The last update of the database could not be determined.

The Gene Transcription Regulation Database (GTRD) is a repository of human and mouse TFs and corresponding binding sites identified by ChIP-seq experiments (Yevshin et al., 2017). The authors of the database started from the raw data obtained from the ENCODE project (Landt et al., 2012), SRA (Kodama et al., 2012), GEO (Barrett et al., 2013) and previous literature; and most importantly, the data has been uniformly processed by the authors' own workflows. The GTRD allows ChIP-seq experiments to be to retrieved by TF, to find TFBSs located in regulatory regions of genes of interest, and conversely, to obtain genes that have binding sites for specific TF. All received results can be exported. The repository has a semi-automatic procedure for updating data, with an update released every six months. It was recently updated on April 2017.

2.2. Drosophila melanogaster databases

The *Drosophila* Core Promoter Database (DCPD) is a simply-constructed repository of 205 *D. melanogaster* core promoter sequences (−45: +30 bp) with TSSs (Kutach and Kadonaga, 2000). The TSSs were empirically determined by nuclease protection, primer extension or 5′ rapid amplification of cDNA ends (RACE). The promoters were extracted from the literature. The DCPD also contains TATA box, downstream promoter element (DPE), Inr data. The last update of the database could not be determined.

FlyFactorSurvey includes 613 PWMs determined by a bacterial one-hybrid (B1H) system or DNase I footprint experiments for 327 TFs in *D. melanogaster* (Zhu et al., 2011). The data were extracted from literature. The user can inquire about TFs, motif sequences and experimental conditions used to gain PWMs. It is possible to query by TF gene or by type of DNA binding domain inside the TF. The last update of the database could not be determined.

The *Drosophila* DNase I footprint database (FlyReg) is a repository of 1367 binding sites for 87 TFs (Bergman et al., 2005). The data were obtained from DNase I footprint experiments and extracted from the literature. FlyReg provides comprehensive knowledge about TFs with external links and references. It was updated recently on October 2007.

OnTheFly contains 436 PWMs (327 from FlyFactorSurvey (Zhu et al., 2011), 87 from FlyReg (Bergman et al., 2005), 22 from JASPAR (Portales-Casamar et al., 2010) and the literature) found in BH1, DNase I, and systematic evolution of ligands by exponential enrichment (SELEX) experiments for over 387 TFs (Shazman et al., 2014). The database supplies general information on TFs, PWM sequences, PWMs of orthologous TFs, and structural data for TFs and their binding sites obtained using high-throughput method based on Monte Carlo simulations (Zhou et al., 2013). The user can search by TF, by TF binding domain and by PWM to identify which TF binds to a specific DNA sequence. OnTheFly was updated on March 2016.

2.3. Saccharomyces cerevisiae databases

The *Saccharomyces cerevisiae* Promoter Database (SCPD) is one of the oldest still functioning yeast databases (Zhu and Zhang, 1999). The 1999 release of the SCPD includes promoter sequences with TSSs, 580 experimentally-mapped TFBSs, and 103 corresponding TFs. The user can search for genes regulated by specific TF and related binding sites on the basis of sequence pattern, which is very useful facility, and identify putative TFBSs in own promoter sequence. Although not every tool from the SCPD currently works, it is nevertheless a valuable database. The last update of the database could not be determined.

Yeast Search for Transcriptional Regulators And Consensus Tracking (YEASTRACT) is a literature-based tool for the analysis of genes and genomic transcription regulation in *S. cerevisiae* (Teixeira et al., 2014). The database gives the opportunity to retrieve promoter sequences (−1000: −1) and TFBSs. YEASTRACT contains DISCOVERER, a set of tools used to recognize overrepresented motifs in the promoters of co-regulated genes. In contrast to YPA described below (Chang et al., 2011), the user can identify documented and potential regulatory associations between TFs and genes in particular environmental conditions. Moreover, it is possible to analyze gene expression changes in response to overexpression, deletion or mutation of TFs. The database was recently updated in 2017.

A similar but more sophisticated database is Yeast Promoter Atlas (YPA) (Chang et al., 2011). It provides wide range of promoter features related to transcriptional regulation: promoter sequences, TSSs, TATA boxes, 5′UTRs, 3′UTRs, DNA bendability, nucleosome occupancy, TFBSs, TFs, physical or genetic interactions between two TFs which

have binding sites in the same promoter, and TF knockout gene expression data. The assembled data were derived from five experimental articles and six databases including MYBS (Tsai et al., 2007), YEASTRACT (Monteiro et al., 2008), BioGRID (Breitkreutz et al., 2008), SwissRegulon (Pachkov et al., 2007) and the SCPD (Zhu and Zhang, 1999). It is very user friendly to operate: the user has many options while searching, the data are highly integrated and the promoter features are displayed simultaneously. It is also possible to download obtained results for further manipulation. YPA was updated on 30th June 2012.

As with YPA, the Yeast Combinatorial Regulation Database (YCRD) yields regulatory associations (434,197) between cooperating TFs (2535) and target genes (6243) (Wu et al., 2016). This is the most comprehensive collection of such interactions in *S. cerevisiae*. To build the database, the creators extracted cooperating TF pairs predicted by 17 different algorithms given in the literature, and empirically confirmed target genes from YEASTRACT platform (Teixeira et al., 2014). In addition, experimental evidence of TF-gene interplays was added to confirm any possible interactions. The mentioned evidence is physical (TF binding to promoter) and genetic (TF perturbation vs. gene expression). It is possible to search for target genes of a chosen TF pair, and conversely, search for cooperating TFs that regulate a gene or genes of interest. The user can also browse the data. The YCRD is well organized and has an easy-to-use interface. It was last revised on 5th June 2016.

The high number of frequently conflicting matrices encountered for specific TF prompted researchers to establish ScerTF (Spivak and Stormo, 2012). The database provides the most accurate PWM for each TF in *S. cerevisiae*. The authors of ScerTF collected 1226 PWMs (experimental and computational) from Transfactome (Foat et al., 2008), SwissRegulon (Pachkov et al., 2007), the SCPD (Zhu and Zhang, 1999), UniPROBE (Robasky and Bulyk, 2011), and the literature for 196 varied TFs and subsequently evaluated the PWMs using results from in vivo chromatin immunoprecipitation-on-chip (ChIP-chip) and TF deletion experiments. Optimal thresholds were also identified to predict TF binding for particular PWM. It was found that the PWMs from ScerTF surpassed PWMs from JASPAR (Portales-Casamar et al., 2010). ScerTF supplies knowledge about TFs, recommended and additional matrices, genes bound to TFs in ChIP-chip experiments, genes affected by TF deletion, and predicted TF-TF regulatory interactions. The user can query the database with TF name, DNA sequencer or matrix. The last update of the database could not be determined.

2.4. Multispecies plant databases

The Plant Promoter Database (PlantProm DB) is a collection of 576 proximal RNA polymerase II promoter sequences (−200:+51 bp) with experimentally determined TSSs (primer extension, 5′RACE) from 86 plant species (Shahmuradov et al., 2003). It provides also 3032 statistically significant TFBSs and corresponding TFs, if available, from *Oryza sativa*, *Z. mays*, *Medicago truncatula*, *Glycine* max and *Vitis vinifera*. The database allows promoters to be classified as either TATA or TATA-less. The results can be downloaded. The PlantPromDB was last updated on March 2016.

Another plant database is the ppdb (Hieno et al., 2014). It comprises information on promoter sequences, TSSs, structure of core promotes (TATA box, Y Patch, Inr, GA and CA elements), and regulatory element groups (REGs) in *A. thaliana*, *O. sativa*, *Physcomitrella patens*, and *Populus trichocarpa*. Genome sequences and TSSs were extracted from databases, i.e., TAIR (Lamesch et al., 2012), RARGE (Seki et al., 2002), the RAP-DB (Sakai et al., 2013), Phytozome (Goodstein et al., 2012) and the literature. Core and regulatory elements were determined by local distribution of short sequences (LDSS) analysis (Yamamoto et al., 2007) and come from the PLACE database (Higo et al., 1999) and from literature. The last update of the database could not be determined.

Plant Promoter Analysis Navigator (PlantPAN 2.0) is a platform for the investigation of promoters and the reconstruction of transcriptional regulatory networks in 76 plant species (Chow et al., 2016). The database gives detailed information on *A. thaliana*, *O. sativa* and *Z. mays* genes, promoter sequences, tandem repeats, CpG islands, TFBSs, the structure of homologous promoters, TFs co-expressed with the query gene under various conditions (environmental stress, hormone treatment, developmental stage), and provides an opportunity to construct GRNs. It is also possible to examine own promoter sequences. Through the TF/TFBS search function the user can obtain wider knowledge about TFs, their target genes and TF binding motifs. In addition, the database allows the co-occurrence of TFs and their binding sites to be determined within promoters of input gene group and construct networks for TF-gene and protein-protein interactions (PPIs). The data incorporated into PlantPAN 2.0 were retrieved from JASPAR (Mathelier et al., 2014), TRANSFAC (Matys et al., 2006), PLACE (Higo et al., 1999) and the literature (TFs, PWMs), from Uniprot (The UniProt Consortium, 2015) (description and functions of TFs), from PAIR (Lin et al., 2011), MINT (Licata et al., 2012) and the BioGRID (Chatr-Aryamontri et al., 2015) (PPIs). Putative TFBSs were found with the Match program (Kel et al., 2003). PlantPAN 2.0 is a comprehensive database with an intuitive and modern interface. It was recently revised on July 2015.

Profound investigation of stress response mechanisms in plants is possible with the Stress-Responsive Transcription Factor Database (STIFDB2) (Naika et al., 2013). It is a repository of 38,798 associations between biotic or abiotic stress signals, stress-responsive genes, and stress-responsive TFs in *A. thaliana*, *O. sativa* subsp. *japonica*, and *O. sativa* subsp. *indica*. Genes, their orthologs, TFBSs, and TFs are very well described here. The data assembled in the STIFDB2 were derived from the literature (TFs), from GEO (Barrett et al., 2011), TAIR (Lamesch et al., 2012), Rice Genome Annotation Project (Ouyang et al., 2007), the RAP-DB (Ohyanagi et al., 2006) (genes, promoter and 5′UTR sequences), and from the STIF algorithm (Shameer et al., 2009) (potential TFBSs). The TFBSs are provided with a z-score to show the strength of prediction. The database gives a range of possibilities to search and browse and has an easy-to-use interface. It was updated on 15th October 2012.

2.5. Arabidopsis thaliana databases

Arabidopsis Gene Regulatory Information Server (AGRIS) contains an essential knowledge for examining GRNs in the model plant *A. thaliana* (Yilmaz et al., 2011). AGRIS consists of three parts: AtcisDB (*Arabidopsis thaliana cis*-regulatory element database), AtTFDB (*Arabidopsis thaliana* transcription factor database), and AtRegNet (*Arabidopsis thaliana* regulatory network). AtcisDB contains 33,239 promoter sequences (from TAIR (Swarbreck et al., 2008)) and experimentally validated or predicted TFBSs (from the literature). AtTFDB includes 1773 TFs grouped into 50 families. Genes encoding the TFs are characterized and information on their contribution to regulatory subnetwork is provided. The TFs were identified by a combination of the Basic Local Alignment Search Tool (BLAST) (McGinnis and Madden, 2004) and motif searches based on the literature with identified TFs or on motifs conserved between TFs from a family. AtRegNet is a tool employed to display the regulatory networks. Utilising the Regulatory Networks Interaction Module (ReIN) it documents and visualizes GRNs created by 64 TFs and 8070 direct target genes. Information given in AtRegNet was taken from published high-throughput in vivo experimental approaches such as ChIP-chip or ChIP-seq. AGRIS was recently updated on 31st October 2016.

Another plant database aimed at a thorough reconstruction of GRNs is *Arabidopsis thaliana* Promoter Analysis Net (AtPAN) (Chen et al., 2012). AtPAN is similar in content and structure to the PlantPAN 2.0 database also described in this review (Chow et al., 2016). The major difference is that AtPAN does not provide information on tandem repeats and CpG islands in promoters. The data incorporated into AtPAN are obtained from other databases, i.e. TAIR (Swarbreck et al., 2008),

the ppdb (Yamamoto and Obokata, 2008), NCBI Homologene (Maglott et al., 2011), TRANSFAC (Matys et al., 2006), JASPAR (Bryne et al., 2008), AGRIS (Yilmaz et al., 2011), ATTED-II (Obayashi et al., 2011), PAIR (Lin et al., 2011), and from the literature. TFBSs were identified by scanning promoter sequences with the Mach tool (Matys et al., 2006) and a homemade Perl program. GRNs were created using the AutoPat method (Tang et al., 2011). The Apriori algorithm was used to discover the co-occurrence of TF/TFBS. The advantage of AtPAN is its intuitive interface. The last update of the database could not be determined.

AthaMap reveals transcriptional and post-transcriptional regulation of gene expression in *A. thaliana* by providing TFBSs and small RNA (sRNA) target genes (Bulow et al., 2012; Hehl et al., 2016). The repository contains 4.9×10^7 PWM-predicted TFBSs for 126 TFs representing 29 families and 41,965 putative miRNA target sites in 10,442 genes. The TFBS were identified with the Patser program (Hertz and Stormo, 1999), and miRNA targets with the psRNATarget web server (Dai and Zhao, 2011). AthaMap incorporates several web tools to address specific questions. The Search tool shows potential TFBSs and sRNA target sites in the selected genomic region (500 bp upstream and downstream) with information on highly-conserved sites. The binding sites are provided with well-characterized corresponding TFs or sRNAs. The Gene Analysis tool enables identification of common binding sites for TFs or sRNAs in the set of chosen genes. The results can be visualized graphically. Using the Gene Identification tool, the user is able to identify predicted binding sites targeted by selected TFs within a 6000 bp region (-2000:$+4000$ bp). Co-localizing TFBSs of two TFs in proximity can be found with the Colocalization tool. Small RNA Targets and MicroRNA Targets tools are used to predict target genes for sRNAs. AthaMap is a very well-organized database. The user can easily download or print all the results. It was updated on 3th July 2017.

A unique database is AtmiRNET (Chien et al., 2015). It aims to extend the knowledge about miRNAs, their transcriptional regulation and targets in *A. thaliana*. AtmiRNET is a repository of miRNA promoters (sequences, TSSs, Y Patch, Inr, REGs, CA and GA elements, TFBSs), corresponding TFs, direct and indirect targets of miRNAs, and miRNA function. Furthermore, by combining TF-miRNA and miRNA-gene interactions with indirect targets, AtmiRNET allows the reconstruction of regulatory networks. The promoter sequences given with the TSSs were obtained experimentally or by using an SVM-based model (Chien et al., 2011). The core promoter elements were obtained from the ppdb database (Yamamoto and Obokata, 2008). TFs and their binding sites were predicted by performing a combination of co-expression analysis with hypergeometric *p*-values. psRNATarget server (Dai and Zhao, 2011) and miRTarBase (Hsu et al., 2014) are sources of predicted and empirically determined miRNA-target interactions. Indirect targets of miRNAs were found using the DFS algorithm. AtmiRNET has a modern and user-friendly interface. The last update was on 3th March 2017.

The information on selected eukaryotic databases is summarized in Table 1.

3. Prokaryotic databases

3.1. Multispecies prokaryotic databases

The Prokaryotic Database of Gene Regulation (PRODORIC) is a platform that concerns transcriptional regulation, signal transduction, protein interactions, and biochemical pathways in 696 bacterial species with a special focus on pathogenic organisms (Grote et al., 2009). In the field of transcriptional regulation, the database contains information on 1586 promoter sequences, 197 PWMs with sequence logos, 2921 TFBSs, regulated genes, operons, and TFs. Several tools are incorporated into PRODORIC. The GBpro Genome Browser allows promoters, TFBSs, and genes to be displayed in parallel as a sequence and graphical map. A prediction tool called Virtual Footprint scans sequences or whole genomes for new TF targets (Munch et al., 2005). Evolutionary conservation of the found targets can be subsequently analyzed with the SMILE tool. PRODORIC is based only on the original scientific literature and is not complemented with in silico predictions. It was last updated on 10th September 2008.

Another prokaryotic repository is the σ^{54} promoter database (Pro54DB) (Liang et al., 2016). It is a unique resource of σ^{54} promoters that control expression of carbon and nitrogen-related genes and respond to environmental changes. The Pro54DB includes 210 experimentally verified σ^{54} promoter sequences with TSSs, 297 regulated genes in 43 prokaryotic species. The incorporated data were manually extracted from 133 scientific publications. Moreover, the database contains the iPro54-PseKNC tool for identifying σ^{54} promoters by pseudo k-tuple nucleotide composition (Lin et al., 2014). The user can search, browse, and even submit his own data. The Pro54DB was last revised on 31st August 2016.

CollecTF represents a collection of 9750 experimentally-validated binding sites for over 240 TFs in over 100 bacterial species (Kilic et al., 2016). After selecting the TF family, bacterial taxonomic unit, and experimental technique, the user obtains information on TFBSs (binding motif, sequence, genomic location), TFs, regulated genes (with type of regulation), experimental techniques, and the supporting literature. It is also possible to browse the data and download them. Moreover, CollecTF allows pairwise comparisons of TFBS and motif similarities across species and taxonomical groups using Levenshtein distance, Kullback-Leibler divergence, and the Pearson correlation coefficient. The database also serves as a submission portal. CollecTF is a well-organized and very intuitive database. The last update of the database could not be determined.

3.2. Bacillus subtilis database

The Database of Transcriptional Regulation in *Bacillus subtilis* (DBTBS) consists of 120 TFs (with type of regulation) classified using Pfam motifs, corresponding TFBSs, regulated genes, and operons (Sierro et al., 2008). All the data were retrieved from 947 experimental publications. The DBTBS provides also a phylogenetic conservation study of both TFs and TFBSs (Makita et al., 2004). Furthermore, the user can obtain information on overrepresented hexameric motifs within the upstream intergenic region of homologous genes from 40 Gram-positive bacterial species. It is possible to search and browse the data. The DBTBS was updated on 5th February 2015.

3.3. Escherichia coli databases

PromEC is a simple text files-based platform of 472 promoter sequences (-75:$+25$ bp) and TSSs in *Escherichia coli*, the best studied bacterial model organism (Hershberg et al., 2001). The promoter sequences were extracted from the RegulonDB (Salgado et al., 2000), complete genome sequence (Blattner et al., 1997), and the literature. The TSSs were determined experimentally by S1 nuclease mapping or by primer extension. PromEC was updated on July 2000.

The RegulonDB is a database concerning regulation of transcription initiation and regulatory network in *E. coli* K-12 (Gama-Castro et al., 2016). It includes thorough knowledge about 8597 promoter sequences, 93 PWMs, 2285 TFBSs, 4652 genes, 280 terminators, 208 TFs (with type of regulation), more complex genome units, i.e., 3547 transcription units, 2632 operons, 53 gensor units, 480 regulons and 3239 interactions (TF-gene, TF-transcription unit, TF-operon, TF-TF, sRNA-gene). The incorporated data were retrieved from experimental evidences and are based on computational predictions. In addition, the RegulonDB offers a range of tools: DrawingTracesTool images elements involved in gene regulation, while the Textpresso tool enables the user to browse the literature curated by the repository. The user can search or browse, download all the results, and submit data to the database to make it publicly available. The RegulonDB has been continuously updated since 1997, with the most recent update on 8th May 2017.

Table 1
Selected eukaryotic databases.

Resource	Organism	Contents/Features	URL	References
EPD	eukaryotes	4806 promoter sequences (−499: +100 bp); TSSs	http://epd.vital-it.ch/EPD_database.php	(Cavin Perier et al., 1998)
EPDnew	*H. sapiens*, *M. musculus*, *D. melanogaster*, *A. mellifera*, *C. elegans*, *D. rerio*, *A. thaliana*, *Z. mays*, *S. cerevisiae*, *S. pombe*	> 105,000 promoter sequences; TATA box; GC box; CCAAT box; Inr; CpG island; selection and analysis tools	http://epd.vital-it.ch/EPDnew_database.php	(Dreos et al., 2013)
ECRbase	*H. sapiens*, *M. mulatta*, *P. troglodytes*, *B. taurus*, *C. familiaris*, *M. domestica*, *R. norvegicus*, *M. musculus*, *G. gallus*, *X. tropicalis*, *D. rerio*, *T. rubripes*	promoter sequences; TFBSs	https://ecrbase.dcode.org/	(Loots and Ovcharenko, 2007)
TRED	*H. sapiens*, *M. musculus*, *R. norvegicus*	139,379 promoter sequences; TSSs; TFBSs; TFs; TF-gene interactions; orthologous genes	http://rulai.cshl.edu/TRED	(Zhao et al., 2005; Jiang et al., 2007)
DIANA miRGen (version 3.0)	*H. sapiens*, *M. musculus*	miRNA genes (276 TSSs, over 19,000 TF binding site logos); information on miRNAs; 202 TFs	http://www.microrna.gr/mirgen	(Georgakilas et al., 2016)
CircuitsDB	*H. sapiens*, *M. musculus*	promoter sequences of miRNAs and protein-coding genes with TFBSs; sites for miRNAs in 3´UTR; interactions (TF-promoter, TF-miRNA, miRNA-3´UTR); information on TFs, target genes, and miRNAs; tissue expression of FFL components	http://biocluster.di.unito.it/circuits/	(Friard et al., 2010)
RegNetwork	*H. sapiens*, *M. musculus*	interactions (TF-gene, TF-miRNA, TF-TF, miRNA-TF, miRNA-gene)	http://www.regnetworkweb.org	(Liu et al., 2015)
microPIR2	*H. sapiens*, *M. musculus*	80 million human, 40 million mouse miRNA binding sites in promoters; conserved miRNA binding sites; information on miRNAs and target genes	http://www4a.biotec.or.th/micropir2	(Piriyapongsa et al., 2014)
miRWalk2.0	*H. sapiens*, *M. musculus*, *R. norvegicus*	miRNA binding sites in promoters; information on miRNAs and target genes	http://mirwalk.uni-hd.de/	(Dweep et al., 2011)
DoOP, DoOPSearch	chordates, plants	promoter sequences; evolutionary conserved promoter sequences (putative TFBSs); possible function of motifs and genes	http://doop.abc.hu/ http://doopsearch.abc.hu/	(Barta et al., 2005; Sebestyen et al., 2009)
JASPAR (version 2016)	vertebrates, plants, insects, fungi, nematodes, urochordata	1082 TF binding profiles; information on TFs; tools for predicting TFBSs within DNA sequences and for analysis and manipulation of TFBSs	http://jaspar.genereg.net/	(Mathelier et al., 2016)
HOCOMOCO (version 10)	*H. sapiens*, *M. musculus*	1135 PWMs; information on TFs	http://hocomoco.autosome.ru/	(Kulakovskiy et al., 2016)
MotifMap	*H. sapiens*, *M. musculus*, *S. cerevisiae*, *D. melanogaster*, *C. elegans*	binding motifs for TFs; binding motifs overlaping with SNP sites	http://motifmap.ics.uci.edu/#Home	(Daily et al., 2011)
GTRD	*H. sapiens*, *M. musculus*	TFBSs for 476 human and 257 mouse TFs	http://gtrd.biouml.org/	(Yevshin et al., 2017)
DCPD	*D. melanogaster*	205 core promoter sequences; TSSs; TATA box; DPE; Inr	http://labs.biology.ucsd.edu/Kadonaga/DCPD.htm	(Kutach and Kadonaga, 2000)
FlyFactorSurvey	*D. melanogaster*	613 PWMs with sequences for 327 TFs; information on TFs	http://mccb.umassmed.edu/ffs/	(Zhu et al., 2011)
Drosophila DNase I footprint database (version 2.0)	*D. melanogaster*	1367 binding sites for 87 TFs; information on TFs	http://www.flyreg.org/	(Bergman et al., 2005)
OnTheFly	*D. melanogaster*	436 PWMs with sequences for over 387 TFs; PWMs of orthologous TFs; structural data for TFs and their binding sites; information on TFs	https://bhapp.c2b2.columbia.edu/OnTheFly/index.php	(Shazman et al., 2014)
SCPD	*S. cerevisiae*	promoter sequences; TSSs; 580 TFBSs; 103 TFs	http://rulai.cshl.edu/SCPD/	(Zhu and Zhang, 1999)
YEASTRACT	*S. cerevisiae*	promoter sequences; 326 TFBSs; tools to determine overrepresented motifs in promoters of co-regulated genes; 206,299 TF-gene interactions in various conditions; gene expression changes in TF perturbation	http://www.yeastract.com/	(Teixeira et al., 2014)

(*continued on next page*)

Table 1 (*continued*)

Resource	Organism	Contents/Features	URL	References
YPA (version 1.6)	*S. cerevisiae*	promoter sequences; TSSs; TATA boxes; 5′UTRs; 3′UTRs; DNA bendability; nucleosome occupancy; TFBSs; interactions (28,826 TF-gene, 100 TF-TF-gene); 21,843 TF knockout gene expression	http://ypa.csbb.ntu.edu.tw/index.html	(Chang et al., 2011)
YCRD	*S. cerevisiae*	434,197 TF-TF-gene interactions	http://cosbi.ee.ncku.edu.tw/YCRD/	(Wu et al., 2016)
ScerTF	*S. cerevisiae*	the most accurate and additional PWMs for 196 TFs; genes bound to TFs; genes affected by TF deletion; TF-TF interactions; information on TFs	http://stormo.wustl.edu/ScerTF/	(Spivak and Stormo, 2012)
PlantProm DB (version 2016.03)	plants (86 species)	576 promoter sequences; TSSs; promoters TATA or TATA-less; 3032 TFBSs; TFs	http://www.softberry.com/berry.phtml?topic=plantp_2016.03&subgroup=plantprom&group=data&no_menu=on	(Shahmuradov et al., 2003)
ppdb (version 3.0)	*A. thaliana*, *O. sativa*, *P. patens*, *P. trichocarpa*	promoter sequences; TSSs; TATA box; Y Patch; Inr; GA and CA elements; REGs	http://ppdb.agr.gifu-u.ac.jp/ppdb/cgi-bin/index.cgi	(Hieno et al., 2014)
PlantPAN 2.0	plants (76 species), mainly: *A. thaliana*, *O. sativa*, *Z. mays*	promoter sequences; tandem repeats; CpG islands; 1143 TFBSs; structure of homologous promoters; 16,960 TFs; co-occurrence of TFs and TFBSs within promoters of gene group; GRNs	http://plantpan2.itps.ncku.edu.tw/	(Chow et al., 2016)
STIFDB2	*A. thaliana*, *O. sativa* subsp. *japonica*, *O. sativa* subsp. *indica*	38,798 associations between stress signals, stress-responsive genes, and stress-responsive TFs; information on genes, their orthologs, TFBSs, TFs	http://caps.ncbs.res.in/stifdb2	(Naika et al., 2013)
AGRIS	*A. thaliana*	33,239 promoter sequences; TFBSs; 1773 TFs; TF-gene interactions	http://agris-knowledgebase.org/	(Yilmaz et al., 2011)
AtPAN	*A. thaliana*	promoter sequences; TSSs; TFBSs; structure of homologous promoters; TFs; co-occurrence of TFs and TFBSs within promoters of gene group; GRNs	http://atpan.itps.ncku.edu.tw/	(Chen et al., 2012)
AthaMap (version 8.0)	*A. thaliana*	4.9×10^7 TFBSs for 126 TFs; 41,965 miRNA target sites in 10,442 genes; information on TFs and sRNAs; web tools	http://www.athamap.de	(Bulow et al., 2012; Hehl et al., 2016)
AtmiRNET	*A. thaliana*	187 miRNA promoter sequences; TSSs; Y Patch; Inr; REGs; CA and GA elements; TFBSs; TFs; targets of miRNAs; regulatory networks; miRNAs functions	http://AtmiRNET.itps.ncku.edu.tw/	(Chien et al., 2015)

Another *E. coli* K-12-centered database is EcoCyc (Keseler et al., 2016). It is a comprehensive source of literature-based experimental and in silico predicted information for the entire genome, transporters, metabolism and transcriptional regulation. In the field of transcriptional regulation, EcoCyc complements the data provided by the RegulonDB (Gama-Castro et al., 2016). It contains information on 3841 promoters, 2836 TFBSs, 4505 genes, 283 terminators, 205 TFs, 3553 transcription units, operons and regulons. In contrast to the RegulonDB, it is not possible to download the data, but the SmartTable tool allows the user to browse the results and save it. Access to the repository is free, but after viewing > 30 pages the user is required to register. EcoCyc is well maintained and often updated, the last update being on 17th December 2016.

The information on selected prokaryotic databases is summarized in Table 2.

4. Common eukaryotic and prokaryotic databases

The footprintDB is a database that collects the knowledge about 9087 DNA binding motifs and 23,174 DNA binding sites for 7196 TFs in prokaryotes and eukaryotes (Sebastian and Contreras-Moreira, 2014). It allows the user to submit DNA sites and motifs for comparison with those from the database and to predict binding TFs, and conversely, predict potential DNA sites and motifs based on the protein sequence of a possible TF. However, the predictive power is higher for eukaryotes than for prokaryotes. Moreover, registered users have access to additional facilities, i.e., to preserve searches and insert new databases. The footprintDB integrates data from the literature and several open access repositories, i.e., *Arabidopsis* cistrome (O'Malley et al., 2016), *Arabidopsis* TFs (Franco-Zorrilla et al., 2014), *Arabidopsis* MYB TFs (Kelemen et al., 2015), human TFs (Jolma et al., 2013), human TFs2 (Jolma et al., 2015), *Drosophila* TFs (Down et al., 2007), *Drosophila* zinc fingers (Enuameh et al., 2013), Cis-BP (Weirauch et al., 2014), JASPAR (Mathelier et al., 2016), 3D-footprint (Contreras-Moreira, 2010), HOCOMOCO (Kulakovskiy et al., 2013), AthaMap (Bulow et al., 2009), the DBTBS (Sierro et al., 2008), the RegulonDB (Salgado et al., 2013), UniPROBE (Robasky and Bulyk, 2011). According to the authors, the footprintDB has a similar coverage of multicellular organisms as commercial TRANSFAC. The database gives the opportunity to download all obtained results. It is constructed in plain fashion and is easy to use. The most recent update was on 20th June 2017.

The second database that provides information on DNA binding specificities of TFs in prokaryotes and eukaryotes is Universal PBM Resource for Oligonucleotide Binding Evaluation (UniPROBE) (Hume et al., 2015). It currently hosts PWM logos and binding sequences for 574 TFs. It includes information for species not supported by the footprintDB (Sebastian and Contreras-Moreira, 2014). The data incorporated into UniPROBE were extracted from the literature (PBM experiments). Furthermore, the database offers tools to search for potential binding sites along nucleotide sequence, to find a similar motif to that input by the user, and to generate negative control sequences that will not bind to the selected TFs. The user can search or browse the database, download the results, and add his own findings. UniPROBE has an intuitive interface and is frequently updated. The last update was on 1st June 2016.

The information on selected eukaryotic and prokaryotic databases is summarized in Table 3.

5. Evaluation of databases

To demonstrate how the databases perform when subjected to the same analysis, the most frequently given parameter was chosen, i.e. TF with binding motif. The number of TF–binding motif pairs was estimated with eleven repositories where the data were available (the

Table 2
Selected prokaryotic databases.

Resource	Organism	Contents/features	URL	References
PRODORIC (version 8.9)	prokaryotes (696 species)	1586 promoter sequences; 197 PWMs; 2921 TFBSs; regulated genes; operons; TFs; tools that scans sequences or whole genomes for TF targets and checks its evolutionary conservation	http://www.prodoric.de	(Grote et al., 2009)
Pro54DB (version 1.0)	prokaryotes (43 species)	210 σ^{54} promoter sequences; TSSs; 297 regulated genes; tool that identifies σ54 promoters	http://lin.uestc.edu.cn/database/pro54db	(Liang et al., 2016)
CollecTF	prokaryotes (over 100 species)	9750 TFBSs for over 240 TFs; regulated genes; pairwise TFBSs and motifs comparisons; submission portal	http://www.collectf.org	(Kilic et al., 2016)
DBTBS (version 5.0)	*B. subtilis*	120 TFs; TFBSs; regulated genes; 736 operons; phylogenetic conservation of TFs and TFBSs; overrepresented hexameric motifs within upstream region of homologous genes	http://dbtbs.hgc.jp	(Makita et al., 2004; Sierro et al., 2008)
PromEC	*E.coli*	472 promoter sequences (−75: +25 bp); TSSs	http://margalit.huji.ac.il/promec	(Hershberg et al., 2001)
Regulon DB (version 9.2)	*E. coli* K-12	8597 promoter sequences; 93 PWMs, 2285 TFBSs; 4652 genes; 280 terminators; 208 TFs; 3547 transcription units; 2632 operons; 480 regulons; 53 gensor units; 3239 interactions (TF-gene, TF- transcription unit, TF-operon, TF-TF, sRNA-gene); submission portal	http://regulondb.ccg.unam.mx	(Gama-Castro et al., 2016)
EcoCyc (version 20.5)	*E. coli* K-12 substr. MG1655	genome; transporters; metabolism; transcriptional regulation (3841 promoters, 2836 TFBSs, 4505 genes, 283 terminators, 205 TFs, 3553 transcription units, operons, regulons)	https://ecocyc.org/	(Keseler et al., 2016)

footprintDB, JASPAR, UniPROBE, MotifMap, HOCOMOCO, FlyFactorSurvey, YEASTRACT, ScerTF, the PRODORIC, CollecTF, the Regulon DB) for nine main species (*H. sapiens*, *M. musculus*, *R. norvegicus*, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, *A. thaliana*, *Z. mays*, *E. coli*). As a consequence, thirty-five results were obtained (Table 4). The greatest number of TF – binding motif pairs was collected by YEASTRACT (759 for *S. cerevisiae*), followed by MotifMap (724 for *M. musculus*, 707 for *H. sapiens*), FlyFactorSurvey (326 for *D. melanogaster*), and the footprintDB (> 200 for *C. elegans*, > 200 for *A. thaliana*, 182 for *E. coli*, 55 for *R. norvegicus*, 11 for *Z. mays*) (Table 4). It is worth underlining that databases with a lower score included unique TF–binding motif pairs in contrast to the leading ones (data not shown). To sum up, conducted analysis did not indicate one comprehensive database which performed perfectly with the given query. These results demonstrate need to employ a few complementary resources in research.

6. Summary and conclusions

The forty databases described above are certainly valuable tools for the researchers to investigate transcription-level regulation of gene expression in various species of eukaryotes and prokaryotes. The repositories provide information on promoter sequences, promoter patterns essential for transcription (TATA box, GC box, CCAAT box, GA and CA elements, Inr, CpG island, DPE, Y Patch, REGs), binding sites for TFs and miRNAs and corresponding TFs and miRNAs, interactions that allow GRNs to be established (TF-gene, miRNA-gene, TF-TF, TF-miRNA), promoter DNA bendability and nucleosome occupancy (DNA accessibility), as well as TSSs. Moreover, the databases give the possibility to examine evolutionarily-conserved TFs, TFBSs, miRNA binding sites in promoters, and statistically overrepresented motifs in promoters of co-regulated genes, thus allowing common mechanisms of regulation of genes transcription to be determined. They also contain information on the influence of external factors (biotic and abiotic environmental stress, hormone treatment) on transcription in plants. The presented findings concern many species, including several model organisms, e.g., *E. coli*, *B. subtilis*, *S. cerevisiae*, *A. thaliana*, *Z. mays*, *O. sativa*, *P. trichocarpa*, *C. elegans*, *D. melanogaster*, *D. rerio* and *M. musculus*. Since the functional analysis of promoters is laborious and expensive, researchers tend to extend at least part of the information gained on model organisms to less-studied organisms, with the degree of the extension associated with the distance between the organisms (Sierro et al., 2008). In this regard, the described platforms offer great potential. The data incorporated into the repositories were derived from both experimental literature and in silico predictions. The repositories are highly diverse in their approach: i.e. they can be simple or more complex, regularly updated or not, and they offer a range of search options. Moreover, the performed analysis indicated the necessity to use multiple resources in research. The databases are clearly powerful tools for the researchers, and there is a continuous need to expand existing repositories and establish new ones.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

M.M. developed the original idea of the work and wrote the manuscript. H.W., Ł.K. and P.S. revised the manuscript. All authors have read and approved the final version of the manuscript.

Funding

Not applicable.

Acknowledgements

Not applicable.

Table 3
Selected eukaryotic and prokaryotic databases.

Resource	Organism	Contents/features	URL	References
footprintDB	eukaryotes and prokaryotes	9087 DNA binding motifs and 23,174 TFBS for 7196 TFs; tool that predicts binding TFs and DNA sites, motifs	http://floresta.eead.csic.es/footprintdb/	(Sebastian and Contreras-Moreira, 2014)
UniPROBE	eukaryotes and prokaryotes	PWM logos and binding sequences for 574 TFs; tools that predict binding sites for TFs, find similar motif, generate negative control sequence	http://uniprobe.org	(Hume et al., 2015)

Table 4
Comparison of databases.

Resource	*H. sapiens*	*M. musculus*	*R. norvegicus*	*D. melanogaster*	*C. elegans*	*S. cerevisiae*	*A. thaliana*	*Z. mays*	*E.coli*
	No. of TF-binding motif pairs								
footprintDB	> 200	> 200	55	> 200	> 200	> 200	> 200	11	182
JASPAR	386	128	11	133	26	175	187	6	–
UniPROBE	55	358	–	16	–	119	18	–	
MotifMap	707	724	–	248	8	507	–	–	–
HOCOMOCO	274	262	–	–	–	–	–	–	–
FlyFactorSurvey	–	–	–	326	–	–	–	–	–
YEASTRACT	–	–	–	–	–	759	–	–	–
ScerTF	–	–	–	–	–	197	–	–	–
PRODORIC	–	–	–	–	–	–	–	–	87
CollecTF	–	–	–	–	–	–	–	–	25
Regulon DB	–	–	–	–	–	–	–	–	93

Note. "–" no data available; " > 200" > 200 results were incorporated into the database but it was not possible to retrieve them.

References

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Muertter, R.N., Holko, M., Ayanbule, O., Yefanov, A., Soboleva, A., 2011. NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res. 39, D1005–D1010.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41, D991–D995.

Barta, E., Sebestyen, E., Palfy, T.B., Toth, G., Ortutay, C.P., Patthy, L., 2005. DoOP: databases of Orthologous promoters, collections of clusters of orthologous upstream sequences from chordates and plants. Nucleic Acids Res. 33, D86–D90.

Bartel, D.P., 2004. microRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.

Bergman, C.M., Carlson, J.W., Celniker, S.E., 2005. *Drosophila* DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*. Bioinformatics 21, 1747–1749.

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1462.

Breitkreutz, B.J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D.H., Bahler, J., Wood, V., Dolinski, K., Tyers, M., 2008. The BioGRID interaction database: 2008 update. Nucleic Acids Res. 36, D637–D640.

Browning, D.F., Busby, S.J., 2004. The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. 2, 57–65.

Bryne, J.C., Valen, E., Tang, M.H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., Sandelin, A., 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. 36, D102–D106.

Bulow, L., Engelmann, S., Schindler, M., Hehl, R., 2009. AthaMap, integrating transcriptional and post-transcriptional data. Nucleic Acids Res. 37, D983–986.

Bulow, L., Bolivar, J.C., Ruhe, J., Brill, Y., Hehl, R., 2012. microRNA Targets', a new AthaMap web-tool for genome-wide identification of miRNA targets in *Arabidopsis thaliana*. Bio. Data Min. 5, 7.

Cavin Perier, R., Junier, T., Bucher, P., 1998. The eukaryotic promoter database EPD. Nucleic Acids Res. 26, 353–357.

Chang, D.T., Huang, C.Y., Wu, C.Y., Wu, W.S., 2011. YPA: an integrated repository of promoter features in *Saccharomyces cerevisiae*. Nucleic Acids Res. 39, D647–D652.

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K., Tyers, M., 2015. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 43, D470–D478.

Chen, Y.A., Wen, Y.C., Chang, W.C., 2012. AtPAN: an integrated system for reconstructing transcriptional regulatory networks in *Arabidopsis thaliana*. BMC Genomics 13, 85.

Chien, C.H., Sun, Y.M., Chang, W.C., Chiang-Hsieh, P.Y., Lee, T.Y., Tsai, W.C., Horng, J.T., Tsou, A.P., Huang, H.D., 2011. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. Nucleic Acids Res. 39, 9345–9356.

Chien, C.H., Chiang-Hsieh, Y.F., Chen, Y.A., Chow, C.N., Wu, N.Y., Hou, P.F., Chang, W.C., 2015. AtmiRNET: a web-based resource for reconstructing regulatory networks of *Arabidopsis* microRNAs. Database (Oxford) 2015 (bav042).

Chow, C.N., Zheng, H.Q., Wu, N.Y., Chien, C.H., Huang, H.D., Lee, T.Y., Chiang-Hsieh, Y.F., Hou, P.F., Yang, T.Y., Chang, W.C., 2016. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res. 44, D1154–D1160.

Contreras-Moreira, B., 2010. 3D-footprint: a database for the structural analysis of protein-DNA complexes. Nucleic Acids Res. 38, D91–D97.

Dai, X., Zhao, P.X., 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 39, W155–W159.

Daily, K., Patel, V.R., Rigor, P., Xie, X., Baldi, P., 2011. MotifMap: integrative genome-wide maps of regulatory motif sites for model species. BMC Bioinformatics 12, 495.

Down, T.A., Bergman, C.M., Su, J., Hubbard, T.J., 2007. Large-scale discovery of promoter motifs in *Drosophila melanogaster*. PLoS Comput. Bio. l 3, e7.

Dreos, R., Ambrosini, G., Cavin Perier, R., Bucher, P., 2013. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. Nucleic Acids Res. 41, D157–D164.

Dweep, H., Sticht, C., Pandey, P., Gretz, N., 2011. miRWalk—database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J. Biomed. Inform. 44, 839–847.

Enuameh, M.S., Asriyan, Y., Richards, A., Christensen, R.G., Hall, V.L., Kazemian, M., Zhu, C., Pham, H., Cheng, Q., Blatti, C., Brasefield, J.A., Basciotta, M.D., Ou, J., McNulty, J.C., Zhu, L.J., Celniker, S.E., Sinha, S., Stormo, G.D., Brodsky, M.H., Wolfe, S.A., 2013. Global analysis of *Drosophila* Cys(2)-His(2) zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. Genome Res. 23, 928–940.

Foat, B.C., Tepper, R.G., Bussemaker, H.J., 2008. TransfactomeDB: a resource for exploring the nucleotide sequence specificity and condition-specific regulatory activity of *trans*-acting factors. Nucleic Acids Res. 36, D125–D131.

Franco-Zorrilla, J.M., Lopez-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., Solano, R., 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc. Natl. Acad. Sci. U. S. A. 111, 2367–2372.

Friard, O., Re, A., Taverna, D., De Bortoli, M., Cora, D., 2010. CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. BMC Bioinformatics 11, 435.

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muniz-Rascado, L., Garcia-Sotelo, J.S., Alquicira-Hernandez, K., Martinez-Flores, I., Pannier, L., Castro-Mondragon, J.A., Medina-Rivera, A., Solano-Lira, H., Bonavides-Martinez, C., Perez-Rueda, E., Alquicira-Hernandez, S., Porron-Sotelo, L., Lopez-Fuentes, A., Hernandez-Koutoucheva, A., Del Moral-Chavez, V., Rinaldi, F., Collado-Vides, J., 2016. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res. 44, D133–D143.

Georgakilas, G., Vlachos, I.S., Paraskevopoulou, M.D., Yang, P., Zhang, Y., Economides, A.N., Hatzigeorgiou, A.G., 2014. microTSS: accurate microRNA transcription start site identification reveals a significant number of divergent pri-miRNAs. Nat. Commun. 5, 5700.

Georgakilas, G., Vlachos, I.S., Zagganas, K., Vergoulis, T., Paraskevopoulou, M.D., Kanellos, I., Tsanakas, P., Dellis, D., Fevgas, A., Dalamagas, T., Hatzigeorgiou, A.G., 2016. DIANA-miRGen v3.0: accurate characterization of microRNA promoters and their regulators. Nucleic Acids Res. 44, D190–D195.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40, D1178–D1186.

Grote, A., Klein, J., Retter, I., Haddad, I., Behling, S., Bunk, B., Biegler, I., Yarmolinetz, S., Jahn, D., Munch, R., 2009. PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes. Nucleic Acids Res. 37, D61–D65.

Hehl, R., Norval, L., Romanov, A., Bulow, L., 2016. Boosting AthaMap database content with data from protein binding microarrays. Plant Cell Physiol. 57, e4.

Hershberg, R., Bejerano, G., Santos-Zavaleta, A., Margalit, H., 2001. PromEC: An updated database of *Escherichia coli* mRNA promoters with experimentally identified transcriptional start sites. Nucleic Acids Res. 29, 277.

Hertz, G.Z., Stormo, G.D., 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics 15, 563–577.

Hieno, A., Naznin, H.A., Hyakumachi, M., Sakurai, T., Tokizawa, M., Koyama, H., Sato, N., Nishiyama, T., Hasebe, M., Zimmer, A.D., Lang, D., Reski, R., Rensing, S.A., Obokata, J., Yamamoto, Y.Y., 2014. ppdb: plant promoter database version 3.0. Nucleic Acids Res. 42, D1188–D1192.

Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. 27, 297–300.

Hsu, S.D., Tseng, Y.T., Shrestha, S., Lin, Y.L., Khaleel, A., Chou, C.H., Chu, C.F., Huang, H.Y., Lin, C.M., Ho, S.Y., Jian, T.Y., Lin, F.M., Chang, T.H., Weng, S.L., Liao, K.W., Liao, I.E., Liu, C.C., Huang, H.D., 2014. miRTarBase update 2014: an information

resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 42, D78–D85.
Huang, V., Place, R.F., Portnoy, V., Wang, J., Qi, Z., Jia, Z., Yu, A., Shuman, M., Yu, J., Li, L.C., 2012. Upregulation of Cyclin B1 by miRNA and its implications in cancer. Nucleic Acids Res. 40, 1695–1707.
Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., Bulyk, M.L., 2015. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. 43, D117–D122.
Jiang, C., Xuan, Z., Zhao, F., Zhang, M.Q., 2007. TRED: a transcriptional regulatory element database, new entries and other development. Nucleic Acids Res. 35, D137–D140.
Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T., Taipale, J., 2013. DNA-binding specificities of human transcription factors. Cell 152, 327–339.
Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., Taipale, J., 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature 527, 384–388.
Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V., Wingender, E., 2003. MATCH: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. 31, 3576–3579.
Kelemen, Z., Sebastian, A., Xu, W., Grain, D., Salsac, F., Avon, A., Berger, N., Tran, J., Dubreucq, B., Lurin, C., Lepiniec, L., Contreras-Moreira, B., Dubos, C., 2015. Analysis of the DNA-binding activities of the *Arabidopsis* R2R3-MYB transcription factor family by one-hybrid experiments in yeast. PLoS One 10, e0141044.
Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velazquez-Ramirez, D.A., Weaver, D., Collado-Vides, J., Paulsen, I., Karp, P.D., 2016. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res. 45, D543–D555.
Kilic, S., Sagitova, D.M., Wolfish, S., Bely, B., Courtot, M., Ciufo, S., Tatusova, T., O'Donovan, C., Chibucos, M.C., Martin, M.J., Erill, I., 2016. From data repositories to submission portals: rethinking the role of domain-specific databases in CollecTF. Database (Oxford) 2016 (baw055).
Kodama, Y., Shumway, M., Leinonen, R., 2012. The sequence read archive: explosive growth of sequencing data. Nucleic Acids Res. 40, D54–D56.
Kruger, J., Rehmsmeier, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34, W451–W454.
Kulakovskiy, I.V., Medvedeva, Y.A., Schaefer, U., Kasianov, A.S., Vorontsov, I.E., Bajic, V.B., Makeev, V.J., 2013. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Res. 41, D195–D202.
Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Soboleva, A.V., Kasianov, A.S., Ashoor, H., Ba-Alawi, W., Bajic, V.B., Medvedeva, Y.A., Kolpakov, F.A., Makeev, V.J., 2016. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. 44, D116–D125.
Kutach, A.K., Kadonaga, J.T., 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. Mol. Cell. Biol. 20, 4754–4764.
Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., Huala, E., 2012. The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40, D1202–D1210.
Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., Snyder, M., 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831.
Latchman, D.S., 1997. Transcription factors: an overview. Int. J. Biochem. Cell Biol. 29, 1305–1312.
Liang, Z.Y., Lai, H.Y., Yang, H., Zhang, C.J., Yang, H., Wei, H.H., Chen, X.X., Zhao, Y.W., Su, Z.D., Li, W.C., Deng, E.Z., Tang, H., Chen, W., Lin, H., 2016. Pro54DB: a database for experimentally verified sigma-54 promoters. Bioinformatics 33, 467–469 (btw630).
Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., Castagnoli, L., Cesareni, G., 2012. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. 40, D857–D861.
Lin, M., Shen, X., Chen, X., 2011. PAIR: the predicted *Arabidopsis* interactome resource. Nucleic Acids Res. 39, D1134–D1140.
Lin, H., Deng, E.Z., Ding, H., Chen, W., Chou, K.C., 2014. iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition. Nucleic Acids Res. 42, 12961–12972.
Liu, Z.P., Wu, C., Miao, H., Wu, H., 2015. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database (Oxford) 2015 (bav095).
Loots, G.G., Ovcharenko, I., 2004. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. Nucleic Acids Res. 32, W217–W221.
Loots, G., Ovcharenko, I., 2007. ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. Bioinformatics 23, 122–124.
Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T., 2011. Entrez gene: gene-centered information at NCBI. Nucleic Acids Res. 39, D52–D57.
Makita, Y., Nakao, M., Ogasawara, N., Nakai, K., 2004. DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. Nucleic Acids Res. 32, D75–D77.
Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., Wasserman, W.W., 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. 42, D142–D147.
Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A., Wasserman, W.W., 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 44, D110–D115.
Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., Kloos, D.U., Land, S., Lewicki-Potapov, B., Michael, H., Munch, R., Reuter, I., Rotert, S., Saxel, H., Scheer, M., Thiele, S., Wingender, E., 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 31, 374–378.
Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., Wingender, E., 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34, D108–D110.
McGinnis, S., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32, W20–W25.
Monteiro, P.T., Mendes, N.D., Teixeira, M.C., d'Orey, S., Tenreiro, S., Mira, N.P., Pais, H., Francisco, A.P., Carvalho, A.M., Lourenco, A.B., Sa-Correia, I., Oliveira, A.L., Freitas, A.T., 2008. YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. Nucleic Acids Res. 36, D132–D136.
Morgenstern, B., 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15, 211–218.
Munch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., Jahn, D., 2005. Virtual footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics 21, 4187–4189.
Naika, M., Shameer, K., Mathew, O.K., Gowda, R., Sowdhamini, R., 2013. STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in *Arabidopsis* and rice. Plant Cell Physiol. 54, e8.
Obayashi, T., Nishida, K., Kasahara, K., Kinoshita, K., 2011. ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. Plant Cell Physiol. 52, 213–219.
Ohler, U., Niemann, H., 2001. Identification and analysis of eukaryotic promoters: recent computational approaches. Trends Genet. 17, 56–60.
Ohyanagi, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T., Fujii, Y., Antonio, B.A., Nagamura, Y., Imanishi, T., Ikeo, K., Itoh, T., Gojobori, T., Sasaki, T., 2006. The rice annotation project database (RAP-DB): hub for *Oryza sativa* ssp. *Japonica* genome information. Nucleic Acids Res. 34, D741–D744.
O'Malley, R.C., Huang, S.S., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., Ecker, J.R., 2016. Cistrome and epicistrome features shape the regulatory DNA landscape. Cell 165, 1280–1292.
Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., Buell, C.R., 2007. The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res. 35, D883–D887.
Pachkov, M., Erb, I., Molina, N., van Nimwegen, E., 2007. SwissRegulon: a database of genome-wide annotations of regulatory sites. Nucleic Acids Res. 35, D127–D131.
Pachkov, M., Balwierz, P.J., Arnold, P., Ozonov, E., van Nimwegen, E., 2013. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res. 41, D214–D220.
Piriyapongsa, J., Bootchai, C., Ngamphiw, C., Tongsima, S., 2014. microPIR2: a comprehensive database for human-mouse comparative study of microRNA-promoter interactions. Database (Oxford) 2014 (bau115).
Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., Sandelin, A., 2010. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. Nucleic Acids Res. 38, D105–D110.
Qiu, P., 2003. Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. Biochem. Biophys. Res. Commun. 309, 495–501.
Robasky, K., Bulyk, M.L., 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. 39, D124–D128.
Sakai, H., Lee, S.S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C.C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., Itoh, T., 2013. Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. Plant Cell Physiol. 54, e6.
Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millan-Zarate, D., Blattner, F.R., Collado-Vides, J., 2000. RegulonDB (version 3.0): transcriptional regulation and operon organization in *Escherichia coli* K-12. Nucleic Acids Res. 28, 65–67.
Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muniz-Rascado, L., Garcia-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martinez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernandez, S., Alquicira-Hernandez, K., Lopez-Fuentes, A., Porron-Sotelo, L., Huerta, A.M., Bonavides-Martinez, C., Balderas-Martinez, Y.I., Pannier, L., Olvera, M., Labastida, A., Jimenez-Jacinto, V., Vega-Alvarado, L., Del Moral-Chavez, V., Hernandez-Alvarez, A., Morett, E., Collado-Vides, J., 2013.

RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. 41, D203–D213.
Sebastian, A., Contreras-Moreira, B., 2014. FootprintDB: a database of transcription factors with annotated *cis*-elements and binding interfaces. Bioinformatics 30, 258–265.
Sebestyen, E., Nagy, T., Suhai, S., Barta, E., 2009. DoOPSearch: a web-based tool for finding and analysing common conserved motifs in the promoter regions of different chordate and plant genes. BMC Bioinformatics 10 (Suppl. 6) (S6).
Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Shinozaki, K., 2002. Functional annotation of a full-length *Arabidopsis* cDNA collection. Science 296, 141–145.
Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M., Solovyev, V.V., 2003. PlantProm: a database of plant promoter sequences. Nucleic Acids Res. 31, 114–117.
Shameer, K., Ambika, S., Varghese, S.M., Karaba, N., Udayakumar, M., Sowdhamini, R., 2009. STIFDB-*Arabidopsis* stress responsive transcription factor database. Int. J. Plant Genomics 2009, 583429.
Shazman, S., Lee, H., Socol, Y., Mann, R.S., Honig, B., 2014. OnTheFly: a database of *Drosophila melanogaster* transcription factors and their binding sites. Nucleic Acids Res. 42, D167–D171.
Sierro, N., Makita, Y., de Hoon, M., Nakai, K., 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. Nucleic Acids Res. 36, D93–D96.
Spivak, A.T., Stormo, G.D., 2012. ScerTF: a comprehensive database of benchmarked position weight matrices for *Saccharomyces* species. Nucleic Acids Res. 40, D162–D168.
Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., Huala, E., 2008. The *Arabidopsis* information resource (TAIR): gene structure and function annotation. Nucleic Acids Res. 36, D1009–D1014.
Tang, Y.T., Li, S.J., Kao, H.Y., Tsai, S.J., Wang, H.C., 2011. Using unsupervised patterns to extract gene regulation relationships for network construction. PLoS One 6, e19633.
Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Goncalves, J.P., Mira, N.P., dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P., Madeira, S.C., Oliveira, A.L., Freitas, A.T., Sa-Correia, I., 2014. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. Nucleic Acids Res. 42, D161–D166.
The UniProt Consortium, 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212.
Tsai, H.K., Chou, M.Y., Shih, C.H., Huang, G.T., Chang, T.H., Li, W.H., 2007. MYBS: a comprehensive web server for mining transcription factor binding sites in yeast. Nucleic Acids Res. 35, W221–W226.
Wasserman, W.W., Sandelin, A., 2004. Applied bioinformatics for the identification of regulatory elements. Nat. Rev. Genet. 5, 276–287.
Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., Zheng, H., Goity, A., van Bakel, H., Lozano, J.C., Galli, M., Lewsey, M.G., Huang, E., Mukherjee, T., Chen, X., Reece-Hoyes, J.S., Govindarajan, S., Shaulsky, G., Walhout, A.J., Bouget, F.Y., Ratsch, G., Larrondo, L.F., Ecker, J.R., Hughes, T.R., 2014. Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158, 1431–1443.
Wu, W.S., Hsieh, Y.C., Lai, F.J., 2016. YCRD: yeast combinatorial regulation database. PLoS One 11, e0159213.
Yamamoto, Y.Y., Obokata, J., 2008. ppdb: a plant promoter database. Nucleic Acids Res. 36, D977–D981.
Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K., Abe, T., 2007. Identification of plant promoter constituents by analysis of local distribution of short sequences. BMC Genomics 8, 67.
Yevshin, I., Sharipov, R., Valeev, T., Kel, A., Kolpakov, F., 2017. GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. Nucleic Acids Res. 45, D61–D67.
Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X., Welch, L., Grotewold, E., 2011. AGRIS: the *Arabidopsis* gene regulatory information server, an update. Nucleic Acids Res. 39, D1118–D1122.
Younger, S.T., Corey, D.R., 2011. Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. Nucleic Acids Res. 39, 5682–5691.
Zawel, L., Reinberg, D., 1995. Common themes in assembly and function of eukaryotic transcription complexes. Annu. Rev. Biochem. 64, 533–561.
Zhao, F., Xuan, Z., Liu, L., Zhang, M.Q., 2005. TRED: a transcriptional regulatory element database and a platform for in silico gene regulation studies. Nucleic Acids Res. 33, D103–D107.
Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A.C., Ghane, T., Di Felice, R., Rohs, R., 2013. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic Acids Res. 41, W56–W62.
Zhu, J., Zhang, M.Q., 1999. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. Bioinformatics 15, 607–611.
Zhu, L.J., Christensen, R.G., Kazemian, M., Hull, C.J., Enuameh, M.S., Basciotta, M.D., Brasefield, J.A., Zhu, C., Asriyan, Y., Lapointe, D.S., Sinha, S., Wolfe, S.A., Brodsky, M.H., 2011. FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. Nucleic Acids Res. 39, D111–D117.

„Isolation and comprehensive *in silico* characterisation of a new *3-hydroxy-3-methylglutaryl-coenzyme A reductase 4* (*HMGR4*) gene promoter from *Salvia miltiorrhiza*: comparative analyses of plant *HMGR* promoters"

(Publikacja II)

plants

MDPI

Article

Isolation and Comprehensive in Silico Characterisation of a New *3-Hydroxy-3-Methylglutaryl-Coenzyme A Reductase 4* (*HMGR4*) Gene Promoter from *Salvia miltiorrhiza*: Comparative Analyses of Plant *HMGR* Promoters

Małgorzata Majewska *, Łukasz Kuźma and Piotr Szymczyk

Department of Biology and Pharmaceutical Botany, Medical University of Lodz, Muszyńskiego 1, 90-151 Lodz, Poland; lukasz.kuzma@umed.lodz.pl (Ł.K.); piotr.szymczyk@umed.lodz.pl (P.S.)
* Correspondence: malgorzata.majewska1@stud.umed.lodz.pl

Abstract: *Salvia miltiorrhiza* synthesises tanshinones with multidirectional therapeutic effects. These compounds have a complex biosynthetic pathway, whose first rate limiting enzyme is 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR). In the present study, a new 1646 bp fragment of the *S. miltiorrhiza HMGR4* gene consisting of a promoter, 5′ untranslated region and part of a coding sequence was isolated and characterised in silico using bioinformatics tools. The results indicate the presence of a TATA box, tandem repeat and pyrimidine-rich sequence, and the absence of CpG islands. The sequence was rich in motifs recognised by specific transcription factors sensitive mainly to light, salicylic acid, bacterial infection and auxins; it also demonstrated many binding sites for microRNAs. Moreover, our results suggest that *HMGR4* expression is possibly regulated during flowering, embryogenesis, organogenesis and the circadian rhythm. The obtained data were verified by comparison with microarray co-expression results obtained for *Arabidopsis thaliana*. Alignment of the isolated *HMGR4* sequence with other plant *HMGR*s indicated the presence of many common binding sites for transcription factors, including conserved ones. Our findings provide valuable information for understanding the mechanisms that direct transcription of the *S. miltiorrhiza HMGR4* gene.

Keywords: *HMGR4*; microRNA; promoter; *Salvia miltiorrhiza*; transcription factor; transcription factor binding site

Citation: Majewska, M.; Kuźma, Ł.; Szymczyk, P. Isolation and Comprehensive in Silico Characterisation of a New *3-Hydroxy-3-Methylglutaryl-Coenzyme A Reductase 4* (*HMGR4*) Gene Promoter from *Salvia miltiorrhiza*: Comparative Analyses of Plant *HMGR* Promoters. *Plants* **2022**, *11*, 1861. https://doi.org/10.3390/plants11141861

Academic Editors: Gitanjali Yadav, Giuseppe Puglia and Anyela Camargo Rodríguez

Received: 20 April 2022
Accepted: 12 July 2022
Published: 16 July 2022

1. Introduction

Salvia miltiorrhiza Bunge, also known as Red sage, or Chinese sage, is an important species used in traditional Chinese medicine. The dried root is used alone or in combination with other herbs to treat various ailments including cardiovascular diseases, menstrual disorders and insomnia [1,2]. In addition, it has been found to have potential in treating cancer [3], Parkinson's [4] and Alzheimer's disease [5], as well as renal deficiency [6], hepatocirrhosis [7], acute lung injury [8], fibrosis [9], neuropathic pain [10], diabetes mellitus [11], or alcohol dependence [12]. The main bioactive compounds responsible for such medical properties are quinone diterpenoids (e.g., tanshinones) and phenolic acids. The tanshinones are biosynthesised from intermediates generated in mevalonate (MVA) and methylerythritol phosphate (MEP) pathways [13]. The key rate-limiting enzyme in the MVA pathway converting 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) to MVA is HMG-CoA reductase (HMGR) [14]. The pivotal role of *HMGR* in plant metabolism is emphasised by the precise regulation of its function at the level of transcription, post-transcription, translation and post-translation [15,16]. To date, five *S. miltiorrhiza HMGR* gene sequences (*HMGR*—*HMGR4*) have been identified and deposited in the GenBank database [17–19]. A combination of cDNA sequence similarity searches with exon/intron structure indicates

that *HMGR* (EU680958.1) and *HMGR4* (JN831103.1), and *HMGR2* (FJ747636.1) and *HMGR3* (JN831102.1) are probably two pairs of duplicated genes, respectively [19]. Although the coding sequences of the *S. miltiorrhiza HMGR* genes have been identified and described, some of their promoter sequences remain unknown. These include the *HMGR4* promoter.

Promoter sequence analysis provides much valuable information for understanding the regulation of gene expression. Motifs such as the TATA box, CpG islands, tandem repeats or transcription factor binding sites (TFBSs) deserve special attention. Genome-wide analyses indicate that most in vivo functional TFBSs are located in the proximal promoter region [20,21]. These sites form clusters, thus improving interactions of corresponding transcription factors (TFs) to ensure a better execution of their regulatory functions [22]. An essential functional linkage exists between TFs and RNA polymerase II, acting as a large, conformationally flexible multiprotein complex known as a Mediator [23]. This complex regulates polymerase activity by transmitting signals from TFs. Groups of TFs form complex networks of dependencies and act in a coordinated manner in response to intracellular and environmental signals, thus directing many biological processes [24]. Gene expression is also regulated by the activity of microRNAs (miRNAs). They are mainly known as post-transcriptional and translational inhibitors of gene expression. The miRNAs cut the mRNA strands, destabilise the mRNA by shortening its poly(A) tail, and reduce the efficiency of the translation process [25,26]. However, studies on human and *Arabidopsis thaliana* indicate that miRNAs can also regulate gene expression during transcription by binding to promoter sequences [27–30]. In conclusion, to understand the mechanisms driving gene expression, it is necessary to also understand the nature of the promoter regions. The first step to achieving this goal requires use of bioinformatics tools.

A. thaliana is the most widely-studied plant in modern biology. Its wide appeal for scientists results from its fast growth rate, easy maintenance and small space requirements [31]. Moreover, the plant indicates a good tolerance to homozygosity and self-fertility [32]. Genetic studies are attracted by the small size (132 Mbp) of its completely sequenced genome [32]. Due to its similarity to other plants, *A. thaliana* has become the starting point for studying numerous aspects of plant cell-, molecular- and system-biology [33]. The abundance of research conducted on *A. thaliana* has led to the creation of numerous clones, cloning vectors, mutant lines, seeds, databases and online tools containing genomic, epigenomic, transcriptomic and proteomic data [34].

This work describes the isolation of a new *S. miltiorrhiza HMGR4* promoter and 5′ untranslated region (5′UTR) sequences, and their in silico characterisation via specialised databases such as PlantPan 2.0, TSSP and miRBase. Furthermore, the comprehensive in silico analysis presented herein presents valuable new information about the regulatory functions of *HMGR* promoters. The data can be used to create modified or synthetic promoters which could be active under certain controlled conditions [35]. In this way, numerous medically-important metabolites may be obtained.

2. Results

2.1. In Silico Analysis of the S. miltiorrhiza HMGR4 Promoter

The 1646 bp sequence obtained by the DNA walking method was deposited in GenBank under accession number KT921337.1 (Figure 1). The sequence contained a 51 bp coding region which perfectly coincided with *HMGR4* gene sequence JN831103.1 identified by Ma et al. [19]. The TATA box was located at bases −28 to −33 from TSS. One tandem repeat (at −1296 to −1353) and no CpG islands were found. A pyrimidine-rich sequence (PRS) was recognised in the 5′UTR.

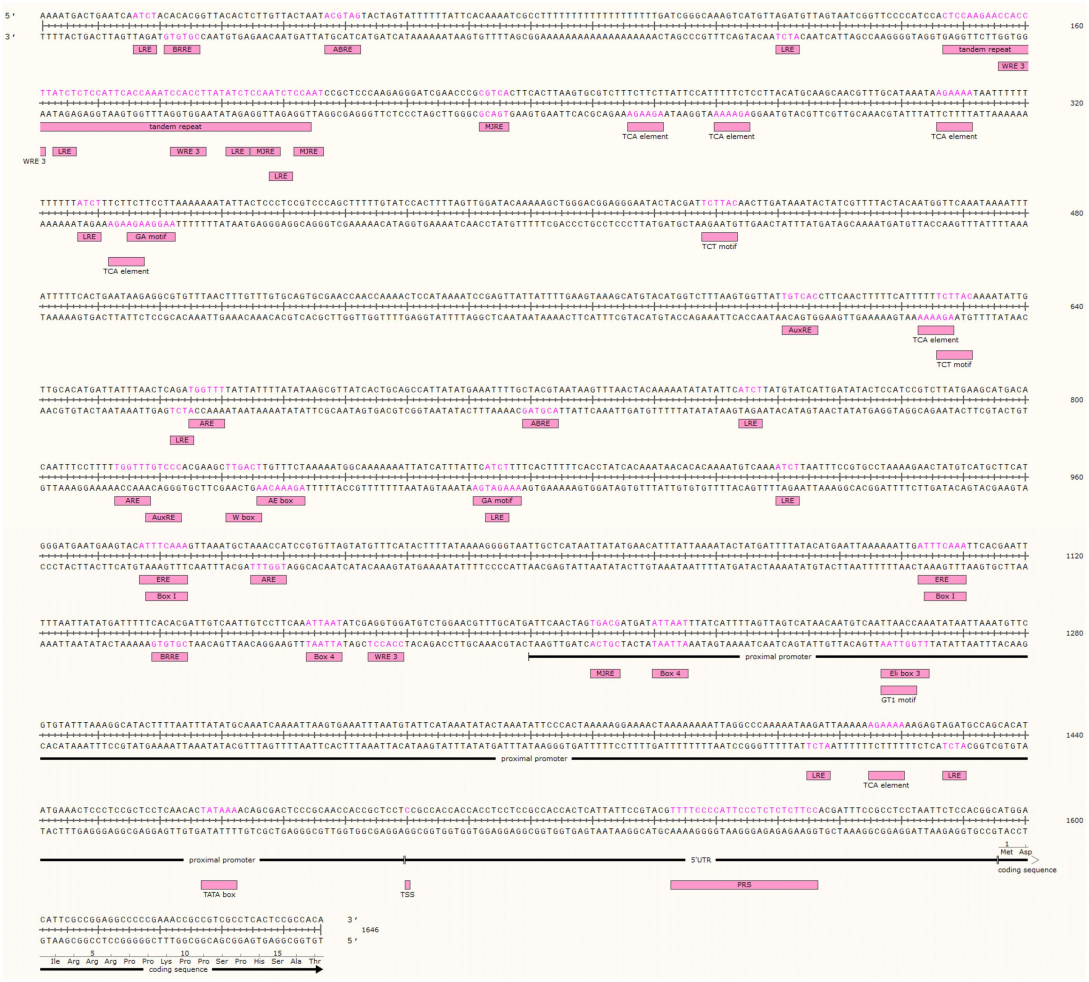


Figure 1. Isolated *S. miltiorrhiza HMGR4* promoter sequence (1499 bp), 5′ untranslated region (5′UTR) (96 bp) and coding sequence fragment (51 bp). The potential TATA box, transcription start site (TSS), pyrimidine-rich sequence (PRS), tandem repeat and consensus sequences for hormone-, pathogen-, wounding-, light-, and anaerobic-responsive elements are signed and marked in pink on the strands. ABRE, abscisic acid-responsive element; ARE, anaerobic-responsive element; AuxRE, auxin-responsive element; BRRE, brassinosteroid-responsive element; ERE, ethylene-responsive element; LRE, light-responsive element; MJRE, methyl jasmonate-responsive element.

Moreover, 5369 TFBSs and 365 potentially interacting TFs described previously in *A. thaliana* were revealed using the PlantPan 2.0 tool (File S1). Each of the obtained TFs could interact with a number of binding sites. The TFBSs were identified at both strands of the examined sequence. The similarity score between the binding sites found in the *S. miltiorrhiza HMGR4* promoter and those detected in *A. thaliana* ranged from 0.7 to 1.0.

Additional analysis of the entire *HMGR4* promoter sequence revealed the following commonly-known consensus sequences: two auxin-responsive elements (AuxREs); seven salicylic acid (SA)-responsive elements, including W box and TCA-elements; two brassinosteroid-responsive elements (BRREs); two ethylene-responsive elements (EREs); two abscisic acid-responsive elements (ABREs); four methyl jasmonate-responsive elements (MJREs); one pathogen-responsive element Eli box 3; three wounding- and pathogen-responsive elements WRE3; three anaerobic-responsive elements (AREs); and twenty-four

light-responsive elements (LREs), including ATCT sequences, Box 4, GT1 motif, TCT motifs, GA motifs, AE box, Box I (Figure 1). The vast majority of these results were in agreement with data provided by PlantPan 2.0, which demonstrated the presence of numerous binding sites for TFs responding to these types of stimulation (Table S1). Only anaerobic-sensitive TFs were not found. The most commonly-observed TFs were those representing the following families: GATA (light); MYB-related and Dof (auxin); WRKY (SA, wounding and pathogen); NAC; NAM (brassinosteroid and abscisic acid (ABA)); MYB-related (ethylene); and CAMTA (MeJa) (Table S1). The consensus sequences listed above were distributed along the entire *HMGR4* promoter, with some being located only a few nucleotides from each other; this allows more precise regulation of gene expression by dimerization of binding TFs. Such sequences included two SA-responsive TCA elements and two LREs (Table 1).

Table 1. Transcription factor binding sites (TFBSs) with the potential to form dimers detected in the *S. miltiorrhiza HMGR4* promoter sequence.

Transcription Factor (TF) Family Name	Fragment of Promoter Sequence with Underlined TFBSs [a] and Potentially Interacting TF Pairs
-	AGAAAAATGGAATAAGAAGA two salicylic acid (SA)-responsive TCA elements (AGAAAA and AGAAGA) spaced by eight nucleotides
-	ATCTCCAATCT two LREs (ATCT) spaced by three nucleotides
HD-ZIP	atTTAATgtaTTCATAAATata ATML1/HDG1 and PDF2/ATML1 spaced by four nucleotides
WRKY	AGTCATAACAATGTCAA WRKY2/WRKY14/WRKY45/WRKY57/WRKY69 and WRKY2/WRKY14/WRKY45/WRKY57/WRKY69 spaced by six nucleotides
Dof	AAAGAAAAAAGA DOF5.4 and DOF5.4 spaced by two nucleotides

[a] Most conserved positions within a matrix were written in capital letters.

The search of the PlantPan 2.0 database revealed about 234 TFs that could interact with 915 TFBSs in the *HMGR4* proximal promoter (File S2). To complement the findings described above, the binding sites located in the proximal promoter were analysed in terms of their response to external factors. Such data were available for 666 TFBSs (File S2). The results indicated that *HMGR*4 transcription may be most dependent on light, SA, bacterial infection and auxins, and less dependent on ABA, gibberellin, chitin, cold or salt stress (Figure 2). The response to these agents was mainly associated with the following TF families: GATA (light); WRKY (SA, bacterial infection, chitin); bZIP (auxins); MYB-related, WRKY and C2H2 (ABA); MYB-related and MADS box (gibberellins); C2H2 and WRKY (cold); and WRKY and MYB-related (salt stress) (Table S2). To determine the stage of *S. miltiorrhiza* development at which the *HMGR4* gene expression regulation is likely to occur, 450 TFBSs and interacting TFs located in its proximal promoter were examined (File S2). The findings indicated flowering, embryogenesis, organogenesis (root, shoot, leaf and flower development) and circadian rhythm (Figure 3).

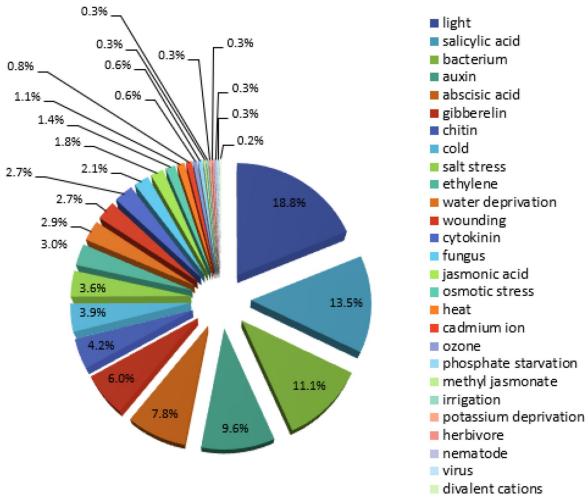


Figure 2. Classification of TFBSs found in the proximal *S. miltiorrhiza HMGR4* promoter with regard to their response to biotic and abiotic factors.

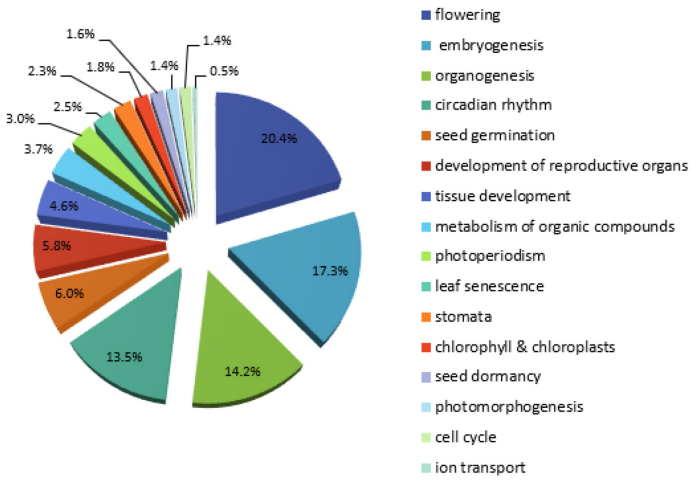


Figure 3. Classification of TFBSs found in the proximal *S. miltiorrhiza HMGR4* promoter with regard to their biological functions.

As *HMGR* genes are crucial for the production of intermediates in the biosynthesis pathway for tanshinones, a literature search was performed for TFs that positively regulate tanshinone production. Following this, based on the results obtained with the PlantPan 2.0 database, the *S. miltiorrhiza HMGR4* promoter sequence was searched for the presence of binding sites for the 20 identified TFs; of these, the results indicated the presence of the following five TFs: BHLH6 (MYC2) (14 binding sites), BHLH74 (2 sites), BZIP20 (32 sites), WRKY2 (8 sites) and WRKY61 (9 sites) (File S1) [36–40]. A number of TFBSs were also found to be located in the proximal promoter region. The obtained results suggest that *HMGR4* may play a role in the biosynthesis of tanshinones.

Furthermore, in silico analysis of the *HMGR4* promoter and 5′UTR revealed potentially interacting miRNAs (Table 2). In total, 12 mature miRNAs were found, 8 binding within the promoter and 4 within the 5′UTR.

Table 2. Identification of microRNAs (miRNAs) potentially interacting with the *S. miltiorrhiza HMGR4* promoter sequence and 5′UTR identified by the miRBase tool.

miRNA Name and Source	miRNA Sequence	Sequence Alignment Position Start/End	Strand	e-Value
***HMGR4* Promoter**				
miR1128 *Saccharum sp.*	UACUACUCCCUCCGUCCCAAA	350/368 405/423	+ –	0.75 4.2
miR6462c-5p *Populus trichocarpa*	AAGGGACAAAAAUGGCAUAAGA	259/279	–	3.5
miR1128 *Triticum aestivum*	UACUACUCCCUCCGUCCGAAA	350/368	+	4.2
miR1436 *Oryza sativa* and *Hordeum vulgare*	ACAUUAUGGGACGGAGGGAGU	354/368	–	6.2
miR5205a *Medicago truncatula*	CAUACAAUUUGGGACGGAGGGAG	355/374	–	9.1
miR8740 *Gossypium raimondii*	UAAUGAUGUGGCACAAUAUUA	634/653	–	9.1
miR11573a and miR11573b *Picea abies*	UUGGGGAGCGUAUUGUAGAUU	197/216	–	9.1
5′UTR of *HMGR4*				
miR477 *Gossypium raimondii*	CGAAGUCUUGGAAGAGAGUAA	59/75	–	3.2
miR6180 *Hordeum vulgare*	AGGGUGGAAGAAAGAGGGCG	55/69	–	3.9
miR4993 *Glycine max*	GAGCGGCGGCGGUGGAGGAUG	13/30	–	6.9
miR12107-5p *Citrus sinensis*	CUGAUGAGAGAGCGAAUGAUA	51/66	–	8.4

2.2. Microarray and Next-Generation Sequencing (NGS) Co-Expression Data Analysis

The Protein BLAST analysis revealed that the coding sequence of *S. miltiorrhiza HMGR4* (AEZ55673.1) was more similar to *A. thaliana HMGR1* (NP_177775.2), with an identity of 73.76%, than to *A. thaliana HMGR2* (NP_179329.1), with one of 69.48%. Additionally, a phylogenetic study of coding sequences indicated that *HMGR4* from *S. miltiorrhiza* (JN831103.1) and *HMGR1* from *A. thaliana* (NM_106299.4) were more closely related to each other (Figure S1). Therefore, the *A. thaliana HMGR1* gene (At1g76490) was used for further co-expression research. As a result of the conducted microarray analysis, 166 TF genes co-expressed with *A. thaliana HMGR1* in the *r* range of 0.5–1.0 were found: 41 in AtGenExpress Hormone and Chemical Compendium, 37 in AtGenExpress Abiotic Stress Compendium, 34 in AtGenExpress Pathogen Compendium, 25 in AtGenExpress Tissue Compendium, and 29 in AtGenExpress Plus—Extended Tissue Compendium (Table S3). The RNA-seq analysis did not identify any TF genes co-expressed with *A. thaliana HMGR1* in the WGCNA correlation range of 0.5–1.0.

2.3. Comparison of the in Silico HMGR4 Analysis Results with Microarray Co-Expression Data

The comparison identified 32 common TFs in the *S. miltiorrhiza HMGR4* promoter (Table 3), with the most well-represented being TFs from the homeodomain-leucine zipper (HD-ZIP) and WRKY families. The common TFs participated mainly in response to hormones (ABA, ethylene, jasmonic acid and cytokinins), other abiotic factors (light, salt stress, water deprivation, heat and iron ion) and biotic agents (bacteria), embryogenesis, organogenesis (root development), flowering or, finally, tissue development (epidermis) (Table 3).

Table 3. TFs common between in silico analysis of the *S. miltiorrhiza HMGR4* promoter and microarray co-expression studies with *A. thaliana HMGR1*.

TF Family Name	TF Gene Name and Locus	Processes in Which TF is Involved [a]	r-Value [b]	TFBS Motif and Localisation [c,d]	
Homeodomain; HD-ZIP	*ATHB-13;* At1g69780	cotyledon and leaf morphogenesis; primary root development; sucrose-signalling pathway	0.594	ATAAT 310; 309	AATAA 308; 307
	ATHB-16; At4g40060	regulation of timing of transition from vegetative to reproductive phase; repression of cell expansion during plant development; response to blue light	0.562	ATAAT 309	
	HDG1; At3g61150	maintenance of floral organ identity	0.507	ATTAA 161	**TTAAT 1218; 1331; 1332**
	ANL2; At4g00730	regulation of tissue-specific accumulation of anthocyanins; cellular organisation of primary root; cuticle hydrocarbon biosynthetic process; plant-type cell wall modification; root hair cell differentiation	0.524	**TTAAT 1218**	ATTAA 1161
	ATML1; At4g21750	cotyledon development; epidermal cell differentiation; seed dormancy and germination	0.583	**TTAAT 1332** **ATTTA** 1057; **1282**	**TAAAT** 987; **1272**; **1346**
	PDF2; At4g04890	cotyledon development; epidermal cell differentiation; seed dormancy and germination; maintenance of floral organ identity	0.587	**ATTTA** 1057; **1282**	**TAAAT** 987; **1272**; **1346**
Homeodomain; bZIP; HD-ZIP	*HAT5;* At3g01470	leaf morphogenesis; response to blue light and salt stress	0.521	ATAAT 307; 310	AATAA 308; 307
bZIP	*BZIP25;* At3g54620	positive regulation of seed maturation	0.666	CCACG 822 TACGT 46; 720 ACGTA 47; 721	AACGT 290;1188 ACGTT 291; 1189
WRKY	*WRKY2;* At5g56270	regulation of basal cell division patterns during early embryogenesis; establishment of cell polarity; longitudinal axis specification; pollen development	0.575/ 0.557	TGACT 5; 832 **AGTCA** 111; **1240**	TTGAC 831 **GTCAA** 913; 1147; **1251**
	WRKY14; At1g30650	-	0.576		

Table 3. *Cont.*

TF Family Name	TF Gene Name and Locus	Processes in Which TF is Involved [a]	r-Value [b]	TFBS Motif and Localisation [c,d]	
	WRKY57; At1g69310	response to osmotic stress, salt stress and water deprivation	0.504		
	WRKY45; At3g01970	phosphate ion transport	0.515	TTGAC 830; 831 TGACT 5; 832	**AGTCA** 111; **1240** **GTCAA** 913; 1147; **1251**
	WRKY69; At3g58710	-	0.564		
Myb/SANT; ARR-B	*ARR2;* At4g16110	His-to-Asp phosphorelay signal transduction system; expression of nuclear genes for components of mitochondrial complex I; ethylene- and cytokinin-activated signalling pathways; promotion of cytokinin-mediated leaf longevity; root meristem growth; seed growth; stomatal movement	0.552/0.611	AATCT 15; 197; 919 **AGATT 1405**	AATCC 179; 204; 548
	ARR14; At2g01760	His-to-Asp phosphorelay signal transduction system; activation of some type-A response regulators in response to cytokinins	0.509/ 0.530		
Myb/SANT; MYB	*MYB6;* At4g09460	response to ethylene, abscisic acid (ABA), indole-3-acetic acid, and *Pseudomonas syringae* pv. *phaseolica*	0.599	ACCTA 886	
MYB-related	*RVE1;* At5g17300	morning-phased TF integrating circadian clock and auxin pathways; regulation of free indole-3-acetic acid level in time-of-day specific manner; negative regulation of freezing tolerance	0.501	ATATC 1166	
	RVE4; At5g02840	regulation of circadian rhythm	0.545/ 0.573	ATATC 1166	**GATAT 1215**
EIN3; EIL	*EIL1;* At2g27050	positive regulation of ethylene response pathway; cellular response to iron ion; defence response to bacterium	0.554	TGTAT 374; 759	ATACA 391
	EIL3; At1g73730	ethylene response pathway; sulphur metabolic process; cellular response to iron ion	0.525	ATGTA 757	

Table 3. *Cont.*

TF Family Name	TF Gene Name and Locus	Processes in Which TF is Involved [a]	r-Value [b]	TFBS Motif and Localisation [c,d]	
MADS box; MIKC	*AGL18;* At3g57390	negative regulation of flowering and short-day photoperiodism; pollen development	0.567	TTTCC 804; 801 TTTTG 805 **CAAAA 1396**	**AGAAA** 293; **1402** **GGAAA 1364**; **1362** TTTTT 77; 78; 79, 80; 81; 82; 83; 84
	SVP; At2g22540	inhibition of floral transition in autonomous flowering pathway; identity of floral meristem; response to temperature stimulus	0.537	TTTCC 801	**GGAAA 1362**
NAC; NAM	*NAC055;* At3g15500	jasmonic acid-mediated signalling pathway; response to water deprivation	0.594	TACGT 44; 718 CGTA 720 ACAT 644	TTGAC 829 ACGTA 44; 718; 720
	NAC072; At4g27410	activator in ABA-mediated dehydration response	0.543	CGTA 720 TTGAC 829	ACAT 644
NF-YB	*NFYB5;* At2g47810		0.558	CTAAT 42 ATCGG 102; 131 CCCAT 139 CCAAG 149; 213 CCAAT 195; 202 GTTGG 389 ATGGG 959 ATTGC 1039 CGAAT 1115 **ATTAG 1389** CCTAT 890 TTTGG 811	**TCAAT** 13; 1149; **1253** AATGG 462; 847 **CCACT** 144; 380; **1364** CCATT 169; 266; 701 CCAAC 530 AGTGG 592 ATTGA 766; 1099 **ACAAT** 460; 800; **1247** ATTGT 599; 637; 1145; 1152 **CCAAA** 176; 534; **1261**; **1396** **CAAAT** 177; 469; 897; 1107; 1161; **1262**; **1313**
NF-YC	*NFYC10;* At5g38140	protein heterodimerization activity	0.523		
TBP	*TBP2;* At1g55520	required for basal transcription (facilitating the recruitment of TFIID to the promoter, forming a preinitiation complex with RNA polymerase)	0.645	ATATA 746; 739 TTTTA 1022; 1078 **TAAAA** 541; **1465**; 1024	**TATAT** 677; **1305**; 743; 736 **ATAAA** 1025; **1466**; **1344**; 298; 439; 1024; **1465** **TTTAT** 669; 1016; 1072; **1297**; **1223**; **1303**; 674; 1077; **1302**
TCP	*TCP21;* At5g08330	positive regulation of circadian clock	0.546	CCCAC 818	

Table 3. *Cont.*

TF Family Name	TF Gene Name and Locus	Processes in Which TF is Involved [a]	r-Value [b]	TFBS Motif and Localisation [c,d]	
AP2; ERF	*RAP2-3;* At3g16770	cell death; heat acclimation; ethylene-activated signalling pathway; response to cytokinin, jasmonic acid and other organism	0.502	TAAGA 494	
C2H2	*AZF2;* At3g19580	inhibition of plant growth under abiotic stress conditions; negative regulation of ABA signalling during seed germination; positive regulation of leaf senescence; jasmonate early signalling response; response to chitin and water deprivation; plants overexpressing AZF2 have increased sensitivity to salt stress and barely survive under high salt conditions	0.503	**ACACT** 29; **1462**	
Dof	*DOF5.4;* At5g60850	metal ion binding; binding of OBF TFs to OCS elements	0.727	CGTTA 685 AACGT 286; 1184 ACGTT 289; 1187 GCCTT 79 CCTTT 80; 807 AAAGT 109; 982 AAGGG 1031 **AAGGA 1372**	ACCTT 158; 184; 605; 612 GCTTT 369 AAAGC 400; 571 **AAAGA** 939; **1413**; **1420** **AAAGG** 1030; **1289**; **1371** TCTTT 252; 328; 586; 874 TCCTT 275; 338; 806; 1156 **ACTTT** 382; 509; 881; 1021; **1291**
SBP	*SPL3;* At2g33810	promotion of vegetative phase change and flowering; vegetative- to reproductive-phase transition of meristem	0.596	AGTAC 51; 972 GTACA 578; 973 TGTAC 577 GTACT 52 ATACG 43 CGTAG 46	CTTAC 275; 427; 625 CGTCC 360 GGACG 405 CGAAC 524 CTACG 717 CGTAA 720

[a] The roles of the TFs were assumed based on the UniProt database. [b] r-value between the TFBSs found in the *S. miltiorrhiza HMGR4* promoter and those detected in *A. thaliana* ranged from 0.5 to 1.0. [c] For TFBSs, only the most conserved positions within a matrix were listed. [d] TFBSs localised in proximal promoter region were put in bold.

These 32 common TFs were scanned with the Genomatix Pathway System for the presence of interactions between them. The identified relationships are presented in Figure 4. It was found that SVP, AGL18 and SPL3 proteins were involved together in the regulation of flowering, and PDF2 and ATML1 in epidermal specification in embryos, respectively. Furthermore, NAC3 and RD26 responded to high salinity, drought and ABA, while NAC3 and ZF2 supported resistance to the herbivore *Spodoptera littoralis*. EIL3 and EIL1 played roles in regulating the response to sulphur deficiency and in ethylene signalling, and EBP transcription was light modulated through the EIN2-EIN3/EIL1 pathway.

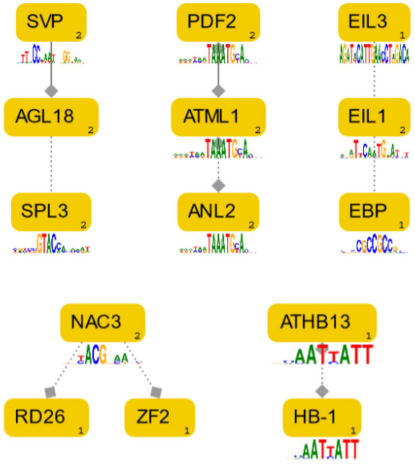


Figure 4. Interactions between TFs potentially binding to the *S. miltiorrhiza HMGR4* promoter found with the Pathway System tool. The presented dependencies are based on co-citation (dashed line) or expert-curation (solid line). Diamond-ended lines indicate that a given TF has a predicted binding site in dependent promoter sequence. The number in the lower right corner of TF indicates the number of interactions within the network (including those not displayed). EBP = RAP 2-3, NAC3 = NAC055, RD26 = NAC072, ZF2 = AZF2, HB-1 = HAT5.

Of the 32 common TFs that were recognised, 18 were found to interact with the TFBSs situated in the proximal promoter region (Table 3). The ability of these 18 TFs to bind to DNA as dimers or multimers was also tested. The TFBSs identified for HD-ZIP (ATML1, PDF2, HDG1), WRKY (WRKY2, WRKY14, WRKY45, WRKY57, WRKY69) and Dof (DOF5.4) are given in Table 1; all are closely located to each other, and only separated by a few nucleotides. The existence of experimentally-determined interactions between ATML1 and PDF2 proteins was confirmed by the BioGRID database.

2.4. Comparison of S. miltiorrhiza HMGR Promoters

The *S. miltiorrhiza HMGR1*, *HMGR2* and *HMGR4* promoter sequences were analysed using the Common TFs tool. Based on the findings, common binding sites for TFs were recognised, and these were classified into 22 matrix families (Figure 5, Table 4), with each single matrix family comprising identical or functionally-similar TFs identified by weight matrices. The following matrix families were found: *Arabidopsis* homeobox proteins (P$AHBP), L1 box (P$L1BX), MYB IIG-type binding sites (P$MIIG), DNA binding with one finger (P$DOFF), GT box elements (P$GTBX), MADS box proteins (P$MADS), MYB-like proteins (P$MYBL), MYB proteins with single DNA binding repeat (P$MYBS), NAC factors with transmembrane motif (P$NTMF), plant specific NAC proteins (P$NACF), transcription repressor KANADI (P$KAN1), W box family (P$WBXF), time-of-day-specific regulatory elements (P$TODS), nodulin consensus sequence 1 (P$NCS1), sweet potato DNA-binding factor with two WRKY domains (P$SPF1), zinc finger proteins (P$ZFAT), light-responsive elements (P$LREM), protein secretory pathway elements (P$PSPE), CGCG

box binding proteins (P$CGCG), proteins involved in programmed cell death response (P$PCDR), plant nitrate-responsive elements (P$PNRE) and finally, stomatal carpenter (P$SCAP). As can be seen in Figure 5, the distribution of the matrices within the *HMGR1* and *HMGR2* promoter sequences was strikingly similar. The above TF family analysis found that the *S. miltiorrhiza HMGR1*, *HMGR2* and *HMGR4* genes can be co-regulated in response to abiotic factors (auxins, gibberellins, ABA, SA, jasmonic acid, brassinosteroids, light, water deprivation, salt stress, cold or phosphate starvation), biotic factors (bacteria, fungi and viruses) and during root, stem, leaf and flower organogenesis (Table 4).

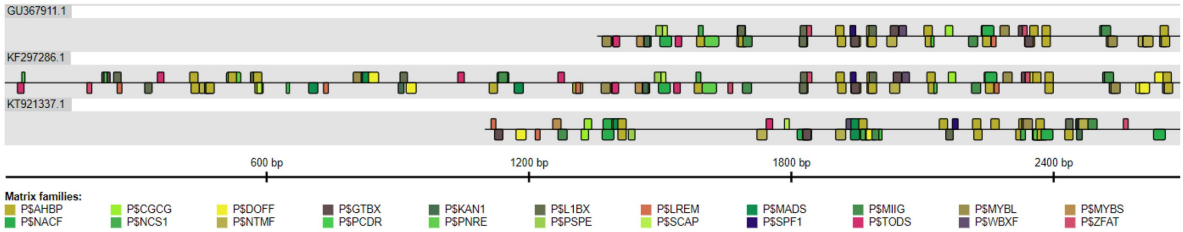


Figure 5. Distribution of matrix families common to the *S. miltiorrhiza* promoter sequences, i.e., *HMGR1* (GU367911.1), *HMGR2* (KF297286.1) and *HMGR4* (KT921337.1) identified by the Common TFs tool. Black lines correspond to the promoter sequences. Each matrix family is marked with a semicircular coloured symbol. The figure shows families found on the positive and negative strands. P$AHBP, *Arabidopsis* homeobox proteins; P$L1BX, L1 box; P$MIIG, MYB IIG-type binding sites; P$DOFF, DNA binding with one finger; P$GTBX, GT box elements; P$MADS, MADS box proteins; P$MYBL, MYB-like proteins; P$MYBS, MYB proteins with single DNA binding repeat; P$NTMF, NAC factors with transmembrane motif; P$NACF, plant specific NAC proteins; P$KAN1, transcription repressor KANADI; P$WBXF, W box family; P$TODS, time-of-day-specific regulatory elements; P$NCS1, nodulin consensus sequence 1; P$SPF1, sweet potato DNA-binding factor with two WRKY domains; P$ZFAT, zinc finger proteins; P$LREM, light-responsive elements; P$PSPE, protein secretory pathway elements; P$CGCG, CGCG box binding proteins; P$PCDR, proteins involved in programmed cell death response; P$PNRE, plant nitrate-responsive elements; P$SCAP, stomatal carpenter.

Table 4. Common TF matrix families found during in silico analysis of the *S. miltiorrhiza HMGR1*, *HMGR2*, *HMGR4* promoter sequences using the Common TFs tool.

TF Matrix Family	Processes in Which TF Is Involved [a]
Arabidopsis homeobox proteins (P$AHBP)	root, leaf and anther development; seed maturation; meristem initiation and growth; xylem and phloem pattern formation; cell differentiation; determination of bilateral symmetry; transition from vegetative to reproductive phase; glucosinolate metabolic process; response to: auxin, gibberellin, ABA, water deprivation, blue light and salt stress
L1 box (P$L1BX)	cotyledon development; seed germination and dormancy; epidermal cell differentiation; maintenance of floral organ identity
MYB IIG-type binding sites (P$MIIG)	root, seed, stamen and xylem development; cellular cadmium ion homeostasis; gibberellin and flavonol biosynthesis; defence response to fungi; response to: ABA, chitin, salt stress, cold, water deprivation, phosphate starvation, potassium ion and light
DNA binding with one finger (P$DOFF)	secondary shoot, cotyledon and seed development; cell wall modification; cell cycle; gibberellin biosynthesis; response to: SA, auxin, chitin, red light and cold
GT box elements (P$GTBX)	shoot system and stomatal complex development; trichome morphogenesis; seed maturation and germination; cell size and growth; response to: auxin and water deprivation

Table 4. *Cont.*

TF Matrix Family	Processes in Which TF Is Involved [a]
MADS box proteins (P$MADS)	flower, ovule and seed coat development; seed maturation; meristem structural organisation; transition from vegetative to reproductive phase; short-day photoperiodism; circadian rhythm; response to auxin
MYB-like proteins (P$MYBL)	integument, anther and pollen development; leaf morphogenesis; seed growth and dormancy; endothelial cell proliferation; vacuole organisation; wax biosynthesis; long-day photoperiodism; defence response to bacteria and fungi; response to: SA, brassinosteroid, gibberellin, ABA, jasmonic acid, chitin, salt, water deprivation and cold
MYB proteins with single DNA binding repeat (P$MYBS)	leaf and lateral root development; leaf senescence; circadian rhythm; peroxidase activity; auxin and sulphate ion homeostasis; response to: ABA, phosphate starvation, absence of light and high light intensity
NAC factors with transmembrane motif (P$NTMF)	leaf and trichome morphogenesis; xylem development; seed germination; photoperiodism; membrane protein proteolysis; response to: gibberellin, salt stress
plant specific NAC proteins (P$NACF)	leaf and secondary shoot development; primary shoot apical meristem specification; formation of organ boundary; regulation of timing of organ formation; response to water deprivation
transcription repressor KANADI (P$KAN1)	phenylpropanoid metabolic process
W box family (P$WBXF)	induced systemic resistance; JA-mediated signalling pathway; phosphate ion transport; defence response to: bacteria, fungi and viruses; response to: SA, chitin and wounding
time-of-day-specific regulatory elements (P$TODS)	circadian rhythm; red or far-red light signalling pathway; response to temperature
nodulin consensus sequence 1 (P$NCS1)	nodule-specific expression
zinc finger proteins (P$ZFAT)	regulation of root development; phosphate ion homeostasis
light-responsive elements (P$LREM)	response to hypoxia
protein secretory pathway elements (P$PSPE)	SA induction of secretion-related genes via NPR1
CGCG box binding proteins (P$CGCG)	leaf senescence; defence response to: bacteria, fungi and insects; response to: cold, auxins and water deprivation
proteins involved in programmed cell death response (P$PCDR)	regulation of expression of vacuolar processing enzyme
plant nitrate-responsive elements (P$PNRE)	nitrate assimilation; stomatal movement; response to: nitrate and water deprivation
stomatal carpenter (P$SCAP)	stomatal movement
sweet potato DNA-binding factor with two WRKY domains (P$SPF1)	-

[a] The roles of the TFs were assumed based on the MatInspector (Genomatix) database.

The FrameWorker tool indicated the existence of 10,000 10-element-frameworks within the *S. miltiorrhiza HMGR1*, *HMGR2* and *HMGR4* promoters. Two selected models are provided in Figure 6. The frameworks were created based on 52 matrix families common to the tested sequences, some of which are mentioned above in the Common TF results section. The matrix families were located on the positive or negative strand of the promoters.

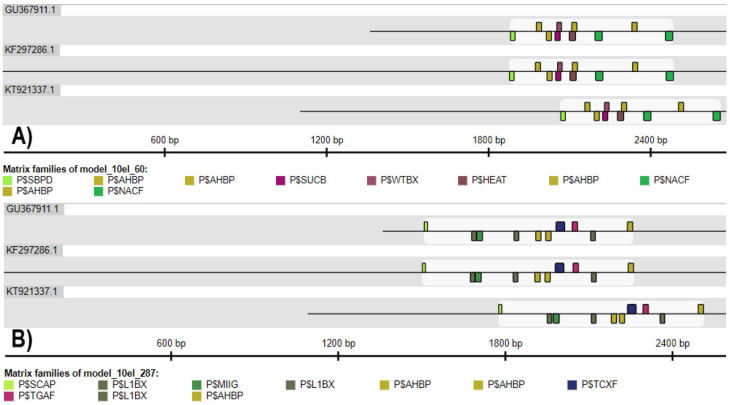


Figure 6. Selected 10-element-frameworks of TFBSs obtained for *S. miltiorrhiza* promoter sequences, i.e., *HMGR1* (GU367911.1), *HMGR2* (KF297286.1) and *HMGR4* (KT921337.1) using the FrameWorker tool. Black lines correspond to the promoter sequences. Each matrix family is marked with a semicircular coloured symbol. The figure shows families found on the positive and negative strands. (**A**) P$SBPD, SBP-domain proteins; P$AHBP, *Arabidopsis* homeobox proteins; P$SUCB, sucrose box; P$WTBX, WT box; P$HEAT, heat shock factors; P$NACF, plant specific NAC proteins; (**B**) P$SCAP, stomatal carpenter; P$L1BX, L1 box; P$MIIG, MYB IIG-type binding sites; P$AHBP, *Arabidopsis* homeobox proteins; P$TCXF, CRC domain containing tesmin/TSO1-like CXC (TCX) factors; P$TGAF, basic/leucine zipper-type TFs of the TGA-family.

The DiAlign TF tool analysis found the *HMGR1* (GU367911.1) and *HMGR2* (KF297286.1) promoters to demonstrate the greatest similarity (97%). In contrast, only 17% similarity was found between *HMGR4* (KT921337.1) and *HMGR2*, and 14% between *HMGR4* and *HMGR1*. The greatest number of identical areas was revealed in the proximal fragments of the analysed promoters, as well as in the beginning and the middle of the distal parts. Common overlapping TFBSs were identified in locations where all three tested sequences showed high local similarity; these included two binding sites for *Arabidopsis* homeobox proteins (P$AHBP), one site for SBP domain proteins (P$SBPD), one site for W box family proteins (P$WBXF), one site for DNA binding with one finger factors (P$DOFF), one GT box element (P$GTBX) and, finally, one L1 box (P$L1BX). The PlantPan 2.0 and MatInspector (Genomatix) databases indicated that P$AHBP proteins are mainly involved in the response to hormones (auxins, ABA, cytokinins and gibberellins) and the initiation and development of shoot, root and flower meristems. P$SBPD TFs are associated with inflorescence development, flowering and leaf epidermis differentiation. In turn, P$WBXF matrix family responds to hormonal stimulation (SA, ABA, ethylene and jasmonic acid), other abiotic factors (salt stress, wounding, osmotic stress, heat, water deprivation and cold) and biotic agents (bacteria, fungi and viruses), and also participate in leaf senescence. P$DOFF proteins are primarily involved in the regulation of flowering, circadian rhythm and in response to hormones (auxins and SA). P$GTBX factors participate in the organogenesis of flowers and shoots. In addition, P$L1BX proteins are needed for epidermis development and seed germination. These data are available in File S1 and Table 4.

2.5. The Conservation of Plant HMGR Promoters

MEGA X software alignment of 36 sequences spanning the proximal promoters and 5′UTRs of the plant *HMGR* genes revealed the presence of conserved regions; these are marked in blue in Figure 7. These regions were detected in both the 5′UTRs and proximal promoters. In most of the tested sequences, PRS was identified within the preserved areas. Additional analysis with the DiAlign TF tool revealed the presence of conserved TFBSs. However, no TFBS was found to be conserved in any of the analysed sequences. The most conserved site was the TATA box, detected in 41.7% of the sequences. Two

preserved binding sites for TFs belonging to the P$AHBP (*Arabidopsis* homeobox protein) and P$GCCF (GCC box family) families were detected in 27.8% of cases. Sites for P$TDTF (transposase-derived proteins), P$MYBL (MYB-like proteins) and P$L1BX (L1 box) proteins were identified in 25% of the sequences. Binding sites for P$DREB (dehydration responsive element binding factors) and P$ROOT (root hair-specific *cis*-elements in angiosperms) families were found in 22.2%. These TFBSs are highlighted in red in Figure 7.

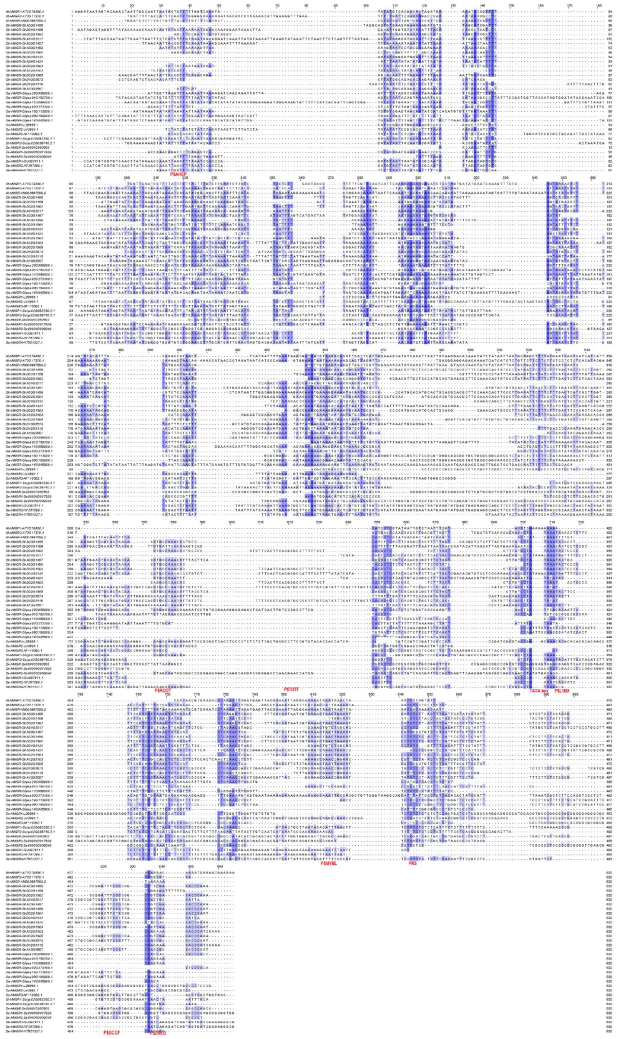


Figure 7. Alignment of proximal promoter regions and 5′UTRs of *Arabidopsis thaliana* (At), *Arabidopsis lyrata* (Al), *Gossypium hirsutum* (Gh), *Glycine max* (Gm), *Oryza sativa* (Os), *Solanum lycopersicum* (Sl), *Zea mays* (Zm) and *Salvia miltiorrhiza* (Sm) *HMGR* genes. Conserved nucleotides are marked in blue. The darker the colour, the greater the degree of conservation within the analysed sequences. The proximal promoters are present at the beginning and the 5′UTRs at the end of the displayed sequences, respectively. Conserved TFBSs identified by DiAlign TF tool are highlighted in red. P$AHBP, *Arabidopsis* homeobox protein; P$GCCF, GCC box family; P$TDTF, transposase-derived proteins; P$MYBL, MYB-like proteins; P$L1BX, L1 box; P$DREB, dehydration responsive element binding factors; P$ROOT, root hair-specific cis-elements in angiosperms.

Apart from the conserved TATA box and PRS motifs, the investigated *S. miltiorrhiza HMGR4* promoter was found to contain several other common binding sites for TFs, belonging to P$AHBP (*Arabidopsis* homeobox protein), P$GTBX (GT box elements), P$DOFF (DNA binding with one finger) and P$L1BX (L1 box); these were shared by 8–13% of the tested sequences. Nucleotide pairwise alignment of *S. miltiorrhiza HMGR4* with the remaining tested sequences found 13 to 31% identity (mean 24.7%).

3. Discussion

Our study presents new data regarding the isolated *S. miltiorrhiza HMGR4* promoter and 5′UTR and compares these sequences with other plant *HMGR*s.

Initially, the sequences were examined for the presence of certain distinctive motifs (Figure 1). One such motif found in the investigated sequences is the TATA box, which is estimated to be present in 30–50% of all known promoters [41] and 29% of *A. thaliana* promoters [42]. It was also detected in the *S. miltiorrhiza HMGR2* promoter [43]. Previous studies in human and yeast models indicate that the TATA box is more common in promoters of highly-regulated genes and in those stimulated by stress factors and extracellular signals [44–48]; in contrast, TATA-less genes are more constitutively expressed and associated with key processes such as cell growth [44–48]. In addition, promoters containing the TATA box appear to have a more conserved sequence than those that do not [49].

The *HMGR4* promoter is also characterised by the presence of a single tandem repeat. This motif is estimated to be present in only 25% of promoters [50], and is absent from the *S. miltiorrhiza HMGR2* promoter [43]. As tandem repeats are more prone to mutation, which affects the length of the repeat and thus local nucleosome positioning and gene expression rate, genes whose promoters have tandem repeats show higher rates of transcription divergence [50].

Both the *HMGR2* promoter and the studied *HMGR4* promoter lack CpG islands [43]. The cytosines in the CG dinucleotides of the islands can be methylated, thus inhibiting gene expression [51,52]. However, the CpG cluster is not required for methylation since, in plants, it can also occur within the CHG and CHH sequences (H = A, T or C) [53].

A PRS is also detected in the 5′UTR of the described sequence. This is a rather rare observation, but not an unprecedented one, as a PRS has also been found in the 5′UTR of the *S. miltiorrhiza HMGR2* gene [43]. It is believed to take part in the organisation of the spliceosomal complex [54].

Furthermore, the examined *S. miltiorrhiza HMGR4* promoter sequence turned out to be rich in TFBSs recognised by specific TFs (File S1). The conducted research indicates that the number of promoter regulatory elements and interacting proteins positively correlates with divergence of gene expression [55]. The *HMGR4* proximal promoter was found to contain consensus sequences mainly related to the response to light, SA, bacterial infection, auxins, ABA, gibberellin, chitin, cold or, finally, salt stress (Figure 2), suggesting that these factors may participate in gene regulation. One previous paper investigating the influence of external agents on *S. miltiorrhiza HMGR4* found that treatment with 200 µM MeJa had no significant effect on *HMGR4* expression in either leaves or roots [19]. It is important to note that the effect of these factors has been examined on other *S. miltiorrhiza HMGR* genes. Chen et al. found that only 100% red light slightly increased the expression of *HMGR* in hairy root culture, while other light types (e.g., 100% far-red, 100% blue, red:far-red, blue:far-red, red:blue, red:blue:UV) had an inhibitory effect [56]. In contrast, Wang et al. noted that UV-B enhanced the expression of *HMGR1* in roots almost 5-fold compared to an untreated control [57]. Incubation of hairy root culture with 100 µM SA raised *HMGR* transcript level, peaking at three-times higher than baseline after 36 h [58]. In turn, 200 µM SA has been found to have a differential effect on *HMGR2* promoter in leaf material [43]. A decrease in its activity was observed after 12, 24 and 48 h of treatment, while a 2.5- to 3-fold rise compared to the calibrator values was observed after 72 and 96 h. Bacteria appeared to be good activators of *HMGR* expression. The addition of *Pseudomonas brassicacearum* subsp. *neoaurantiaca* and *Pseudomonas thivervalensis* to *S. miltiorrhiza* root culture resulted in

2.1- and 1.5-fold enhancements in HMGR enzyme activity, respectively [59]. In addition, *Streptomyces pactum* Act12 increased *HMGR1* expression by more than a factor of 35 on day 14 relative to the calibrator [60]. Exposure to 2.85 µM IAA and 2.88 µM gibberellic acid improved the activity of the *HMGR2* promoter, resulting in manifold higher expression compared to the calibrator in 96 h [43]. In turn, 200 µM and 10 µM ABA upregulated *HMGR1* and *HMGR2*, respectively [43,61]. Salt stress (50 mM, 100 mM, 200 mM, 300 mM NaCl) enhanced the expression and enzymatic activity of *HMGR* in leaves and roots over 48 h of exposure [62]. It was also found that 200 mM NaCl inhibited the level of *HMGR1* transcript in leaves and roots as compared to the calibrator [63].

As non-coding regions are generally not highly conserved, from an evolutionary perspective, finding such motifs in the promoter or 5′UTR sequences suggest they have functional importance [64].

Within the studied 36 *HMGR* sequences, the most frequently-identified conserved motif was the TATA box (Figure 7). This is a known sequence that has been conserved from *Archaebacteria* to humans [65]. The other TFBSs discussed in the Results section were shared by a much smaller number of tested sequences (27.8% or fewer).

The study also examined the possibility that more complex structures could be created by TFs interacting with the *HMGR4* promoter. TFs participate in the regulation of gene expression as monomers, dimers (homo- and heterodimers) or multimers. Dimers and multimers are often preferred by nature because they allow specific interactions with the promoter sequence and bind with high affinity [66]. One TF monomer can create dimers or multimers with different functions, thus mediating the regulation of various genes, by forming bonds with multiple, but not random, protein partners [67]. Our analyses revealed the presence of closely-related TFBSs for the following TFs in the *S. miltiorrhiza HMGR4* proximal promoter: HD-ZIP (ATML1, PDF2 and HDG1), WRKY (WRKY2, WRKY14, WRKY45, WRKY57 and WRKY69) and Dof (DOF5.4) (Table 1). The HD-ZIP proteins are unique to the plant kingdom. TFs from the family are unable to bind to DNA as monomers [68]. They form homo- and heterodimers via the leucine zipper motif [67]. Meanwhile, ATML1 was able to create a heterodimer with its paralogue PDF2 in studies on *Nicotiana benthamiana* and *A. thaliana* [69,70], and to form homodimers in vitro [69,71]. It has been shown that WRKY TFs can interact with DNA as monomers or create homo- and heterodimers, especially those with a leucine zipper motif [72–74], WRKY2 protein was found to form homodimers in *Hordeum vulgare* [75], while WRKY45 created homodimers in vitro by exchanging β4-β5 strands in *Oryza sativa* [72]. The Dof TFs have a multifunctional domain that allows them to bind to DNA and interact with other proteins [76] and to establish homo- and heterodimers.

As miRNA is believed to regulate plant promoter activity at the transcription level, the investigated *HMGR4* sequence was searched for miRNA binding sites and interacting miRNAs [27]. Of the 12 miRNAs potentially binding to the *HMGR4* promoter and 5′UTR sequences (Table 2), non-conserved miR1128 and miR1436 were detected during deep sequencing in *S. miltiorrhiza* [77]. However, their significance in the regulation of gene expression has not yet been investigated at the experimental level.

The results of our present in silico analysis of the *HMGR4* promoter and 5′UTR sequences constitute a strong basis for planning future necessary experiments on *S. miltiorrhiza*.

4. Materials and Methods

4.1. Plant Material

S. miltiorrhiza plants were cultivated from seeds provided by the Garden of Medicinal Plants of the Medical University of Lodz. The plants were grown in pots containing composite soil at 26 ± 2 °C under natural light. Six-month-old plants were used for the experiment.

4.2. Isolation of the S. miltiorrhiza HMGR4 Promoter Sequence

Genomic DNA used for isolation of the *HMGR4* promoter was obtained from young, fresh *S. miltiorrhiza* leaves and stems according to the method proposed by Khan et al. [78]. The DNA was analysed using a NanoPhotometer P300 (Implen, Munich, Germany) to determine its quantity and quality based on A_{260}/A_{280} and A_{260}/A_{230} ratios. The *HMGR4* promoter region was isolated using GenomeWalker Universal Kit (Takara Bio, Kusatsu, Japan) according to the manufacturer's instructions. A 5′-terminal fragment of the *HMGR4* gene, deposited in GenBank under accession number JN831103.1, was used as a target for designing GSP1 and GSP2 specific primers (Table S4). The PCR reactions were performed using the Advantage 2 PCR Kit (Takara Bio, Kusatsu, Japan). The amplified DNA fragments were TOPO-TA cloned into a pCRII-TOPO vector (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The inserts were Sanger sequenced (CoreLab, Medical University of Lodz, Lodz, Poland) with specific primers listed in Table S4. The *HMGR4* promoter sequence was assembled using CodonCode Aligner software version 8.0.2 (CodonCode Corporation, Centerville, MA, USA).

The isolation and sequencing of the *S. miltiorrhiza HMGR4* promoter took approximately one month.

4.3. In Silico Analysis of the S. miltiorrhiza HMGR4 Promoter Sequence

The obtained *HMGR4* promoter sequence was characterised in silico using available tools and databases [79]. The promoter, TATA box, TSS and 5′UTR positions were identified with TSSP software (Softberry Inc., Mount Kisco, NY, USA) [80]. Tandem repeats, CpG islands, TFBSs and TFs were detected using PlantPan 2.0 [81]. The promoter sequence was screened for the presence of commonly-known consensus motifs reported in the published literature. Assuming that the functional TFBSs are concentrated mainly in the proximal promoters, special attention was paid to the promoter region lying within 300 bp from the TSS. The miRBase tool was used to search for miRNA binding sites and interacting miRNAs in the obtained promoter and 5′UTR sequences [82].

4.4. Microarray and NGS Co-Expression Data Analysis

Protein BLAST (NCBI, Bethesda, MD, USA) and MEGA X version 10.2.6 (Pennsylvania State University, State College, PA, USA) [83] were employed to determine which of the *A. thaliana HMGR* genes is a homologue of the *S. miltiorrhiza HMGR4* gene. Analyses were performed on coding sequences. Expression Angler (BAR, Toronto, ON, Canada) [84] and *Arabidopsis* RNA-seq Database [85] were utilised to find TF genes co-expressed with the selected *A. thaliana HMGR* gene. The Expression Angler tool has access to the expression results for approximately 22,000 *Arabidopsis* genes, while the *Arabidopsis* RNA-seq Database integrates 28,164 publicly available *Arabidopsis* RNA-seq libraries. The following microarray dataset compendiums were used during the study: AtGenExpress Hormone and Chemical, AtGenExpress Abiotic Stress, AtGenExpress Pathogen, AtGenExpress Tissue, and AtGenExpress Plus—Extended Tissue. The Pearson's correlation coefficient (r) ranging from 0.50 to 1.00 (moderate to strong positive correlation) was applied to identify co-regulated genes. Information on the detected TFs was obtained from the UniProt database [86]. The collected results were compared with the in silico data found by PlantPan 2.0. The occurrence of interactions between the received common TFs was determined using Pathway System (Genomatix, Munich, Germany) and the BioGRID database version 4.4.201 [87].

4.5. Comparison of S. miltiorrhiza HMGR Promoters

The entire available *S. miltiorrhiza HMGR* promoter sequences, i.e., *HMGR1* (GU367911.1), *HMGR2* (KF297286.1), and *HMGR4* (KT921337.1) were analysed with Common TFs, FrameWorker and DiAlign TF tools from Genomatix, Munich, Germany. Common TFs was used for preliminary analysis of the common TFBSs and interacting TFs located anywhere in the investigated promoters. The search only included sites that were common to all three sequences. The similarity of the matrix to the tested sequences was set to the highest

possible value, i.e., 0.05. The FrameWorker tool permitted the common, most complex framework of TFBSs to be extracted from the input promoters. Frameworks are defined as TFBSs that occur in the same order and in a specificed space range in all of the sequences. The DiAlign TF allowed for multiple alignment of the studied *HMGR* promoters, and revealed conserved regions and TFBSs located therein. The analyses using the Genomatix tools were based on matrix library version 11.3 and default search criteria.

4.6. Assessment of Conservation of Plant HMGR Promoters

The conservation of the 36 available *HMGR* promoters derived from plants such as *A. thaliana*, *Arabidopsis lyrata*, *Glycine max*, *Gossypium hirsutum*, *Oryza sativa*, *Solanum lycopersicum*, *Zea mays* and *S. miltiorrhiza* was assessed by aligning their sequences. Proximal promoters and 5′UTR sequences (each sequence 500 bases long) were obtained from PlantPan 3.0 [88] and NCBI Nucleotide databases with the participation of the UniProt [86]. Alignments were performed using the MUSCLE algorithm from the MEGA X software, version 10.2.6 [83]. TFBSs located in the conserved regions of the compared sequences were recognised with the DiAlign TF tool (Genomatix, Munich, Germany).

5. Conclusions

Regulation of *S. miltiorrhiza HMGR4* gene expression can occur during flowering, embryogenesis, organogenesis and circadian rhythm, and are influenced mainly by factors such as light, SA, bacterial infection and auxins.

The presence of binding sites for TFs that promote the biosynthesis of tanshinones may indicate that the *S. miltiorrhiza HMGR4* gene plays an important role in the production of these metabolites.

A comparison of TFBSs and TFs in the *S. miltiorrhiza HMGR1*, *HMGR2*, and *HMGR4* promoter sequences indicates that these genes can be co-regulated in response to abiotic and biotic factors, and during organogenesis.

The *S. miltiorrhiza HMGR4* promoter is not highly conserved.

Future research on the *S. miltiorrhiza HMGR4* promoter could be developed towards preparing promoter deletion mutants, and studying their transcriptional activity [89]. Moreover, mutagenesis of particular TFBSs could be suitable for experimental verification of their importance in response to biotic or abiotic factors [89]. The TFs or other regulatory proteins could be isolated using a yeast-one hybrid (Y1H) system and the promoter segments as bait [90]. Isolated TFs could be functionally characterised by studying their DNA binding properties, and their potential to increase expression of specific genes [91,92]. These studies could be verified by chromatin immunoprecipitation-sequencing (ChIP-seq) of DNA fragments that are associated with particular proteins [93]. Finally, regulatory networks of TFs and other proteins playing a pivotal role in the response to certain external factors could be built using transcriptomic RNA sequencing, and weighted gene co-expression network analysis (WGCNA) [94,95].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants11141861/s1, Figure S1: Unrooted dendrogram of *HMGR* sequences from *S. miltiorrhiza* and *A. thaliana*; Table S1: TFs responsive to light, hormone, wounding and pathogen stimulation found in the entire *S. miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 database; Table S2: TFs responsive to major abiotic and biotic factors found in the proximal *S. miltiorrhiza HMGR4* promoter using the PlantPan 2.0 database; Table S3: TF genes co-expressed with *A. thaliana HMGR1* found with the Expression Angler tool; Table S4: Primers used in the study.

Author Contributions: Conceptualization, M.M. and P.S.; methodology, M.M. and P.S.; formal analysis, M.M. and P.S.; investigation, M.M. and P.S.; writing—original draft preparation, M.M.; writing—review and editing, P.S. and Ł.K.; visualization, M.M.; supervision, P.S. and Ł.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Medical University of Lodz (grant nos. 502-03/3-012-02/502-34-103, 503/3-012-01/503-31-001-19-00 and 503/3-012-02/503-31-001-19-00).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article and supplementary material.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, Y.; Wan, R.; Lin, Y.; Yang, L.; Chen, Y.; Liu, C. Recent advance on research and application of Salvia miltiorrhiza. *Asian J. Pharmacodyn. Pharmacokinet.* **2007**, *7*, 99–130.
2. Cheng, T.O. Cardiovascular effects of Danshen. *Int. J. Cardiol.* **2007**, *121*, 9–22. [CrossRef] [PubMed]
3. Zhang, H.S.; Wang, S.Q. Salvianolic acid B from Salvia miltiorrhiza inhibits tumor necrosis factor-alpha (TNF-alpha)-induced MMP-2 upregulation in human aortic smooth muscle cells via suppression of NAD(P)H oxidase-derived reactive oxygen species. *J. Mol. Cell. Cardiol.* **2006**, *41*, 138–148. [CrossRef] [PubMed]
4. Zhang, H.N.; An, C.N.; Zhang, H.N.; Pu, X.P. Protocatechuic acid inhibits neurotoxicity induced by MPTP in vivo. *Neurosci. Lett.* **2010**, *474*, 99–103. [CrossRef] [PubMed]
5. Wong, K.K.; Ho, M.T.; Lin, H.Q.; Lau, K.F.; Rudd, J.A.; Chung, R.C.; Fung, K.P.; Shaw, P.C.; Wan, D.C. Cryptotanshinone, an acetylcholinesterase inhibitor from Salvia miltiorrhiza, ameliorates scopolamine-induced amnesia in Morris water maze task. *Planta Med.* **2010**, *76*, 228–234. [CrossRef]
6. Kang, D.G.; Oh, H.; Sohn, E.J.; Hur, T.Y.; Lee, K.C.; Kim, K.J.; Kim, T.Y.; Lee, H.S. Lithospermic acid B isolated from Salvia miltiorrhiza ameliorates ischemia/reperfusion-induced renal injury in rats. *Life Sci.* **2004**, *75*, 1801–1816. [CrossRef]
7. Wu, Z.M.; Wen, T.; Tan, Y.F.; Liu, Y.; Ren, F.; Wu, H. Effects of salvianolic acid a on oxidative stress and liver injury induced by carbon tetrachloride in rats. *Basic Clin. Pharmacol. Toxicol.* **2007**, *100*, 115–120. [CrossRef]
8. Li, J.; Xu, M.; Fan, Q.; Xie, X.; Zhang, Y.; Mu, D.; Zhao, P.; Zhang, B.; Cao, F.; Wang, Y.; et al. Tanshinone IIA ameliorates seawater exposure-induced lung injury by inhibiting aquaporins (AQP) 1 and AQP5 expression in lung. *Respir. Physiol. Neurobiol.* **2011**, *176*, 39–49. [CrossRef]
9. Che, X.H.; Park, E.J.; Zhao, Y.Z.; Kim, W.H.; Sohn, D.H. Tanshinone II A induces apoptosis and S phase cell cycle arrest in activated rat hepatic stellate cells. *Basic Clin. Pharmacol. Toxicol.* **2010**, *106*, 30–37. [CrossRef]
10. Isacchi, B.; Fabbri, V.; Galeotti, N.; Bergonzi, M.C.; Karioti, A.; Ghelardini, C.; Vannucchi, M.G.; Bilia, A.R. Salvianolic acid B and its liposomal formulations: Anti-hyperalgesic activity in the treatment of neuropathic pain. *Eur. J. Biochem.* **2011**, *44*, 552–558. [CrossRef]
11. Yang, X.Y.; Qiang, G.F.; Zhang, L.; Zhu, X.M.; Wang, S.B.; Sun, L.; Yang, H.G.; Du, G.H. Salvianolic acid A protects against vascular endothelial dysfunction in high-fat diet fed and streptozotocin-induced diabetic rats. *J. Asian Nat. Prod. Res.* **2011**, *13*, 884–894. [CrossRef] [PubMed]
12. Serra, S.; Vacca, G.; Tumatis, S.; Carrucciu, A.; Morazzoni, P.; Bombardelli, E.; Colombo, G.; Gessa, G.L.; Carai, M.A. Anti-relapse properties of IDN 5082, a standardized extract of Salvia miltiorrhiza, in alcohol-preferring rats. *J. Ethnopharmacol.* **2003**, *88*, 249–252. [CrossRef]
13. Lu, S. *The Salvia Miltiorrhiza Genome*; Springer: Cham, Switzerland, 2019; pp. 129–139.
14. Liao, P.; Hemmerlin, A.; Bach, T.J.; Chye, M.L. The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnol. Adv.* **2016**, *34*, 697–713. [CrossRef] [PubMed]
15. Hemmerlin, A. Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant Sci.* **2013**, *203–204*, 41–54. [CrossRef] [PubMed]
16. Leivar, P.; Antolín-Llovera, M.; Ferrero, S.; Closa, M.; Arró, M.; Ferrer, A.; Boronat, A.; Campos, N. Multilevel control of Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase by protein phosphatase 2A. *Plant Cell* **2011**, *23*, 1494–1511. [CrossRef]
17. Liao, P.; Zhou, W.; Zhang, L.; Wang, J.; Yan, X.; Zhang, Y.; Zhang, R.; Li, L.; Zhou, G.; Kai, G. Molecular cloning, characterization and expression analysis of a new gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from Salvia miltiorrhiza. *Acta Physiol. Plant.* **2009**, *31*, 565–572. [CrossRef]
18. Dai, Z.; Cui, G.; Zhou, S.F.; Zhang, X.; Huang, L. Cloning and characterization of a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from Salvia miltiorrhiza involved in diterpenoid tanshinone accumulation. *J. Plant Physiol.* **2011**, *168*, 148–157. [CrossRef]
19. Ma, Y.; Yuan, L.; Wu, B.; Li, X.; Chen, S.; Lu, S. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in Salvia miltiorrhiza. *J. Exp. Bot.* **2012**, *63*, 2809–2823. [CrossRef]

20. Hernandez-Garcia, C.M.; Finer, J.J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* **2014**, *217–218*, 109–119. [CrossRef]
21. Yu, C.P.; Lin, J.J.; Li, W.H. Positional distribution of transcription factor binding sites in Arabidopsis thaliana. *Sci. Rep.* **2016**, *6*, 25164. [CrossRef]
22. Hussain, T.; Rehman, N.; Inam, S.; Ajmal, W.; Afroz, A.; Muhammad, A.; Zafar, Y.; Ali, G.M.; Khan, M.R. Homotypic clusters of transcription factor binding sites in the first large intron of AGL24 MADS-box transcription factor are recruited in the enhancement of floral expression. *Plant Mol. Biol. Rep.* **2019**, *37*, 24–40. [CrossRef]
23. Allen, B.L.; Taatjes, D.J. The Mediator complex: A central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 155–166. [CrossRef] [PubMed]
24. Khan, D.; Ziegler, D.J.; Kalichuk, J.L.; Hoi, V.; Huynh, N.; Hajihassani, A.; Parkin, I.A.P.; Robinson, S.J.; Belmonte, M.F. Gene expression profiling reveals transcription factor networks and subgenome bias during Brassica napus seed development. *Plant J.* **2022**, *109*, 477–489. [CrossRef]
25. Engels, B.M.; Hutvagner, G. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* **2006**, *25*, 6163–6169. [CrossRef] [PubMed]
26. Pillai, R.S.; Bhattacharyya, S.N.; Filipowicz, W. Repression of protein synthesis by miRNAs: How many mechanisms? *Trends Cell Biol.* **2007**, *17*, 118–126. [CrossRef]
27. Yang, G.; Li, Y.; Wu, B.; Zhang, K.; Gao, L.; Zheng, C. MicroRNAs transcriptionally regulate promoter activity in Arabidopsis thaliana. *J. Integr. Plant Biol.* **2019**, *61*, 1128–1133. [CrossRef]
28. 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. [CrossRef]
29. Place, R.F.; Li, L.C.; Pookot, D.; Noonan, E.J.; Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1608–1613. [CrossRef]
30. Kim, D.H.; Saetrom, P.; Snøve, O., Jr.; Rossi, J.J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16230–16235. [CrossRef]
31. Holland, C.K.; Jez, J.M. Arabidopsis: The original plant chassis organism. *Plant Cell Rep.* **2018**, *37*, 1359–1366. [CrossRef]
32. Woodward, A.W.; Bartel, B. Biology in Bloom: A Primer on the Arabidopsis thaliana Model System. *Genetics* **2018**, *208*, 1337–1349. [CrossRef] [PubMed]
33. Provart, N.J.; Alonso, J.; Assmann, S.M.; Bergmann, D.; Brady, S.M.; Brkljacic, J.; Browse, J.; Chapple, C.; Colot, V.; Cutler, S.; et al. 50 years of Arabidopsis research: Highlights and future directions. *New Phytol.* **2016**, *209*, 921–944. [CrossRef]
34. Cantó-Pastor, A.; Mason, G.A.; Brady, S.M.; Provart, N.J. Arabidopsis bioinformatics: Tools and strategies. *Plant J.* **2021**, *108*, 1585–1596. [CrossRef] [PubMed]
35. Dey, N.; Sarkar, S.; Acharya, S.; Maiti, I.B. Synthetic promoters in planta. *Planta* **2015**, *242*, 1077–1094. [CrossRef]
36. Shi, M.; Wang, Y.; Wang, X.; Deng, C.; Cao, W.; Hua, Q.; Kai, G. Simultaneous promotion of tanshinone and phenolic acid biosynthesis in Salvia miltiorrhiza hairy roots by overexpressing Arabidopsis MYC2. *Ind. Crops Prod.* **2020**, *155*, 112826. [CrossRef]
37. Deng, C.; Hao, X.; Shi, M.; Fu, R.; Wang, Y.; Zhang, Y.; Zhou, W.; Feng, Y.; Makunga, N.P.; Kai, G. Tanshinone production could be increased by the expression of SmWRKY2 in Salvia miltiorrhiza hairy roots. *Plant Sci.* **2019**, *284*, 1–8. [CrossRef]
38. Zhang, Y.; Xu, Z.; Ji, A.; Luo, H.; Song, J. Genomic survey of bZIP transcription factor genes related to tanshinone biosynthesis in Salvia miltiorrhiza. *Acta Pharm. Sin. B* **2018**, *8*, 295–305. [CrossRef] [PubMed]
39. Zhang, X.; Luo, H.; Xu, Z.; Zhu, Y.; Ji, A.; Song, J.; Chen, S. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in Salvia miltiorrhiza. *Sci. Rep.* **2015**, *5*, 11244. [CrossRef]
40. Chen, Y.; Wang, Y.; Guo, J.; Yang, J.; Zhang, X.; Wang, Z.; Cheng, Y.; Du, Z.; Qi, Z.; Huang, Y.; et al. Integrated Transcriptomics and Proteomics to Reveal Regulation Mechanism and Evolution of SmWRKY61 on Tanshinone Biosynthesis in Salvia miltiorrhiza and Salvia castanea. *Front. Plant Sci.* **2021**, *12*, 820582. [CrossRef]
41. Shahmuradov, I.A.; Umarov, R.K.; Solovyev, V.V. TSSPlant: A new tool for prediction of plant Pol II promoters. *Nucleic Acids Res.* **2017**, *45*, e65. [CrossRef]
42. Molina, C.; Grotewold, E. Genome wide analysis of Arabidopsis core promoters. *BMC Genom.* **2005**, *6*, 25. [CrossRef]
43. Szymczyk, P.; Grąbkowska, R.; Skała, E.; Żebrowska, M.; Balcerczak, E.; Jeleń, A. Isolation and characterization of a 3-hydroxy-3-methylglutaryl coenzyme A reductase 2 promoter from Salvia miltiorrhiza. *J. Plant Biochem. Biotechnol.* **2018**, *27*, 223–236. [CrossRef]
44. Tirosh, I.; Weinberger, A.; Carmi, M.; Barkai, N. A genetic signature of interspecies variations in gene expression. *Nat. Genet.* **2006**, *38*, 830–834. [CrossRef]
45. Yang, C.; Bolotin, E.; Jiang, T.; Sladek, F.M.; Martinez, E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **2007**, *389*, 52–65. [CrossRef] [PubMed]
46. Basehoar, A.D.; Zanton, S.J.; Pugh, B.F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **2004**, *116*, 699–709. [CrossRef]
47. Bae, S.H.; Han, H.W.; Moon, J. Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PLoS ONE* **2015**, *10*, e0120848. [CrossRef] [PubMed]

48. de Jonge, W.J.; O'Duibhir, E.; Lijnzaad, P.; van Leenen, D.; Groot Koerkamp, M.J.; Kemmeren, P.; Holstege, F.C. Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *EMBO J.* **2017**, *36*, 274–290. [CrossRef] [PubMed]
49. Cliften, P.; Sudarsanam, P.; Desikan, A.; Fulton, L.; Fulton, B.; Majors, J.; Waterston, R.; Cohen, B.A.; Johnston, M. Finding functional features in Saccharomyces genomes by phylogenetic footprinting. *Science* **2003**, *301*, 71–76. [CrossRef] [PubMed]
50. Vinces, M.D.; Legendre, M.; Caldara, M.; Hagihara, M.; Verstrepen, K.J. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* **2009**, *324*, 1213–1216. [CrossRef]
51. Davey, C.; Pennings, S.; Allan, J. CpG methylation remodels chromatin structure in vitro. *J. Mol. Biol.* **1997**, *267*, 276–288. [CrossRef] [PubMed]
52. Bell, A.C.; Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **2000**, *405*, 482–485. [CrossRef] [PubMed]
53. Henderson, I.R.; Jacobsen, S.E. Epigenetic inheritance in plants. *Nature* **2007**, *447*, 418–424. [CrossRef] [PubMed]
54. Saulière, J.; Sureau, A.; Expert-Bezançon, A.; Marie, J. The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol. Cell. Biol.* **2006**, *26*, 8755–8769. [CrossRef] [PubMed]
55. Tirosh, I.; Barkai, N. Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet.* **2008**, *24*, 109–113. [CrossRef]
56. Chen, I.J.; Lee, M.S.; Lin, M.K.; Ko, C.Y.; Chang, W.T. Blue light decreases tanshinone IIA content in Salvia miltiorrhiza hairy roots via genes regulation. *J. Photochem. Photobiol. B* **2018**, *183*, 164–171. [CrossRef]
57. Wang, C.H.; Zheng, L.P.; Tian, H.; Wang, J.W. Synergistic effects of ultraviolet-B and methyl jasmonate on tanshinone biosynthesis in Salvia miltiorrhiza hairy roots. *J. Photochem. Photobiol. B* **2016**, *159*, 93–100. [CrossRef]
58. Hao, X.; Shi, M.; Cui, L.; Xu, C.; Zhang, Y.; Kai, G. Effects of methyl jasmonate and salicylic acid on tanshinone production and biosynthetic gene expression in transgenic Salvia miltiorrhiza hairy roots. *Biotechnol. Appl. Biochem.* **2015**, *62*, 24–31. [CrossRef]
59. Yan, Y.; Zhang, S.; Zhang, J.; Ma, P.; Duan, J.; Liang, Z. Effect and mechanism of endophytic bacteria on growth and secondary metabolite synthesis in Salvia miltiorrhiza hairy roots. *Acta Physiol. Plant.* **2014**, *36*, 1095–1105. [CrossRef]
60. Yan, Y.; Zhang, S.; Yang, D.; Zhang, J.; Liang, Z. Effects of Streptomyces pactum Act12 on Salvia miltiorrhiza hairy root growth and tanshinone synthesis and its mechanisms. *Appl. Biochem. Biotechnol.* **2014**, *173*, 883–893. [CrossRef]
61. Yang, D.; Ma, P.; Liang, X.; Wei, Z.; Liang, Z.; Liu, Y.; Liu, F. PEG and ABA trigger methyl jasmonate accumulation to induce the MEP pathway and increase tanshinone production in Salvia miltiorrhiza hairy roots. *Physiol. Plant.* **2012**, *146*, 173–183. [CrossRef]
62. Yu, W.; Yu, Y.; Wang, C.; Zhang, Z.; Xue, Z. Mechanism by which salt stress induces physiological responses and regulates tanshinone synthesis. *Plant Physiol. Biochem.* **2021**, *164*, 10–20. [CrossRef] [PubMed]
63. Li, X.; Li, J.; Islam, F.; Najeeb, U.; Pan, J.; Hou, Z.; Shou, J.; Qin, J.; Xu, L. 5-Aminolevulinic acid could enhance the salinity tolerance by alleviating oxidative damages in Salvia miltiorrhiza. *Food Sci. Technol.* **2021**, *42*, 1–9. [CrossRef]
64. Setlow, J.K. *Genetic Engineering. Principles and Methods*, 1st ed.; Springer Science Business Media LLC: New York, NY, USA, 1997; pp. 15–47.
65. Reeve, J.N. Archaeal chromatin and transcription. *Mol. Microbiol.* **2003**, *48*, 587–598. [CrossRef]
66. Matthews, J.M. *Protein Dimerization and Oligomerization in Biology*; Springer Science+Business Media: New York, NY, USA, 2012; pp. 1–170.
67. Amoutzias, G.D.; Robertson, D.L.; Van de Peer, Y.; Oliver, S.G. Choose your partners: Dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* **2008**, *33*, 220–229. [CrossRef] [PubMed]
68. Ariel, F.D.; Manavella, P.A.; Dezar, C.A.; Chan, R.L. The true story of the HD-Zip family. *Trends Plant Sci.* **2007**, *12*, 419–426. [CrossRef]
69. San-Bento, R.; Farcot, E.; Galletti, R.; Creff, A.; Ingram, G. Epidermal identity is maintained by cell-cell communication via a universally active feedback loop in Arabidopsis thaliana. *Plant J.* **2014**, *77*, 46–58. [CrossRef]
70. Rombolá-Caldentey, B.; Rueda-Romero, P.; Iglesias-Fernández, R.; Carbonero, P.; Oñate-Sánchez, L. Arabidopsis DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box cis-element. *Plant Cell* **2014**, *26*, 2905–2919. [CrossRef]
71. Abe, M.; Takahashi, T.; Komeda, Y. Identification of a cis-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. *Plant J.* **2001**, *26*, 487–494. [CrossRef]
72. Cheng, X.; Zhao, Y.; Jiang, Q.; Yang, J.; Zhao, W.; Taylor, I.A.; Peng, Y.L.; Wang, D.; Liu, J. Structural basis of dimerization and dual W-box DNA recognition by rice WRKY domain. *Nucleic Acids Res.* **2019**, *47*, 4308–4318. [CrossRef]
73. Eulgem, T.; Rushton, P.J.; Robatzek, S.; Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **2000**, *5*, 199–206. [CrossRef]
74. Ciolkowski, I.; Wanke, D.; Birkenbihl, R.P.; Somssich, I.E. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol. Biol.* **2008**, *68*, 81–92. [CrossRef] [PubMed]
75. Shen, Q.H.; Saijo, Y.; Mauch, S.; Biskup, C.; Bieri, S.; Keller, B.; Seki, H.; Ulker, B.; Somssich, I.E.; Schulze-Lefert, P. Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science* **2007**, *315*, 1098–1103. [CrossRef] [PubMed]

76. Yanagisawa, S. Dof DNA-binding domains of plant transcription factors contribute to multiple protein-protein interactions. *Eur. J. Biochem.* **1997**, *250*, 403–410. [CrossRef] [PubMed]
77. Xu, X.; Jiang, Q.; Ma, X.; Ying, Q.; Shen, B.; Qian, Y.; Song, H.; Wang, H. Deep sequencing identifies tissue-specific microRNAs and their target genes involving in the biosynthesis of tanshinones in Salvia miltiorrhiza. *PLoS ONE* **2014**, *9*, e111679. [CrossRef]
78. Khan, S.; Qureshi, M.I.; Kamaluddin; Alam, T.; Abdin, M.Z. Protocol for isolation of genomic DNA from dry and fresh roots of medicinal plants suitable for RAPD and restriction digestion. *Afr. J. Biotechnol.* **2007**, *6*, 175–178. [CrossRef]
79. Majewska, M.; Wysokińska, H.; Kuźma, Ł.; Szymczyk, P. Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: A comprehensive overview. *Gene* **2018**, *644*, 38–48. [CrossRef]
80. Shahmuradov, I.A.; Gammerman, A.J.; Hancock, J.M.; Bramley, P.M.; Solovyev, V.V. PlantProm: A database of plant promoter sequences. *Nucleic Acids Res.* **2003**, *31*, 114–117. [CrossRef]
81. Chow, C.N.; Zheng, H.Q.; Wu, N.Y.; Chien, C.H.; Huang, H.D.; Lee, T.Y.; Chiang-Hsieh, Y.F.; Hou, P.F.; Yang, T.Y.; Chang, W.C. PlantPAN 2.0: An update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* **2016**, *44*, D1154–D1160. [CrossRef]
82. Griffiths-Jones, S.; Grocock, R.J.; van Dongen, S.; Bateman, A.; Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **2006**, *34*, D140–D144. [CrossRef]
83. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
84. Toufighi, K.; Brady, S.M.; Austin, R.; Ly, E.; Provart, N.J. The Botany Array Resource: E-Northerns, Expression Angling, and promoter analyses. *Plant J.* **2005**, *43*, 153–163. [CrossRef] [PubMed]
85. Zhang, H.; Zhang, F.; Yu, Y.; Feng, L.; Jia, J.; Liu, B.; Li, B.; Guo, H.; Zhai, J. A Comprehensive Online Database for Exploring ~20,000 Public Arabidopsis RNA-Seq Libraries. *Mol. Plant* **2020**, *13*, 1231–1233. [CrossRef] [PubMed]
86. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]
87. Oughtred, R.; Rust, J.; Chang, C.; Breitkreutz, B.J.; Stark, C.; Willems, A.; Boucher, L.; Leung, G.; Kolas, N.; Zhang, F.; et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **2021**, *30*, 187–200. [CrossRef]
88. Chow, C.N.; Lee, T.Y.; Hung, Y.C.; Li, G.Z.; Tseng, K.C.; Liu, Y.H.; Kuo, P.L.; Zheng, H.Q.; Chang, W.C. PlantPAN3.0: A new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res.* **2019**, *47*, D1155–D1163. [CrossRef]
89. Luján, M.A.; Soria-García, Á.; Claver, A.; Lorente, P.; Rubio, M.C.; Picorel, R.; Alfonso, M. Different Cis-Regulatory Elements Control the Tissue-Specific Contribution of Plastid ω-3 Desaturases to Wounding and Hormone Responses. *Front. Plant Sci.* **2021**, *12*, 727292. [CrossRef]
90. Yuan, L.; Perry, S.E. *Plant Transcription Factors*, 1st ed.; Humana Press: Totowa, NJ, USA, 2011; pp. 45–66.
91. Zhang, X.; Ge, F.; Deng, B.; Shah, T.; Huang, Z.; Liu, D.; Chen, C. Molecular Cloning and Characterization of PnbHLH1 Transcription Factor in Panax notoginseng. *Molecules* **2017**, *22*, 1268. [CrossRef]
92. Chen, R.; Cao, Y.; Wang, W.; Li, Y.; Wang, D.; Wang, S.; Cao, X. Transcription factor SmSPL7 promotes anthocyanin accumulation and negatively regulates phenolic acid biosynthesis in Salvia miltiorrhiza. *Plant Sci.* **2021**, *310*, 110993. [CrossRef]
93. Chen, X.; Bhadauria, V.; Ma, B. ChIP-Seq: A Powerful Tool for Studying Protein-DNA Interactions in Plants. *Curr. Issues Mol. Biol.* **2018**, *27*, 171–180. [CrossRef]
94. Chen, H.; Yang, Q.; Fu, H.; Chen, K.; Zhao, S.; Zhang, C.; Cai, T.; Wang, L.; Lu, W.; Dang, H.; et al. Identification of Key Gene Networks and Deciphering Transcriptional Regulators Associated With Peanut Embryo Abortion Mediated by Calcium Deficiency. *Front. Plant Sci.* **2022**, *13*, 814015. [CrossRef]
95. Lei, X.; Liu, Z.; Xie, Q.; Fang, J.; Wang, C.; Li, J.; Wang, C.; Gao, C. Construction of two regulatory networks related to salt stress and lignocellulosic synthesis under salt stress based on a Populus davidiana × P. bolleana transcriptome analysis. *Plant Mol. Biol.* **2022**. [CrossRef] [PubMed]

Supplementary materials

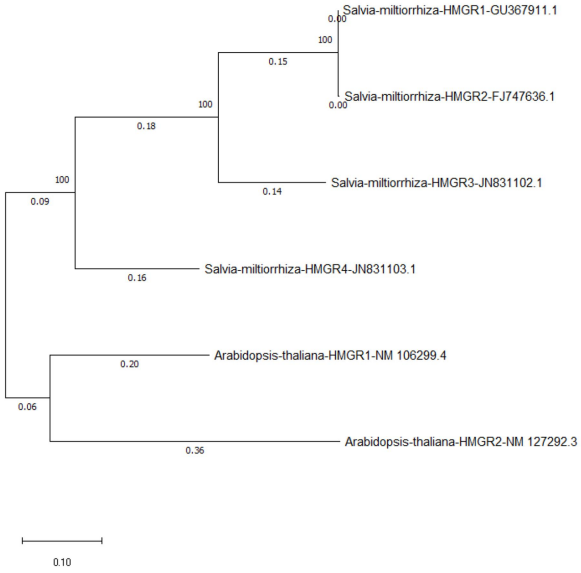


Figure S1. Unrooted dendrogram of *HMGR* sequences from *S. miltiorrhiza* and *A. thaliana* constructed by Maximum Likelihood method and bootstrap test with 1,000 replicates using the MEGA X program.

Table S1. TFs responsive to light, hormone, wounding and pathogen stimulation found in the entire *S. miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 database.

TF family name	TF name
light-responsive TFs	
bHLH	BHLH63, PIF1
bZIP	BZIP54
Dehydrin	COR15A
Dof	DOF1.5, DOF1.10, DOF2.5, DOF3.3, DOF5.2
GATA	GATA1, GATA2, GATA3, GATA4, GATA5, GATA8, GATA9, GATA21, GATA22, GATA23
Homeodomain; bZIP; HD-ZIP	ATHB-1
Homeodomain; HD-ZIP	ATHB-2, ATHB-4, ATHB-16
MYB-related/Myb/SANT	RVE7
NAC; NAM	NAC081
auxin-responsive TFs	
bHLH	BEE1, BEE3
Dof	DOF1.1, DOF3.4, DOF3.6
Homeodomain; bZIP; HD-ZIP	ATHB-15
Homeodomain; HB-PHD	PRH
Homeodomain; HD-ZIP	ATHB-2, ATHB-20
MADS box ;MIKC	AGL14, AGL15
MYB-related	CCA1, RVE4, RVE5, RVE8
Myb/SANT; MYB	MYB6
NAC; NAM	NAC030, NAC101
WRKY	WRKY23
SA-responsive TFs	
bHLH	BHLH66
CAMTA	CAMTA2, CAMTA4, CAMTA6
CSD	CSP2
Dof	DOF1.1, DOF3.4, DOF3.6
MYB	MYB46
MYB-related	CCA1, RVE4, RVE5, RVE6, RVE8
Myb/SANT	MYB3, MYB6
NAC; NAM	NAC062, NAC081
WRKY	WRKY3, WRKY4, WRKY6, WRKY7, WRKY8, WRKY15, WRKY18, WRKY21, WRKY26, WRKY30, WRKY40, WRKY53, WRKY54, WRKY60, WRKY70
brassinosteroid-responsive TFs	
AP2; B3; RAV	RAV1
bHLH	BEE1, BEE3
NAC; NAM	NAC030, NAC081, NAC101
ethylene-responsive TFs	
AP2; ERF	RAP2-2
bHLH	BEE1, BEE3, BHLH66
C2H2	AZF2, AZF3
CAMTA	CAMTA1, CAMTA4
CG-1; CAMTA	CAMTA3
CSD	CSP2
MYB-related	CCA1, RVE4, RVE5, RVE6, RVE8
MYB-related/Myb/SANT	RVE7
Myb/SANT;ARR-B	ARR2
Myb/SANT; MYB	MYB6
NAC; NAM	NAC029
WRKY	WRKY4, WRKY6
ABA-responsive TFs	

bHLH	BEE1, BEE3, MYC2, SCRM
bZIP	ABF3, ABF4
C2H2	AZF2, AZF, ZAT10
CG-1; CAMTA	CAMTA2, CAMTA4, CAMTA6
CSD	CSP2
Dehydrin	COR15A
Homeodomain; HD-ZIP	HAT22
MYB-related	CCA1, RVE4, RVE5, RVE6, RVE8
Myb/SANT	MYB7
Myb/SANT; MYB	MYB3
NAC; NAM	NAC019, NAC029, NAC032, NAC055, NAC062, NAC072
WRKY	WRKY8
ZF-HD	ZHD5
MeJa-responsive TFs	
AP2; ERF	RAP2-3
bHLH	MYC2
CAMTA	CAMTA1, CAMTA4
CG-1; CAMTA	CAMTA3
NAC; NAM	NAC081
wounding-responsive TFs	
C2H2	ZAT10
CAMTA	CAMTA1, CAMTA4, CAMTA6
CG-1; CAMTA	CAMTA2, CAMTA3
NAC; NAM	NAC002, NAC081
WRKY	WRKY6, WRKY8, WRKY20, WRKY40
pathogen-responsive TFs	
CG-1; CAMTA	CAMTA3
Homeodomain; HB-PHD	PRH
Myb/SANT; MYB	MYB6
NAC; NAM	NAC081, NAC091
SBP	SPL14
WRKY	WRKY4, WRKY6, WRKY8, WRKY11, WRKY17, WRKY18, WRKY23, WRKY27, WRKY33, WRKY38, WRKY48, WRKY53, WRKY60

Table S2. TFs responsive to major abiotic and biotic factors found in the proximal *S. miltiorrhiza HMGR4* promoter using the PlantPan 2.0 database.

TF family name	TF name
light-responsive TFs	
bHLH	MYC2
Dehydrin	COR15A
Dof	DOF1.5, DOF2.5
GATA	GATA1, GATA2, GATA3, GATA4, GATA5, GATA8, GATA9, GATA12, GATA21, GATA22, GATA23
MADF;Trihelix	GT-1
MYB-related/Myb/SANT	RVE7
NF-YA	NFYA5
NF-YB	NFYB9
SA-responsive TFs	
bZIP	TGA1, TGA2, TGA3, TGA5, TGA7
Dof	DOF1.1, DOF3.4, DOF3.6
MYB	MYB46
MYB-related	RVE4, RVE6, RVE8
WRKY	WRKY3, WRKY4, WRKY6, WRKY7, WRKY8 WRKY15, WRKY18, WRKY21, WRKY26, WRKY30, WRKY38, WRKY40, WRKY53, WRKY54, WRKY60, WRKY70
bacterium-responsive TFs	
bZIP	TGA3, TGA7
WRKY	WRKY8, WRKY11, WRKY17, WRKY18, WRKY27, WRKY33, WRKY38, WRKY40, WRKY48, WRKY52, WRKY53, WRKY60, WRKY70
auxin-responsive TFs	
bZIP	TGA1, TGA2, TGA3, TGA5, TGA7
C2H2	AZF2
Dof	DOF1.1, DOF3.4, DOF3.6
Homeodomain; HB-PHD	PRH
MADS box; MIKC	AGL14, AGL15
MYB-related	RVE4, RVE5, RVE8
WRKY	WRKY23
ABA-responsive TFs	
bHLH	SCRM, MYC2
C2H2	AZF2, AZF3, ZAT10
Dehydrin	COR15A
MYB-related	RVE4, RVE5, RVE6, RVE8
NF-YA	NFYA5
NF-YB	NFYB6, NFYB9
NF-YC	NFYC3
WRKY	WRKY8, WRKY25, WRKY33
gibberellin-responsive TFs	
MADS box	AGL42, AGL71, AGL72
MYB-related	RVE4, RVE5, RVE6, RVE8
WRKY	WRKY27
chitin-responsive TFs	
C2H2	ZAT10
Dof	DOF1.7
WRKY	WRKY6, WRKY11, WRKY17, WRKY18, WRKY22, WRKY33, WRKY40, WRKY46, WRKY48, WRKY53, WRKY70
cold-responsive TFs	
bHLH	SCRM
C2H2	AZF2, AZF3, ZAT6, ZAT10
Dehydrin	COR15A

Dof	DOF2.5
MADS box; MIKC	AGL19
WRKY	WRKY25, WRKY33, WRKY34
	salt stress-responsive TFs
bHLH	MYC2, SCRM
C2H2	AZF3, ZAT10
Dehydrin	COR15A
MYB-related	RVE3, RVE4, RVE5, RVE6, RVE8
MYB-related/Myb/SANT	RVE7
WRKY	WRKY8, WRKY25, WRKY33, WRKY46, WRKY57

Table S3. TF genes co-expressed with *A. thaliana HMGR1* found with the Expression Angler tool.

TF locus name	TF gene name	*r*-value [a]
AtGenExpress Hormone and Chemical Compendium		
At5g17300	*RVE1*	0.501
At1g06850	*AtbZIP52*	0.501
At2g01760	*ARR14*	0.509
At1g33240	*GTL1*	0.509
At1g19700	*BLH10*	0.516
At1g64380	*ERF061*	0.516
At5g44190	*GLK2*	0.520
At5g54630	*MRB17.13*	0.521
At1g06040	*BBX24*	0.522
At5g25810	*TINY*	0.527
At3g23050	*IAA7*	0.538
At5g13770	Pentatricopeptide repeat (PPR-like) superfamily protein	0.544
At1g22590	*AGL87*	0.547
At1g69690	*TCP15*	0.549
At2g27050	*EIL1*	0.554
At3g06590	*BHLH148*	0.561
At4g40060	*ATHB-16*	0.562
At5g57660	*COL5*	0.563
At4g36540	*BEE2*	0.566
At4g00150	*SCL6*	0.567
At5g02840	*RVE4*	0.573
At1g30650	*WRKY14*	0.576
At1g13300	*HRS1*	0.578
At3g07650	*COL9*	0.578
At1g69780	*ATHB-13*	0.594
At2g18160	*BZIP2*	0.597
At4g09460	*MYB6*	0.599
At2g42380	*BZIP34*	0.605
At3g05800	*BHLH150*	0.609
At2g28200	*ZAT5*	0.61
At1g74840	Homeodomain-like superfamily protein	0.61
At3g17100	*BHLH147*	0.618
At1g68520	*BBX14*	0.621
At5g47390	*KUA1*	0.623
At3g60490	*ERF035*	0.63
At3g62420	*BZIP53*	0.635
At3g58120	*BZIP61*	0.64
At2g23760	*BLH4*	0.648
At3g48360	*BT2*	0.659
At3g47620	*TCP14*	0.704
At5g60850	*DOF5.4*	0.727
AtGenExpress Abiotic Stress Compendium		
At3g16770	*RAP2-3*	0.502
At5g46690	*BHLH071*	0.502
At1g68190	*BBX27*	0.502
At3g47620	*TCP14*	0.502
At1g69690	*TCP15*	0.503
At2g20570	*GLK1*	0.504
At3g12730	Homeodomain-like superfamily protein	0.506
At3g61150	*HDG1*	0.507

At5g67180	*TOE3*	0.508
At5g44190	*GLK2*	0.512
At2g42300	*BHLH48*	0.518
At5g38140	*NFYC12*	0.523
At4g00730	*ANL2*	0.524
At2g46530	*ARF11*	0.524
At2g02080	*IDD4*	0.525
At3g48590	*NFYC1*	0.527
At4g36870	*BLH2*	0.53
At2g01760	*ARR14*	0.53
At3g17100	*BHLH147*	0.537
At2g22540	*SVP*	0.537
At5g02840	*RVE4*	0.545
At1g72740	Homeodomain-like/winged-helix DNA-binding family protein	0.545
At3g57800	*BHLH60*	0.545
At5g08330	*TCP21*	0.546
At5g05550	*ENAP2*	0.552
At5g24930	*COL4*	0.557
At1g10610	*BHLH90*	0.558
At2g33500	*BBX12*	0.575
At5g08520	*MYBS2*	0.575
At1g54060	*ASIL1*	0.579
At4g21750	*ATML1*	0.583
At4g04890	*PDF2*	0.587
At2g33810	*SPL3*	0.596
At3g19860	*BHLH121*	0.611
At2g43010	*PIF4*	0.625
At1g14920	*GAI*	0.627
At5g62000	*ARF2*	0.63
	AtGenExpress Pathogen compendium	
At3g19580	*AZF2*	0.503
At1g69310	*WRKY57*	0.504
At1g02220	*NAC003*	0.505
At1g07530	*SCL14*	0.516
At1g19850	*ARF5*	0.515
At3g01970	*WRKY45*	0.515
At4g39100	*SHL*	0.52
At3g19860	*BHLH121*	0.521
At1g62990	*KNAT7*	0.522
At1g73730	*EIL3*	0.525
At5g52510	*SCL8*	0.526
At2g36080	*ABS2*	0.53
At1g79180	*MYB63*	0.534
At1g48000	*MYB112*	0.535
At2g21240	*BPC4*	0.536
At3g61890	*ATHB-12*	0.536
At5g49700	*AHL17*	0.537
At3g06490	*MYB108*	0.538
At3g51960	*BZIP24*	0.543
At4g27410	*NAC072*	0.543
At2g28200	*ZAT5*	0.547
At1g56010	*NAC021*	0.554
At1g06180	*MYB13*	0.557
At3g17100	*BHLH147*	0.559

At3g58710	*WRKY69*	0.564
At4g31420	*REIL1*	0.565
At2g46680	*ATHB-7*	0.571
At3g20770	*EIN3*	0.572
At5g13330	*RAP2.6L*	0.573
At5g24800	*BZIP9*	0.59
At3g15500	*NAC055*	0.594
At1g10170	*NFXL1*	0.599
At2g38340	*DREB2E*	0.611
At5g39610	*NAC92*	0.643
	AtGenExpress Tissue Compendium	
At1g77980	*AGL66*	0.506
At3g15540	*IAA19*	0.512
At2g40620	*BZIP18*	0.517
At1g60240	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	0.52
At2g03060	*AGL30*	0.521
At3g01470	*HAT5*	0.521
At2g32460	*MYB101*	0.521
At1g72740	Homeodomain-like/winged-helix DNA-binding family protein	0.525
At3g10470	C2H2-type zinc finger family protein	0.534
At1g18750	*AGL65*	0.535
At3g16350	Homeodomain-like superfamily protein	0.539
At2g42380	*BZIP34*	0.546
At4g08250	*SCL26*	0.548
At4g16110	*ARR2*	0.552
At2g47810	*NFYB5*	0.558
At4g14410	*BHLH104*	0.563
At5g56270	*WRKY2*	0.575
At4g35700	*DAZ3*	0.579
At1g35490	bZIP family transcription factor	0.582
At5g45710	*HSFA4C*	0.584
At4g20380	*LSD1*	0.598
At4g31420	*REIL1*	0.609
At1g53320	*TLP7*	0.64
At3g54620	*BZIP25*	0.666
At1g50640	*ERF3*	0.721
	AtGenExpressPlus Extended Tissue Compendium	
At4g26930	*MYB97*	0.505
At1g18750	*AGL65*	0.51
At4g35700	*DAZ3*	0.512
At2g23340	*DEAR3*	0.516
At2g01930	*BPC1*	0.519
At5g67580	*TRB2*	0.525
At2g34440	*AGL29*	0.53
At3g20310	*ERF7*	0.534
At2g32460	*MYB101*	0.545
At2g03060	*AGL30*	0.547
At1g35490	bZIP family transcription factor	0.553
At5g56270	*WRKY2*	0.557
At1g34190	*NAC017*	0.558
At3g57390	*AGL18*	0.567
At2g42380	*BZIP34*	0.572
At3g20770	*EIN3*	0.572
At2g40620	*BZIP18*	0.572

At3g16350	Homeodomain-like superfamily protein	0.601
At4g37180	*HHO5*	0.601
At5g46910	*JMJ13*	0.604
At4g16110	*ARR2*	0.611
At5g54680	*BHLH105*	0.64
At4g13980	*HSFA5*	0.644
At1g55520	*TBP2*	0.645
At4g31420	*REIL1*	0.661
At4g20380	*LSD1*	0.665
At1g50640	*ERF3*	0.666
At1g53320	*TLP7*	0.667
At5g45710	*HSFA4C*	0.790

[a] The *r*-value was 0.5 - 1.0.

Table S4. Primers used in the study.

Primer name	Sequence
Genome Walker gene-specific primers (GSPs)	
GSP1	GGTGAAGAAAATGCCGTTGGTTAGGAAGA
GSP2	TGTGGCGGAGTGAGGCGACGGCGGTTT
Sanger sequencing of *HMGR4* promoter	
M13_F	GTAAAACGACGGCCAG
M13_R	CAGGAAACAGCTATGAC
HMGR4_F	TTTGTGCAGTGCGAACCAACCAA
HMGR4_R	GCATGACATAGTTCTTTTAGG

File S1 i **File S2** https://www.mdpi.com/article/10.3390/plants11141861/s1

„The expression profiles of the *Salvia miltiorrhiza 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4* gene and its influence on the biosynthesis of tanshinones" (Publikacja III)

molecules

MDPI

Article

The Expression Profiles of the *Salvia miltiorrhiza 3-Hydroxy-3-methylglutaryl-coenzyme A Reductase 4* Gene and Its Influence on the Biosynthesis of Tanshinones

Małgorzata Majewska [1,*], Piotr Szymczyk [1], Jan Gomulski [1], Agnieszka Jeleń [2], Renata Grąbkowska [1], Ewa Balcerczak [2] and Łukasz Kuźma [1,*]

[1] Department of Biology and Pharmaceutical Botany, Medical University of Lodz, Muszyńskiego 1, 90-151 Lodz, Poland; piotr.szymczyk@umed.lodz.pl (P.S.); jan.gomulski@umed.lodz.pl (J.G.); renata.grabkowska@umed.lodz.pl (R.G.)

[2] Laboratory of Molecular Diagnostics and Pharmacogenomics, Department of Pharmaceutical Biochemistry and Molecular Diagnostics, Medical University of Lodz, Muszyńskiego 1, 90-151 Lodz, Poland; agnieszka.jelen@umed.lodz.pl (A.J.); ewa.balcerczak@umed.lodz.pl (E.B.)

* Correspondence: malgorzata.majewska1@stud.umed.lodz.pl (M.M.); lukasz.kuzma@umed.lodz.pl (Ł.K.)

Abstract: *Salvia miltiorrhiza* is a medicinal plant that synthesises biologically-active tanshinones with numerous therapeutic properties. An important rate-limiting enzyme in the biosynthesis of their precursors is 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR). This study presents the organ-specific expression profile of the *S. miltiorrhiza HMGR4* gene and its sensitivity to potential regulators, viz. gibberellic acid (GA_3), indole-3-acetic acid (IAA) and salicylic acid (SA). In addition, it demonstrates the importance of the *HMGR4* gene, the hormone used, the plant organ, and the culture environment for the biosynthesis of tanshinones. *HMGR4* overexpression was found to significantly boost the accumulation of dihydrotanshinone I (DHTI), cryptotanshinone (CT), tanshinone I (TI) and tanshinone IIA (TIIA) in roots by 0.44 to 5.39 mg/g dry weight (DW), as well as TIIA in stems and leaves. *S. miltiorrhiza* roots cultivated in soil demonstrated higher concentrations of the examined metabolites than those grown in vitro. GA_3 caused a considerable increase in the quantity of CT (by 794.2 µg/g DW) and TIIA (by 88.1 µg/g DW) in roots. In turn, IAA significantly inhibited the biosynthesis of the studied tanshinones in root material.

Keywords: *Salvia miltiorrhiza*; *HMGR4*; expression; overexpression; tanshinone; GA_3; IAA; SA

check for updates

Citation: Majewska, M.; Szymczyk, P.; Gomulski, J.; Jeleń, A.; Grąbkowska, R.; Balcerczak, E.; Kuźma, Ł. The Expression Profiles of the *Salvia miltiorrhiza 3-Hydroxy-3-methylglutaryl-coenzyme A Reductase 4* Gene and Its Influence on the Biosynthesis of Tanshinones. *Molecules* **2022**, *27*, 4354. https://doi.org/10.3390/molecules27144354

Academic Editors: Raffaele Capasso and Enrico Millo

Received: 4 May 2022
Accepted: 29 June 2022
Published: 7 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Salvia miltiorrhiza Bunge, also known as Chinese sage or Red sage, is one of the basic elements of traditional Chinese medicine used in treating diverse conditions, such as cardiovascular diseases, menstrual disorders and insomnia [1,2]. The medical properties of this plant result from the biosynthesis of various bioactive compounds, including tanshinones. Recent research indicates that tanshinones provide cardiovascular protection [3], regulate metabolic functions [4], and possess a range of anticancer [5], neuroprotective [6], anti-inflammatory [7], antioxidant [8], phytoestrogenic [9], antiosteoporotic [10], antibacterial [11] and anti-aggregation [12] properties.

Among the several dozen tanshinones isolated from *S. miltiorrhiza* so far [13], the most studied are dihydrotanshinone I (DHTI), cryptotanshinone (CT), tanshinone I (TI) and tanshinone IIA (TIIA). The starting point for the production of the tanshinone diterpene backbone is the synthesis of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) precursors through the mevalonate (MVA) and methylerythritol phosphate (MEP) pathways [14]. The key rate-limiting enzyme in the MVA pathway, catalysing the conversion of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) to MVA, is HMG-CoA reductase (HMGR) [14]. A recent search found five sequences of *S. miltiorrhiza HMGR* genes

(*HMGR* to *HMGR4*) currently deposited in the GenBank database [15–17]. Among them is *HMGR4*, which, unlike other genes, has not been extensively studied, and its importance in the biosynthesis of tanshinones has not been investigated.

Whereas the annual consumption of *S. miltiorrhiza* in China exceeds 16 million kg, the low concentration of tanshinones in plant material and limited arable land available for cultivation, make the meeting of growing demand more and more problematic [14]. Therefore, alternative sources of *S. miltiorrhiza* plant material, such as hairy roots or cell and callus cultures are under development [18]. These sources offer continuous biosynthesis of tanshinones, and their concentration may be boosted by the application of chemical or physical components known as elicitors [19–21].

Gibberellic acid (GA_3) regulates vegetative and reproductive growth by triggering the degradation of DELLA proteins, these being master repressors of its signaling [22,23]. Initially, a bioactive hormone binds to the gibberellin-insensitive DWARF 1 (GID1) receptor and induces a conformational change in its N-terminal fragment, enabling DELLA binding. The DELLA proteins are then polyubiquitinated by E3 ubiquitin ligases such as SLEEPY1 and constitutively photomorphogenic 1 (COP1), and directed towards the destruction in 26S proteasome [24]. Removing the DELLA proteins releases repressed transcription factors (TFs), enabling gene expression regulation. Dominant TFs regulated by GA_3 are GAI-RGA-SCR (GRAS) proteins [25]. In *S. miltiorrhiza* most of the 35 identified GRAS TFs are induced by GA_3 [25]. It has been proven that the overexpression of *GRAS1* and *GRAS2* in *S. miltiorrhiza* hairy roots increases the accumulation of DHTI, CT, TI and TIIA [26].

In the auxin signaling pathway, the transcription of effector genes is controlled by the interaction of auxin/indole-3-acetic acid (Aux/IAA) repressor with transport inhibitor resistant 1/auxin signaling F-box (TIR1/AFB) proteins [27]. Ubiquitination of these complexes by suppressor of kinetochore protein 1 (SKP1)/cullin1/F-box (SCF) E3 ubiquitin ligase complex and subsequent proteasome-dependent degradation, enable auxin response factors (ARFs) to regulate gene transcription [28]. The proper initiation of gene expression usually requires dimerization of ARFs, that bind to closely-located TGTCGG inverted repeats and TGTCTC or TGTCGG direct repeats [29]. The complexity of auxin-dependent gene regulation in *S. miltiorrhiza* is increased by the fact that most of the 25 studied ARFs have an inhibitory effect on the transcription rate [30].

In the salicylic acid (SA) signaling route, the non-expressor of pathogenesis-related genes 1 (NPR1) acts as the master regulator of the plant response [31]. In the absence of SA, the N-terminal BTB domain of NPR1 interacts with the C-terminal transactivation domain to inhibit NPR1 transcription activity [32]. NPR1 is activated through copper-dependent binding of SA [32]. This protein lacks its own DNA-binding domain and expresses its transactivatory function through interaction with bZIP family TFs [33]. Such SA-responsive TGACG transcription factor binding sites (TFBSs) have been found in numerous plant promoters [34–36]. Another group of TFs controlled by SA are WRKY; among these, WRKY1 strongly induces the genes of the tanshinone biosynthesis pathway through interaction with the W-box (T)TGAC(C/T) element [37].

This work examines the organ-specific expression pattern of the *S. miltiorrhiza HMGR4* gene and the influence of selected phytohormones (GA_3, indole-3-acetic acid (IAA), SA) on its transcription level. These experiments were carried out on wild, in vitro-grown plants. In silico analysis of the *S. miltiorrhiza HMGR4* promoter performed with PlantPan 2.0 tool was used to select the appropriate hormones. Moreover, this study investigates the importance of the *HMGR4* gene, the hormone used, the plant organ and the growth environment for the biosynthesis of DHTI, CT, TI, TIIA and total tanshinone using transgenic *S. miltiorrhiza* plants grown in vitro and in soil. The use of hormones is aimed at modulating the *HMGR4* gene expression and thus obtaining information on the presumed role of this gene in the biosynthesis of tanshinones, as well as influencing their content.

2. Results

2.1. Organ-Specific Expression of S. miltiorrhiza HMGR4 Gene

Real-time qPCR results showed that *HMGR4* gene was expressed in all analysed *S. miltiorrhiza* organs, but with different intensities. The leaves and stems demonstrated higher levels of the *HMGR4* transcript than the reference, with R = 1.14 ± 0.08 and R = 1.05 ± 0.01, respectively; in roots, the level was lower than in the reference with R = 0.95 ± 0.07.

Due to their level of transcript and high availability of material for research, it was decided to use the leaves to study the effect of hormones on *S. miltiorrhiza HMGR4* activity.

2.2. Potential Regulators of S. miltiorrhiza HMGR4 Gene Expression

Sequence analysis of the *S. miltiorrhiza HMGR4* promoter using the PlantPan 2.0 tool showed the existence of 5369 potential TFBSs and 365 interacting TFs previously detected in the *Arabidopsis thaliana* model plant. The similarity score between the TFBSs found in *HMGR4* promoter and those identified in *A. thaliana* was set to 0.7–1.0. Of all the TFs detected, a large group was able to respond to hormonal agents; many of these were sensitive to GA_3, IAA and SA (Table 1, Tables S1–S3). It is worth emphasising that these TFs also had potential binding sites in the *HMGR4* proximal promoter region (Tables S1–S3), where most functional TFBSs are believed to be located [38,39]. Therefore, it was decided to investigate the importance of these hormones on the expression of *S. miltiorrhiza HMGR4*.

Table 1. Potential transcription factors (TFs) that bind to the *Salvia miltiorrhiza HMGR4* promoter sequence and respond to gibberellic acid (GA_3), indole-3-acetic acid (IAA), salicylic acid (SA) signals found using PlantPan 2.0 tool.

TF Family Name	TF Gene Name and Locus	TFBSs Number
	GA_3	
AT-Hook	*AHL25*; At4g35390	6
MYB-related	*CCA1*; At2g46830	1
	RVE8; At3g09600	2
	RVE4; At5g02840	
Homeodomain; HD-ZIP	*ATHB-23*; At1g26960	1
bHLH	*PIF3*; At1g09530	2
GATA	*GATA22*; At4g26150	62
	GATA21; At5g56860	
MADS box; MIKC	*SOC1*; At2g45660	8
	AGL24; At4g24540	15
MADS box	*AGL71*; At5g51870	13
	AGL72; At5g51860	14
	AGL42; At5g62165	
NF-YC	*NFYC3*; At1g54830	48
	IAA	
MYB-related	*CCA1*; At2g46830	1
	RVE 8; At3g09600	2
Homeodomain; HD-ZIP	*HAT2*; At5g47370	2
	ATHB-20; At3g01220	3
CAMTA	*CAMTA1*; At5g09410	2
MADS box; MIKC	*AGL14*; At4g11880	17

Table 1. *Cont.*

TF Family Name	TF Gene Name and Locus	TFBSs Number
	SA	
NAC; NAM	*NAC081*; At5g08790	1
	NAC062; At3g49530	2
MYB	*MYB46*; At5g12870	4
Myb/SANT; MYB	*MYB3*; At1g22640	1
MYB-related	*CCA1*; At2g46830	1
	RVE8; At3g09600	2
	RVE4; At5g02840	
Dof	*DOF1.1*; At1g07640	34
	DOF3.4; At3g50410	
bZIP	*TGA2*; At5g06950	32
bHLH	*LRL1*; At2g24260	2
CG-1; CAMTA	*CAMTA2*; At5g64220	3
CAMTA	*CAMTA4*; At1g67310	2
	CAMTA6; At3g16940	
CSD	*CSP2*; At4g38680	2
WRKY	*WRKY6*; At1g62300	9
	WRKY40; At1g80840	
	WRKY4; At1g13960	8
	WRKY60; At2g25000	
	WRKY21; At2g30590	
	WRKY54; At2g40750	
	WRKY70; At3g56400	
	WRKY53; At4g23810	
	WRKY18; At4g31800	
	WRKY26; At5g07100	
	WRKY38; At5g22570	
	WRKY30; At5g24110	

2.3. Effect of GA_3, IAA, SA on S. miltiorrhiza HMGR4 Gene Expression

The hormones used in the experiment changed the expression of *HMGR4* in treated leaves compared to control (leaves not incubated with hormones) (Figure 1). At the beginning of each study, a lower *HMGR4* transcript level was observed in the test materials than in corresponding control (R < 1). Treatment with GA_3 or IAA or SA for 12 and 24 h resulted in the stimulation of *HMGR4* expression against untreated samples (R > 1). It is worth noting that the exposure of leaves to GA_3 resulted in an approximately 2.86-fold increase in *HMGR4* expression between 12 and 24 h, and the stimulation effect was also maintained at 48 h. From 48 h, the level of *HMGR4* transcript in the hormone-treated samples decreased compared to the control (R < 1). In the final part of the testing (72 h for SA and 96 h for GA_3 and IAA) the level of *HMGR4* mRNA increased again in leaves incubated with hormones compared to untreated samples (R > 1).

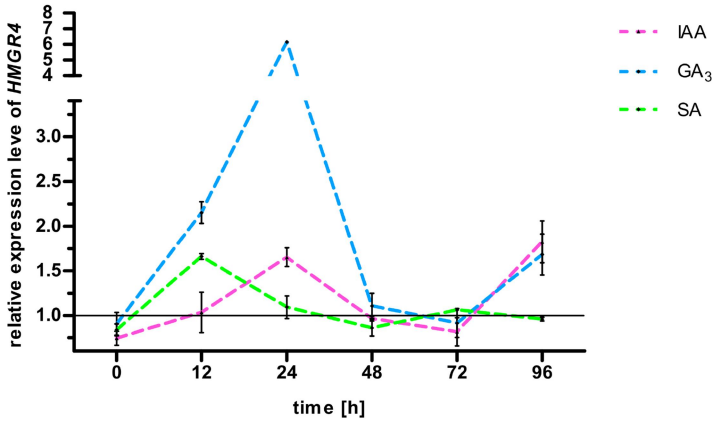


Figure 1. Effect of GA_3, IAA and SA on *HMGR4* gene expression in *S. miltiorrhiza* leaves. The expression was analysed by real-time PCR. The results are presented as mean ± SD.

Based on the obtained results, GA_3 and IAA were selected for experiments determining the tanshinone content.

2.4. Impact of pRI201-AN-HMGR4 Transformation on S. miltiorrhiza HMGR4 Gene Expression

Higher levels of the *HMGR4* transcript were observed in all *S. miltiorrhiza* organs taken from plants transformed with the pRI201-AN-HMGR4 construct compared to control (R > 1). The R values for stems, roots and leaves, were 1.28 ± 0.19, 1.25 ± 0.14 and 1.10 ± 0.14, respectively.

2.5. Influence of HMGR4 Overexpression on the Biosynthesis of Tanshinones in S. miltiorrhiza

The conducted studies indicate that *S. miltiorrhiza HMGR4* overexpression had a significant influence on the quantity of measured tanshinones.

Roots with *HMGR4* overexpression, both soil-grown and cultivated in vitro, demonstrated significantly higher accumulation of DHTI, CT, TI, TIIA and higher total tanshinone content compared to control roots without *HMGR4* overexpression ($p < 0.01$) (Figure 2A–E). The differences described above ranged from 1.51-fold to 2.43-fold, i.e., 0.59–5.39 mg/g dry weight (DW), and from 1.82- to 3.62-fold (0.44–2.40 mg/g DW), respectively, depending on the type of tanshinone.

Clear differences in the levels of individual tanshinones were observed between roots with *HMGR4* overexpression and those without. More specifically, these values were as follows (the first value indicating soil conditions and the second in vitro): 2.43- or 3.62-fold (5.39 or 2.40 mg/g DW) for CT, 2.19- or 2.47-fold (0.59 or 0.44 mg/g DW) for DHTI, 1.86- or 2.21-fold (0.71 or 0.65 mg/g DW) for TI, 1.51- or 1.82-fold (1.88 or 0.55 mg/g DW) for TIIA, and reflected the place of cultivation.

Moreover, overexpression of *HMGR4* gene induced the quantity of TIIA to about 50 μg/g DW in in vitro and in soil-grown stems and leaves (Figure 2D).

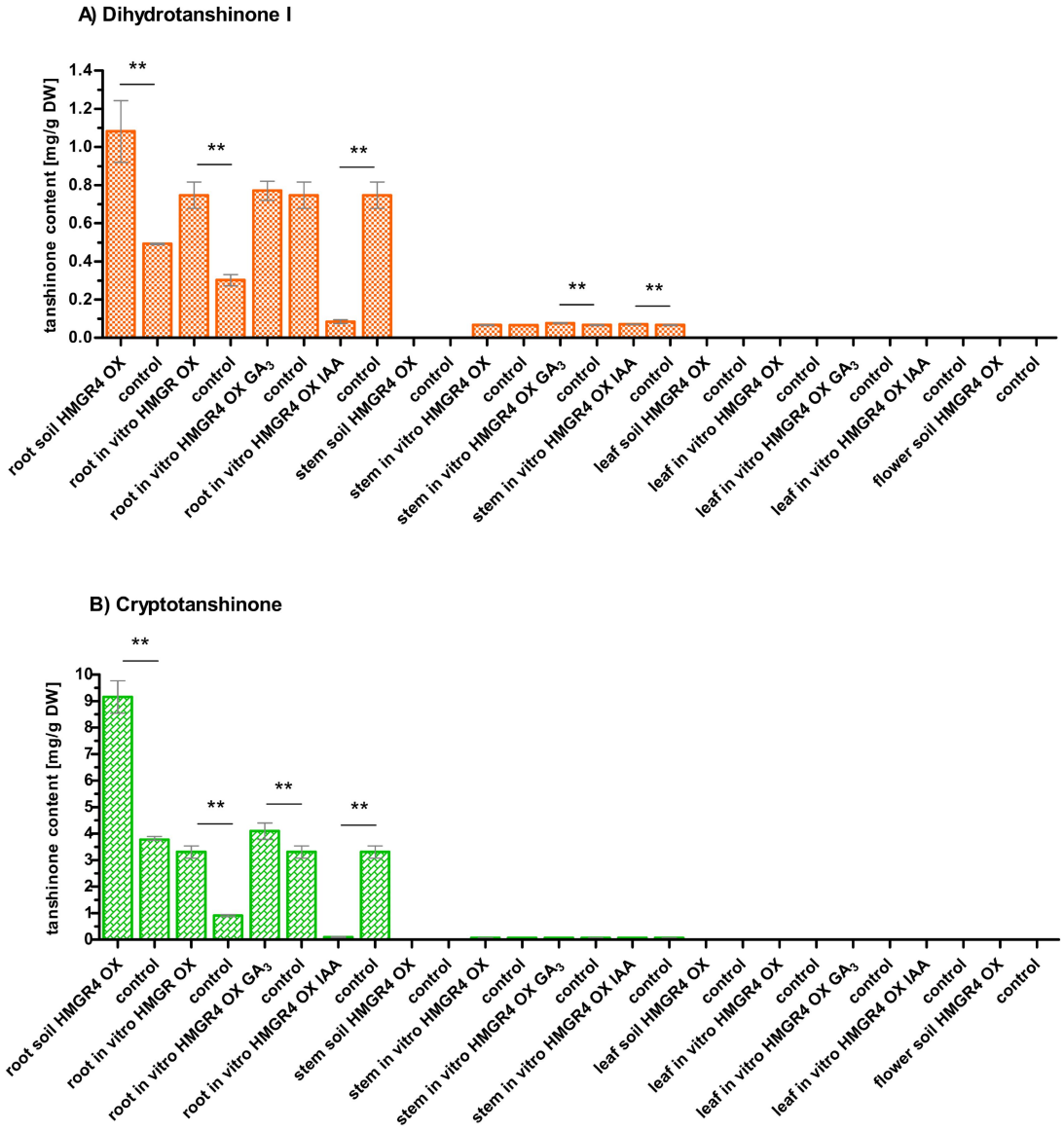


Figure 2. *Cont.*

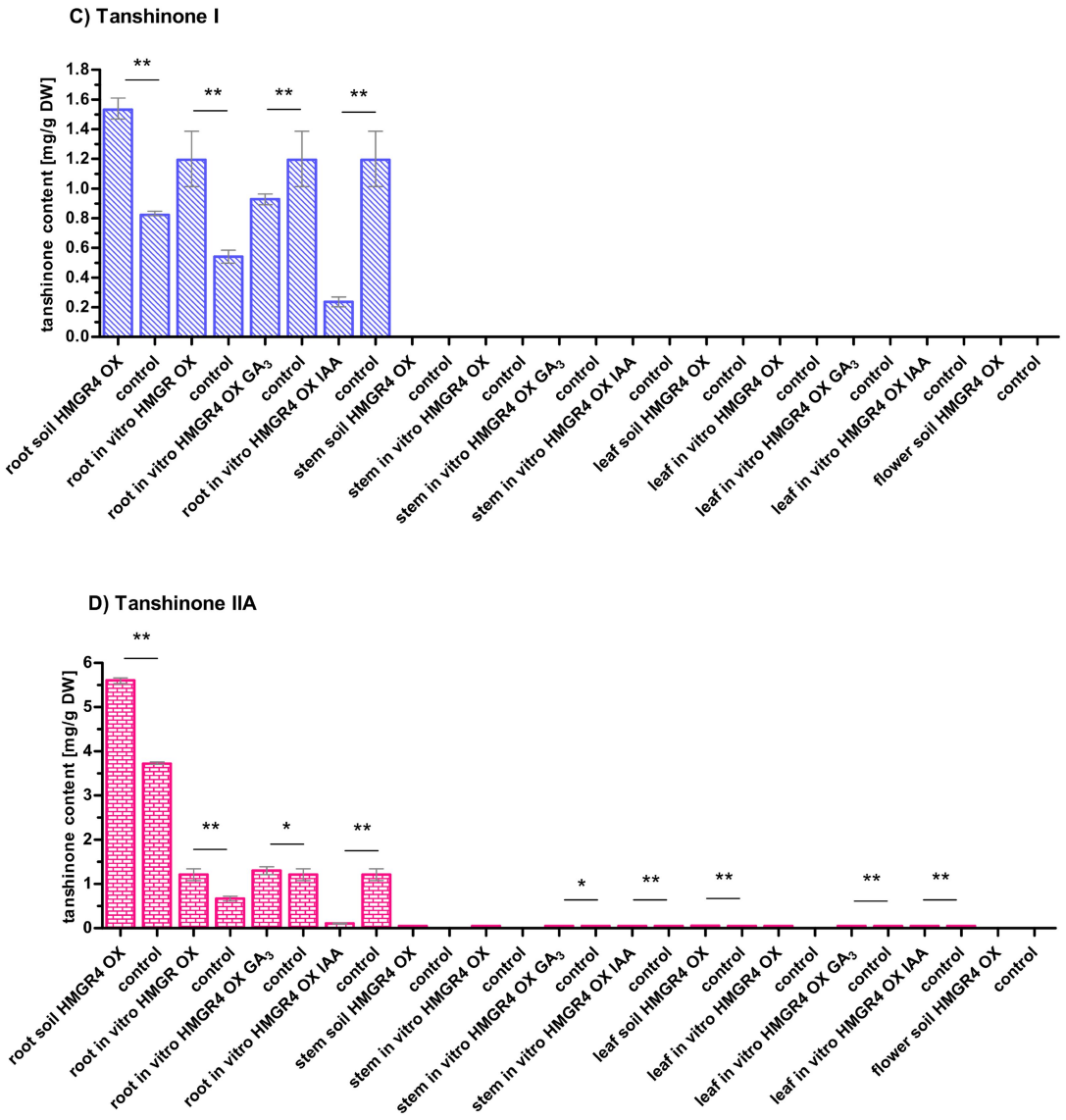


Figure 2. *Cont.*

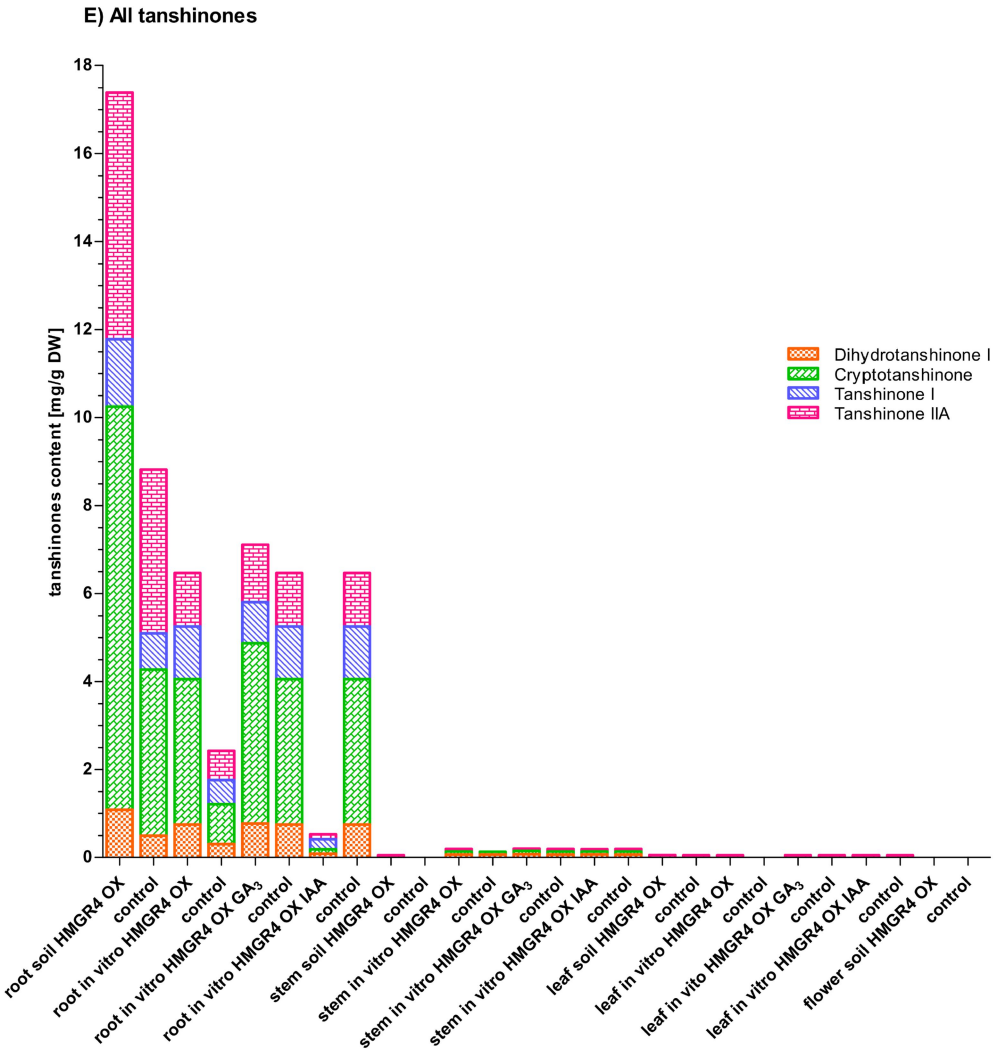


Figure 2. Quantitative UHPLC analysis of the content of individual tanshinones (**A**–**D**) and the total amount of all tested tanshinones (**E**) in extracts obtained from roots, stems, leaves and flowers of *S. miltiorrhiza*. A control is shown for each test sample (non-*HMGR4* overexpressing plant material or, in the case of hormone treatment, non-hormone-treated *HMGR4* overexpressing plant material). Bars are medians with first and third quartile. ** significant difference at $p < 0.01$ compared to control; * significant difference at $0.01 < p < 0.05$ compared to control; DW, dry weight; OX, overexpression.

2.6. Organ-Dependent Accumulation of Tanshinones in S. miltiorrhiza

Roots appeared to be the main site of accumulation of all studied metabolites in *S. miltiorrhiza*. All of the examined roots were found to contain all tested tanshinones (Figure 2A–D). CT was present at the highest levels (0.91–9.17 mg/g DW), while lower amounts were found for TIIA (0.67–5.61 mg/g DW), TI (0.54–1.53 mg/g DW) and DHI (0.30–1.08 mg/g DW) (Figure 2A–D). The quantity of the identified metabolites was highest in soil-grown roots overexpressing *HMGR4*.

Some tanshinones were detected in stems and leaves with median values ranging from 50 to 73.5 μg/g DW (Figure 2A,B,D,E). The most common tanshinone present in the tested stems and leaves was TIIA. The TIIA content was typically 104.5-fold higher (by

5.55 mg/g DW) in roots than in stems or leaves in the soil-grown plants, and 23.4-fold higher (by 1.16 mg/g DW) in in vitro roots (Figure 2D). No tanshinones were detected in flowers (Figure 2A–E). Hence, apart from slight changes in TIIA level in stems and leaves, *HMGR4* overexpression did not appear to significantly change the organ-specific pattern of accumulation of the compounds in *S. miltiorrhiza*.

2.7. Impact of Growth Environment on the Biosynthesis of Tanshinones in S. miltiorrhiza

The soil environment favoured a significantly higher production of all tested tanshinones in the root material compared to in vitro conditions ($p < 0.01$). This was true both in the group of roots with and without *HMGR4* overexpression, and the differences were from 1.45- to 4.62-fold (0.34–5.86 mg/g DW) and from 1.63- to 5.58-fold (0.19–3.06 mg/g DW), respectively, depending on the type of metabolite. The quantities of individual tanshinones varied considerably between the soil-grown roots and those grown in vitro. More specifically, these differences were as follows (first value = *HMGR4* overexpression; the second value = without): 4.62- or 5.58-fold (4.39 or 3.06 mg/g DW) for TIIA, 2.77- or 4.14-fold (5.86 or 2.87 mg/g DW) for CT, 1.45- or 1.63-fold (0.34 or 0.19 mg/g DW) for DHTI, 1.28- or 1.52-fold (0.34 or 0.28 mg/g DW) for TI. It is worth noting that the in vitro roots with *HMGR4* overexpression demonstrated 1.51-fold higher DHTI (0.25 mg/g DW) and 1.45-fold higher TI (0.37 mg/g DW) than the soil-grown roots without overexpression.

In leaf material, the content of TIIA was significantly higher in soil than in in vitro conditions ($p = 0.0000$), amounting to 50 or 56.8 μg/g DW, depending on *HMGR4* overexpression status.

DHTI and CT were detected in stems grown in vitro but not in stems grown in soil (Figure 2A,B). Median levels were 66.8 or 65.9 μg/g DW for DHTI, and 72 or 67.3 μg/g DW for CT, depending on the presence or absence of *HMGR4* overexpression.

2.8. Effect of GA_3 and IAA on the Biosynthesis of Tanshinones in S. miltiorrhiza

The addition of GA_3 to *S. miltiorrhiza* in vitro root culture significantly increased CT, TIIA and total tanshinone levels in comparison to untreated roots ($p = 0.0000$, $p = 0.0404$, $p = 0.0404$, respectively) (Figure 2B,D). The observed increases were 1.24-fold (0.79 mg/g DW) for CT and 1.07-fold (88.1 μg/g DW) for TIIA. Treatment had no effect on DHTI and significantly decreased the amount of TI by 1.29-fold (0.27 mg/g DW) ($p = 0.0000$) (Figure 2A,C).

In vitro cultivation of stems grown in the presence of GA_3 showed a significant 1.15-fold (9.9 μg/g DW) rise in DHTI ($p = 0.0000$) and a significant 1.04-fold (1.9 μg/g DW) reduction in TIIA ($p = 0.0235$) compared to untreated controls (Figure 2A,D). However, in vitro cultivation of leaves with GA_3 resulted in a significant 1.02-fold (1.2 μg/g DW) increase in TIIA compared to control ($p = 0.0000$) (Figure 2D).

The use of IAA resulted in a significant decrease in the content of all tested tanshinones in in vitro root culture compared to untreated roots ($p = 0.0000$) (Figure 2A–D): 34.06-fold (3.21 mg/g DW) for CT, 11.49-fold (1.11 mg/g DW) for TIIA, 8.84-fold (0.66 mg/g DW) for DHTI and 5.05-fold (0.96 mg/g DW) for TI.

IAA treatment only appeared to have a slight influence on the quantity of tanshinones in stems and leaves: a significant 1.05-fold (3.2 μg/g DW) rise in DHTI and a significant 1.05-fold (2.3 μg/g DW) fall in TIIA were observed in stems compared to control ($p = 0.0000$) (Figure 2A,D), while a significant 1.01-fold (0.7 μg/g DW) increase in TIIA was noted in leaves relative to control ($p = 0.0001$) (Figure 2D).

3. Discussion

This work analyses the expression profiles of the *S. miltiorrhiza HMGR4* gene and its influence on the biosynthesis of tanshinones.

Previous studies have shown that *S. miltiorrhiza HMGR* genes are expressed in the roots, stems and leaves, but with different intensities in each organ. *HMGR* showed the strongest activity in roots, and weaker in stems and leaves [17]. The level of the *HMGR2*

transcript was about four-fold higher in leaves than in stems, and about two-fold higher in stems than in roots [15]. *HMGR3* was vigorously expressed in stems and root steles, and to a much greater degree than in root cortices and leaves [16]. *HMGR4* activity was the highest in flowers, lower in stems and leaves, and lowest in root steles and root cortices [16]. In the present study, a higher level of *HMGR4* mRNA was noted in leaves and stems than in the control; however, this was not observed in roots. Previous transcriptomic analyses have indicated that within the *S. miltiorrhiza* root, the strongest expression of *HMGR4* occurred in xylem [40].

The *S. miltiorrhiza HMGR4* gene showed a biphasic response to GA_3 treatment. After initial stimulation of its expression relative to the control at 12, 24 and 48 h, it then decreased and subsequently increased at 96 h (Figure 1). It has been found that 2.89 μM GA_3 has a similar influence on the *S. miltiorrhiza HMGR2* gene; however, in this case, *HMGR2* expression increased compared to control at 12 h, followed by a fall and a second increase at 72 and 96 h [41]. Elsewhere, stimulation with 400 μM GA_3 resulted in an initial rise in *Malus domestica HMGR1* transcripts against control until four hours, followed by a decrease at six h [42]. We hypothesise that stimulation of *S. miltiorrhiza HMGR4* gene expression by GA_3 and subsequent enzyme production could activate the next stages of the MVA pathway and the production of mediators necessary for the biosynthesis of endogenous gibberellins, such as ent-kaurene [43]. The newly-produced endogenous GA_3 could stimulate the *HMGR4* transcription which decreased as a result of metabolising the exogenous hormone. However, this hypothesis needs to be verified by monitoring endogenous GA_3 levels during the course of an experiment.

The impact of IAA on *S. miltiorrhiza HMGR4* expression was very similar to that induced by GA_3 (Figure 1). Although the effect of IAA on plant *HMGR* genes has not been widely studied, we have noticed some similarities in our results with previous research. IAA at a final concentration of 100 μM first raised the level of *M. domestica HMGR4* transcripts relative to control, and then lowered them [42]. The biphasic effect, which we observed in our experiment, may result from the stimulation of various TFs, some of which increase expression of the gene, while others reduce it.

The use of SA caused a rise in *HMGR4* expression at 12, 24 and 72 h and a fall at 48 and 96 h in relation to untreated material (Figure 1). A similar effect was observed for 10 mM SA against *HMGR3* in *Ginkgo biloba* leaves; however, in contrast to our present findings, the level of *HMGR3* mRNA rose against control values in the final phase of the study (96 and 120 h) [44]. Elsewhere, SA treatment was found to result in continually elevated *HMGR* transcript levels versus untreated controls in *S. miltiorrhiza* hairy roots throughout the experiment [20] and *Salvia przewalskii* hairy roots [45]. A maximum three-fold increase in *HMGR* expression was noted after 36 h of stimulation [20], and an eight-fold rise after six days [45].

The present study is the first investigation of the role of *HMGR4* in the biosynthesis of tanshinones in *S. miltiorrhiza*. Overexpression of this gene resulted in a significant increase in DHTI, CT, TI and TIIA content: by 1.51- to 2.43-fold (0.59–5.39 mg/g DW) in soil-grown roots, and by 1.82- to 3.62-fold (0.44–2.40 mg/g DW) in in vitro roots (Figure 2A–D). Of all tanshinones tested, CT showed the highest rise relative to control: 2.43-fold (5.39 mg/g DW) for roots grown in soil and 3.62-fold (2.40 mg/g DW) for in vitro roots. The results are in agreement with data received for other *S. miltiorrhiza* HMGR enzymes. Kai et al. reported that overexpression of the *HMGR* gene led to an increase in CT, TI, TIIA quantity in hairy root culture ranging from 1.17- to 3.19-fold (0.844–1.515 mg/g DW) compared to control [46]. As in our research, CT showed the highest rise in all seven transgenic lines tested. In another study, *HMGR2* overexpression significantly enhanced the amount of DHTI, CT, TI, TIIA by 1.23- to 2.46-fold (0.99–3.16 mg/L) at day 40 of root culture relative to control [15]. In the experiment, CT demonstrated the greatest increase, i.e., by 2.46-fold (3.16 mg/L).

Our results indicate that tanshinone accumulation in *S. miltiorrhiza* was organ-dependent, with roots as the primary storage place for DHTI, CT, TI, TIIA (Figure 2E). Li et al. specif-

ically indicate the root periderm of *S. miltiorrhiza* as the main site of accumulation of all tested tanshinones, viz. DHTI, CT, TI, TIIA, Tanshinone IIB, Dehydrotanshinone IIA, Dashenxinkun B, Trijuganone A, Trijuganone C; the inner layer of the roots and the outer part of stems contained much smaller amounts [47]. Subsequent research also pointed to *S. miltiorrhiza* root periderm as the main storage place for TIIA, although traces were also detected in root phloem [40]. In addition, transcriptomic analyses of the MVA and MEP pathway genes and other enzymes leading to the production of tanshinones indicated that the strongest expression of most of the tested genes (*AACT1* to *AACT6*, *HMGS2*, *HMGR1*, *HMGR2*, *MK*, *PMK*, *MDC1*, *MDC2*, *IPI1*, *GGPPS3*, *DXS2*, *DXS4*, *DXR*, *MCT*, *CMK*, *MDS*, *HDS*, *HDR1* to *HDR3*, *CPS1*, *CPS5*, *KSL1*, *KSL7*, *KSL8*, *CYP76AH1*) occurred in the periderm of *S. miltiorrhiza* roots [40]. Hence, the root periderm layer appears to be not only the main storage site, but also the main place of biosynthesis of tanshinones. The examined stems turned out to be a better source of the metabolites than leaves, but their content was quite low (several dozen μg/g DW) (Figure 2E). These results are in line with previously-performed studies [48]. Organ-specific accumulation and production of tanshinones may result from the existence of various mechanisms regulating the activity of enzymes involved in the biosynthesis of these compounds [49].

Our findings indicate that soil cultivation favoured 1.28- to 5.58-fold (0.19–5.86 mg/g DW) higher production of DHTI, CT, TI and TIIA in roots and 1.12-fold (6.2 μg/g DW) greater production of TIIA in leaves compared to in vitro conditions. This may be due to the community of microorganisms naturally present in the rhizosphere, phyllosphere and endosphere; it is possible that these may affect the biosynthesis of metabolites [50,51]. According to Yan et al., the endophytic bacteria *Pseudomonas brassicacearum* subsp. *neoaurantiaca* raised the activity of HMGR and DXS enzymes by 2.1- and 4.2-fold, respectively, in *S. miltiorrhiza* hairy root culture. This resulted in a significant increase in the content of all tanshinones tested, with particular gains found for DHTI (19.2-fold) CT (11.3-fold) and total tanshinones (3.7-fold) compared to controls [52]. In addition, the polysaccharide fraction isolated from rhizobacterium *Bacillus cereus* stimulated the accumulation of tanshinones in *S. miltiorrhiza* root culture by about seven-fold (1.59 vs. 0.19 mg/g DW) compared to control [53]. Another potential reason for the lower in vitro yields of tanshinones may be changes occurring in the morphology, anatomy and physiology of plants during in vitro cultivation [54,55].

Additionally, our findings provide further information about the influence of hormones on the biosynthesis of tanshinones in *S. miltiorrhiza*. GA_3 stimulated CT and TIIA production, but had no significant effect on DHTI content and decreased TI in in vitro root culture compared to untreated controls (Figure 2A–D). We hypothesize that the presence of GA_3 may strongly induce the expression of some key enzyme/-s involved in the terminal stage of CT biosynthesis. This could be the reason for the higher TIIA content which arises from CT; however, as GA_3 may not have a similar effect on DHTI production, the resulting TI does not rise, and may even fall [14]. GA_3 has been found to increase DHTI, CT, TI and TIIA levels in most *GRAS3*-overexpressing *S. miltiorrhiza* hairy root culture lines and in untransformed controls [56]; however, these results cannot be directly compared to ours, as the experiment used a 34.6-fold higher concentration of the hormone (100 μM) and a much shorter incubation time with GA_3, of only six days. The second hormone used, IAA, significantly reduced the accumulation of CT by 34.06-fold (3.21 mg/g DW), TIIA by 11.49-fold (1.11 mg/g DW), DHTI by 8.84-fold (0.66 mg/g DW) and TI by 5.05-fold (0.96 mg/g DW) in an in vitro root culture versus control (Figure 2A–D). Reduced CT, TI and TIIA synthesis was also observed in *S. miltiorrhiza* hairy roots treated with 5.71 μM IAA: 1.61-fold decrease (82 μg/g DW) for TI, 1.50-fold decrease (125 μg/g DW) for CT, and 1.24-fold decrease (23 μg/g DW) for TIIA, compared to control [57].

4. Materials and Methods

4.1. Establishment of S. miltiorrhiza Culture and Treatments

S. miltiorrhiza plants were cultivated from seeds provided by the Garden of Medicinal Plants of the Medical University of Lodz. To establish in vitro plant cultures, the seeds were surface sterilised utilising 70% ethanol for 1 min and subsequent 1% sodium hypochlorite solution for 5 min, and then rinsed three times with sterile distilled water for 5 min. The seeds were thereafter transferred aseptically onto Murashige and Skoog (MS) basal medium [58] with 3% sucrose (Chempur, Piekary Śląskie, Poland) and 0.65% agar (Sigma-Aldrich, Saint Louis, MO, USA) and a final pH of 5.7. Germination was carried out in the dark at 26 ± 2 °C. After germination, aerial parts of *S. miltiorrhiza* were grown in solid MS medium at 26 ± 2 °C under 16/8 h (light/dark) photoperiod at a cool fluorescent light with intensity of 40 μmol m^{-2} s^{-1}. Roots were cultivated in the dark at 26 ± 2 °C in Gamborg B5 liquid medium [59] agitated at 70 rpm. Subcultures were carried out every five weeks.

Five-week-old leaves, stems and roots, grown as described above, were used to study organ-specific expression of the *HMGR4* gene.

The effect of hormones on *HMGR4* activity was determined in five-week-old leaves. *S. miltiorrhiza* plants were incubated in sterile distilled water containing 1 mg/L (2.89 μM) GA_3 or 0.5 mg/L (2.85 μM) IAA or 20 mg/L (144.80 μM) SA and 0.01% non-ionic detergent Triton X-100 (Sigma-Aldrich, Saint Louis, MO, USA). Plants treated with sterile distilled water supplemented with 0.01% Triton X-100 were used as controls. Samples were collected after 0, 12, 24, 48, 72 and 96 h.

4.2. Selection of Potential Regulators of S. miltiorrhiza HMGR4 Gene Expression

The *S. miltiorrhiza HMGR4* promoter sequence deposited in GenBank under accession number KT921337.1 was scanned with PlantPan 2.0 tool (http://plantpan2.itps.ncku.edu.tw/, accessed on 5 June 2021) for TFBSs and interacting TFs [60]. UniProt database (https://www.uniprot.org/, accessed on 5 June 2021) was used to acquire information on received TFs [61].

4.3. Preparation of pRI201-AN-HMGR4 Overexpression Construct

The *S. miltiorrhiza HMGR4* coding sequence (1653 bp) was synthesised on the basis of JN831103.1 sequence and inserted into a pUC57 vector (Gene Universal Inc., Newark, DE, USA). The correctness of the insert was determined by double-strand Sanger sequencing. Afterwards, the *HMGR4* insert was excised from pUC57 and inserted into a pRI201-AN binary expression vector (Takara Bio Inc., Kusatsu, Japan) at NdeI/SalI sites of MCS1 (Eurofins Genomics, Ebersberg, Germany). *HMGR4* gene overexpression was driven by the strong and constitutive promoter of Cauliflower Mosaic Virus 35S (CaMV), which facilitates high levels of RNA transcription in a wide variety of plants. Analysis of the *HMGR4* sequence and flanking regions was performed by double-strand Sanger sequencing. A map of the prepared pRI201-AN-HMGR4 construct is presented in Figure 3.

4.4. Transformation, Selection, Regeneration and Treatments of S. miltiorrhiza Culture

Agrobacterium tumefaciens (*Rhizobium radiobacter*) GV2260 ($C58C1Rif^R$ with pGV2260) competent cells were transformed with the pRI201-AN-HMGR4 construct or the empty pRI201-AN vector using the freeze/thaw method [62]. The transformed bacteria were firstly grown for 84 h at 26 °C on solid selective YEB medium containing 50 mg/L kanamycin, 100 mg/L carbenicillin and 30 mg/L rifampicin (Chem-Impex International, Wood Dale, IL, USA) and then on liquid selective YEB medium with shaking at 140 rpm until OD_{600} reached 0.4–0.8. To confirm the transformation, plasmid DNA was isolated by alkaline lysis and extracted with a phenol/chloroform/isoamyl alcohol mixture [63]; this was then subjected to PCR amplification using GoTaq Hot Start Green Master Mix (Promega, Madison, WI, USA) and Kanamycin primers (Table 2). The PCR reactions were carried out in an MJ Mini Personal Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following

parameters: initial denaturation (95 °C, 5 min), denaturation (95 °C, 45 s), primer annealing (60 °C, 30 s), extension (72 °C, 30 s), final extension (72 °C, 5 min). In total, 40 PCR cycles were conducted. The obtained products were separated by 2% agarose gel electrophoresis.

Figure 3. Map of the expression construct pRI201-AN-HMGR4.

Table 2. Primers used in the study.

Primer Name	Sequence	Product Size [bp]
	Confirmation of Transformation	
Kanamycin_F	TGATCTCGTCGTGACCCAT	234
Kanamycin_R	AGAAGGCGATAGAAGGCGA	
	Real-Time PCR	
HMGR4_F	CTCAACCTGCTTGGCGTAA	185
HMGR4_R	AGTCTCGTGATGTCCCTGCT	
ACT7_F	TCCGTCTTGATCTTGCTGGT	170
ACT7_R	CGTCTTTGCAGTTTCGAGCT	

To induce virulence, bacterial cultures with confirmed transformation were collected by centrifugation and resuspended to OD_{600} = 0.1 in sterile induction medium, i.e., liquid MS medium supplemented with 100 µM acetosyringone (Sigma-Aldrich, Saint Louis, MO, USA), and then agitated on a rotary shaker at 140 rpm for five hours at 26 °C [64].

Three-month-old leaves of *S. miltiorrhiza* grown in pots were surface sterilised using the same protocol described earlier for the seeds; however, 0.8% sodium hypochlorite solution was applied. Preparation, infection of leaves and co-cultivation were performed according to Dandekar and Fisk with some modifications [64]. The composition of the induction medium was as mentioned above. Co-cultivation solid MS medium was supplemented with 1 mg/L 6-benzylaminopurine (BAP), 0.2 mg/L 1-naphthaleneacetic acid (NAA) and 100 µM acetosyringone (Sigma-Aldrich, Saint Louis, MO, USA). Overall regeneration frequency, non-transgenic regeneration under selection and non-transgenic controls were included in the research. After 72 h of incubation, leaf discs were transferred every two weeks onto fresh *A. tumefaciens* (*R. radiobacter*) killing medium, i.e., solid MS medium with 1 mg/L BAP,

0.2 mg/L NAA and 250 mg/L cefuroxime. After another six weeks, the obtained calluses were moved onto solid MS medium supplemented with 0.5 mg/L BAP, 0.2 mg/L IAA and 250 mg/L cefuroxime. In the following weeks, cefuroxime was gradually phased out and the selection antibiotic kanamycin (Biological Industries, Kibbutz Beit-Haemek, Israel) was introduced (10–50 mg/L). The aerial parts of the *S. miltiorrhiza* transformants and controls were cultivated in solid MS medium at 26 ± 2 °C under 16/8 h (light/dark) photoperiod using a cool fluorescent light with intensity of 40 µmol m^{-2} s^{-1} and their roots in the dark in liquid Gamborg B5 medium agitated at 70 rpm. Subcultures were carried out every five weeks. Additionally, in order to compare the influence of different growth environments on the biosynthesis of tanshinones, the transformants and control were transferred from in vitro cultures to pots containing sterile composite soil. The plants were covered with a transparent glass jar for three weeks and grown at 26 ± 2 °C under natural light. Figure 4 shows *S. miltiorrhiza* cultures at various stages of the experiment.

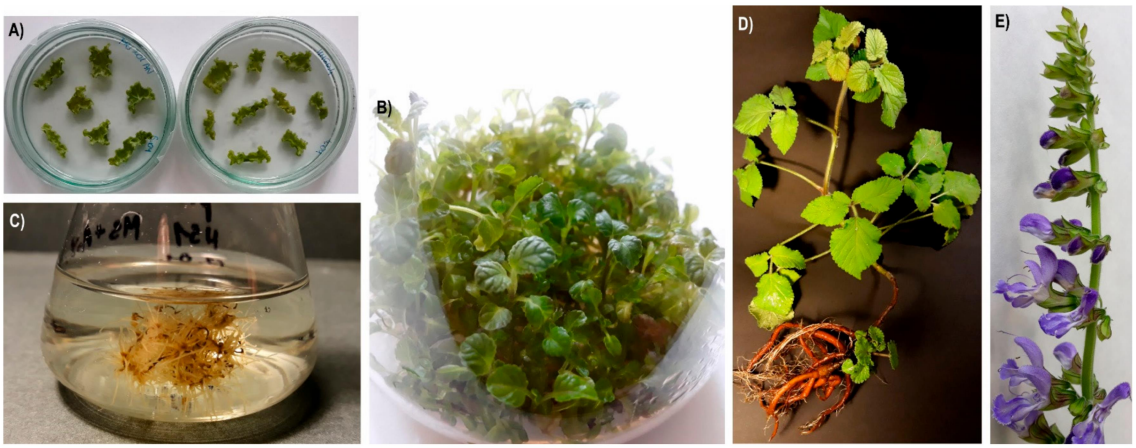


Figure 4. Cultures of *S. miltiorrhiza* at various stages of transformation, regeneration and growth. (**A**) callus resistant to selection antibiotics after two weeks of growth, (**B**) shoots regenerated from callus three months after the experiment began, (**C**) culture of roots transformed with the pRI201-AN-HMGR4 construct, (**D**) plant transformed with the pRI201-AN-HMGR4 construct after several months of growing in soil, (**E**) flowering of the plant transformed with the pRI201-AN-HMGR4 construct.

The transformation of *S. miltiorrhiza* plants was confirmed by PCR analysis of genomic DNA isolated from five-week-old leaves using Isolate II Plant DNA kit (Bioline, Taunton, MA, USA) according to the manufacturer's instructions, with the use of GoTaq Hot Start Green Master Mix (Promega, Madison, WI, USA) and Kanamycin primers (Table 2). The concentration and purity of the DNA were assessed based on A_{260}/A_{280} and A_{260}/A_{230} ratios using a Nanophotometer P300 (Implen, Munich, Germany). PCR reaction parameters were as mentioned above. The obtained products were separated via 2% agarose gel electrophoresis.

The effect of the pRI201-AN-HMGR4 construct or the empty pRI201-AN vector (control) on *S. miltiorrhiza HMGR4* expression was investigated in five-week-old leaves, stems and roots.

The importance of the *HMGR4* gene for tanshinone biosynthesis was assessed in five-week-old *S. miltiorrhiza* roots, leaves, and stems growing in soil and in vitro and overexpressing *HMGR4* relative to plant material that did not overexpress *HMGR4*.

The role of the growth environment for the tanshinone content was evaluated in five-week-old *S. miltiorrhiza* roots, leaves, and stems with and without *HMGR4* overexpression growing in soil in relation to the plant material grown in in vitro conditions.

The effect of GA_3 (1 mg/L, 2.89 µM) or IAA (0.5 mg/L, 2.85 µM) on the production of tanshinones was estimated in five-week-old roots, leaves, and stems of *S. miltiorrhiza*

overexpressing *HMGR4*, grown in vitro and treated with the hormones against untreated plant material.

The role of plant organ for the accumulation of tanshinones was assessed in five-week-old *S. miltiorrhiza* roots, leaves, and stems with and without *HMGR4* overexpression growing in soil and in in vitro conditions.

4.5. RNA Isolation, Reverse Transcription and Quantitative Real-Time PCR

Total RNA was isolated in accordance with the protocol given in NucleoSpin RNA Plant and Fungi kit (Macherey-Nagel, Duren, Germany). Plant material was ground under liquid nitrogen to a fine powder using mortar and pestle. The samples were digested by RNase-free rDNase (Macherey-Nagel, Duren, Germany) to assure removal of genomic DNA. Isolated RNA was stored at −80 °C. The concentration and purity of the RNA were evaluated using Nanophotometer P300 (Implen, Munich, Germany). The obtained A_{260}/A_{280} ratios were within the range of 1.9–2.1 and A_{260}/A_{230} ratios were ~2.

The reverse transcription reactions were carried out using Maxima H Minus Reverse Transcriptase, Oligo(dT)18 Primer, dNTP Mix, RiboLock RNase Inhibitor, and nuclease-free water (Thermo Scientific, Waltham, MA, USA) according to the manufacturer's protocol. The quantity of RNA was adjusted to achieve the same final RNA concentration in a given experiment. No reverse transcriptase and no template controls were applied. The prepared cDNA was stored at −20 °C.

In order to investigate the expression of *HMGR4*, real-time PCR reactions were performed. *Actin* (*ACT7*) was used as a reference gene [65]. Gene-specific primers (Table 2) were created based on JN831103.1 and HM051058.1 sequences using Primer3web version 4.1.0 (https://primer3.ut.ee/, accessed on 5 June 2022) and Jellyfish version 1.5 tools. Expected sizes of *HMGR4* and *ACT7* fragments (Table 2) were confirmed by agarose gel electrophoresis. Amplification reactions were conducted using SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich, Saint Louis, MO, USA) according to the manufacturer's instructions on Rotor-Gene 6000 (Corbett Research, Manchester, United Kingdom). The real-time PCR reaction parameters were as follows: initial denaturation (95 °C, 10 min), 40 cycles of denaturation (95 °C, 20 s), primer annealing (60 °C, 30 s), extension (72 °C, 10 s). The obtained products were melted in the temperature range of 72–95 °C with an increment of 1 °C. Ct values for *HMGR4* were normalised to Ct values for *ACT7* and calculated relative to a calibrator according to Pfaffl method [66].

4.6. Quantitative Analysis of Tanshinones

Roots, stems and leaves of *S. miltiorrhiza* were freeze-dried in a lyophiliser Alpha 1-2 LD (Martin Christ, Osterode am Harz, Germany) under 0.1 mbar pressure and ground with a pestle and mortar to a fine powder. The obtained powder (50 mg) was extracted with methanol (2 mL) under ultrasonic treatment for one hour at room temperature. The mixture was then centrifuged at 14,000× *g* for 5 min and then the supernatant was filtered through a 0.2 μm organic membrane filter (Millipore, Burlington, MA, USA) [67].

The tanshinone content was determined with a UHPLC 1290 Infinity instrument (Agilent Technologies, Santa Clara, CA, USA). Chromatographic separation was performed on a Zorbax Eclipse XDB-C18 column (3 mm × 100 mm, 1.8 μm particle size; Agilent Technologies, Santa Clara, CA, USA) with a Zorbax Eclipse XDB-C18 pre-column (3 mm × 5 mm, 1.8 μm particle size; Agilent Technologies, Santa Clara, CA, USA) thermostatted at 30 °C. The mobile phase consisted of 0.1% (*v/v*) formic acid acetonitrile solution (A) and 0.1% (*v/v*) formic acid aqueous solution (B), and the flow rate was 0.4 mL/min. The following gradient was used (all concentrations are *v/v*): 0–2 min, 40–55% A; 2–12 min, 55–50% A; 12–13 min, 50–80% A; 13–17 min, 80–95% A; 17–20 min, 95% A. The column and pre-column were equilibrated to 40% A for 1.5 min. The samples were injected in a volume of 1 μL and the wavelength of 270 nm was applied for the detection of tanshinones. HPLC grade DHTI, CT, TI, TIIA standards (Sigma-Aldrich, Saint Louis, MO, USA) were used for calibration. HPLC grade methanol, acetonitrile, and water were provided by Sigma-Aldrich (Saint

Louis, MO, USA). More details on the UHPLC analysis are provided in Table 3. The data were collected and processed using ChemStation 3D software.

Table 3. Details of UHPLC tanshinone analysis in *S. miltiorrhiza*.

Analyte	Retention Time [min]	Standard Curve	R^2
DHTI	6.025	y = 190.163x − 16.690	0.99973
CT	10.009	y = 146.106x − 10.926	0.99992
TI	10.938	y = 241.276x − 21.816	0.99989
TIIA	14.808	y = 327.209x − 20.014	0.99996

4.7. Statistical Analysis

Statistica 13.3 software (TIBCO Software Inc, Palo Alto, CA, USA) was used for analyses. The collected UHPLC results were checked for normal distribution using the Shapiro–Wilk test, and subsequent analyses were performed using the Kruskal–Wallis test, Mann–Whitney U test and *t*-test. Values with $p < 0.05$ were considered statistically significant. Expression values were given as mean ± SD.

5. Conclusions

The most important observation from the conducted research concerns the important role played by *HMGR4* in the biosynthesis of tanshinones, which is reflected in the content of DHTI, CT, TI, TIIA in the roots and TIIA in the stems and leaves with gene overexpression.

Other conclusions:

GA_3, IAA and SA regulated the expression of the *S. miltiorrhiza HMGR4* gene, confirming the results of the in silico promoter analysis.

The soil environment promoted a higher accumulation of all tested metabolites in roots and TIIA in leaves compared to in vitro conditions. However, it is worth noting that the amounts of DHTI and TI in in vitro roots with *HMGR4* overexpression were higher than in soil-grown roots without overexpression.

Apart from the positive effect on the appearance of TIIA in the studied stems and leaves of *S. miltiorrhiza*, *HMGR4* overexpression did not change the characteristic organ-dependent pattern of tanshinone accumulation, i.e., the main source was the root, with trace amounts observed in stems and leaves.

GA_3 increased CT and TIIA production in roots, while IAA reduced the biosynthesis of all tested metabolites.

The greatest efficiency of tanshinone biosynthesis was found to result from a combination of three traits, namely *HMGR4* gene overexpression, root organ, and cultivation in soil conditions.

Future research could investigate the mechanisms controlling *S. miltiorrhiza HMGR4* gene expression. TFs regulating *HMGR4* expression could be isolated using the yeast-one hybrid (Y1H) system and then functionally characterised [68]. The role of specific TFBSs in the response of *HMGR4* to abiotic or biotic factors could be verified by its mutagenesis [69]. TF networks that play a key role in the regulation of *HMGR4* gene expression could be explored through transcriptomic RNA sequencing and weighted gene co-expression network analysis (WGCNA) [70,71].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/molecules27144354/s1, Table S1: Potential transcription factor binding sites (TFBSs) and interacting transcription factors (TFs) that respond to gibberellic acid (GA_3) signals found in *Salvia miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 tool; Table S2: Potential TFBSs and interacting TFs that respond to indole-3-acetic acid (IAA) signals found in *S. miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 tool; Table S3: Potential TFBSs and interacting TFs that respond to salicylic acid (SA) signals found in *S. miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 tool.

Author Contributions: Conceptualization, M.M., P.S. and Ł.K.; methodology, M.M., P.S., J.G., A.J., R.G., E.B. and Ł.K.; formal analysis, M.M., A.J. and Ł.K.; investigation, M.M., J.G. and Ł.K.; writing—original draft preparation, M.M.; writing—review and editing, P.S., A.J., E.B. and Ł.K.; visualization, M.M.; supervision, P.S. and Ł.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by by the Medical University of Lodz (grant nos. 502-03/3-012-02/502-34-103, 503/3-012-01/503-31-001-19-00, 503/3-015-02/503-31-001-19-00).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are included in the article.

Acknowledgments: The authors would like to thank Dirk Inzé (Gent University, Belgium) for *A. tumefaciens* (*R. radiobacter*) GV2260 cells.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. Cheng, T.O. Cardiovascular effects of Danshen. *Int. J. Cardiol.* **2007**, *121*, 9–22. [CrossRef] [PubMed]
2. Xu, Y.; Wan, R.; Lin, Y.; Yang, L.; Chen, Y.; Liu, C. Recent advance on research and application of Salvia miltiorrhiza. *Asian J. Pharmacodyn. Pharmacokinet.* **2007**, *7*, 99–130.
3. Zhang, H.; Long, M.; Wu, Z.; Han, X.; Yu, Y. Sodium tanshinone IIA silate as an add-on therapy in patients with unstable angina pectoris. *J. Thorac. Dis.* **2014**, *6*, 1794–1799. [CrossRef] [PubMed]
4. Jia, L.Q.; Zhang, N.; Xu, Y.; Chen, W.N.; Zhu, M.L.; Song, N.; Ren, L.; Cao, H.M.; Wang, J.Y.; Yang, G.L. Tanshinone IIA affects the HDL subfractions distribution not serum lipid levels: Involving in intake and efflux of cholesterol. *Arch. Biochem. Biophys.* **2016**, *592*, 50–59. [CrossRef]
5. Chiu, S.C.; Huang, S.Y.; Chen, S.P.; Su, C.C.; Chiu, T.L.; Pang, C.Y. Tanshinone IIA inhibits human prostate cancer cells growth by induction of endoplasmic reticulum stress in vitro and in vivo. *Prostate Cancer Prostatic Dis.* **2013**, *16*, 315–322. [CrossRef]
6. Wang, S.; Jing, H.; Yang, H.; Liu, Z.; Guo, H.; Chai, L.; Hu, L. Tanshinone I selectively suppresses pro-inflammatory genes expression in activated microglia and prevents nigrostriatal dopaminergic neurodegeneration in a mouse model of Parkinson's disease. *J. Ethnopharmacol.* **2015**, *164*, 247–255. [CrossRef]
7. Kim, S.Y.; Moon, T.C.; Chang, H.W.; Son, K.H.; Kang, S.S.; Kim, H.P. Effects of tanshinone I isolated from Salvia miltiorrhiza bunge on arachidonic acid metabolism and in vivo inflammatory responses. *Phytother. Res.* **2002**, *16*, 616–620. [CrossRef]
8. Tang, F.; Wu, X.; Wang, T.; Wang, P.; Li, R.; Zhang, H.; Gao, J.; Chen, S.; Bao, L.; Huang, H.; et al. Tanshinone II A attenuates atherosclerotic calcification in rat model by inhibition of oxidative stress. *Vascul. Pharmacol.* **2007**, *46*, 427–438. [CrossRef]
9. Wang, C.; Du, X.; Yang, R.; Liu, J.; Xu, D.; Shi, J.; Chen, L.; Shao, R.; Fan, G.; Gao, X.; et al. The prevention and treatment effects of tanshinone IIA on oestrogen/androgen-induced benign prostatic hyperplasia in rats. *J. Steroid Biochem. Mol. Biol.* **2015**, *145*, 28–37. [CrossRef]
10. Wang, L.; Cheng, L.; Zhang, B.; Wang, N.; Wang, F. Tanshinone prevents alveolar bone loss in ovariectomized osteoporosis rats by up-regulating phosphoglycerate dehydrogenase. *Toxicol. Appl. Pharmacol.* **2019**, *376*, 9–16. [CrossRef]
11. Honda, G.; Koezuka, Y.; Tabata, M. Isolation of an antidermatophytic substance from the root of Salvia miltiorrhiza. *Chem. Pharm. Bull.* **1988**, *36*, 408–411. [CrossRef] [PubMed]
12. Lee, A.R.; Wu, W.L.; Chang, W.L.; Lin, H.C.; King, M.L. Isolation and bioactivity of new tanshinones. *J. Nat. Prod.* **1987**, *50*, 157–160. [CrossRef] [PubMed]
13. Mei, X.D.; Cao, Y.F.; Che, Y.Y.; Li, J.; Shang, Z.P.; Zhao, W.J.; Qiao, Y.J.; Zhang, J.Y. Danshen: A phytochemical and pharmacological overview. *Chin. J. Nat. Med.* **2019**, *17*, 59–80. [CrossRef]
14. Lu, S. *The Salvia Miltiorrhiza Genome*; Springer: Cham, Switzerland, 2019; pp. 129–139.
15. Dai, Z.; Cui, G.; Zhou, S.F.; Zhang, X.; Huang, L. Cloning and characterization of a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from Salvia miltiorrhiza involved in diterpenoid tanshinone accumulation. *J. Plant Physiol.* **2011**, *168*, 148–157. [CrossRef]
16. Ma, Y.; Yuan, L.; Wu, B.; Li, X.; Chen, S.; Lu, S. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in Salvia miltiorrhiza. *J. Exp. Bot.* **2012**, *63*, 2809–2823. [CrossRef]
17. Liao, P.; Zhou, W.; Zhang, L.; Wang, J.; Yan, X.; Zhang, Y.; Zhang, R.; Li, L.; Zhou, G.; Kai, G. Molecular cloning, characterization and expression analysis of a new gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from Salvia miltiorrhiza. *Acta Physiol. Plant.* **2009**, *31*, 565–572. [CrossRef]
18. Wang, J.W.; Wu, J.Y. Tanshinone biosynthesis in Salvia miltiorrhiza and production in plant tissue cultures. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 437–449. [CrossRef]

19. Cheng, Q.; He, Y.; Li, G.; Liu, Y.; Gao, W.; Huang, L. Effects of combined elicitors on tanshinone metabolic profiling and SmCPS expression in Salvia miltiorrhiza hairy root cultures. *Molecules* **2013**, *18*, 7473–7485. [CrossRef]
20. Hao, X.; Shi, M.; Cui, L.; Xu, C.; Zhang, Y.; Kai, G. Effects of methyl jasmonate and salicylic acid on tanshinone production and biosynthetic gene expression in transgenic Salvia miltiorrhiza hairy roots. *Biotechnol. Appl. Biochem.* **2015**, *62*, 24–31. [CrossRef]
21. Wu, C.T.; Mulabagal, V.; Nalawade, S.M.; Chen, C.L.; Yang, T.F.; Tsay, H.S. Isolation and quantitative analysis of cryptotanshinone, an active quinoid diterpene formed in callus of Salvia miltiorrhiza BUNGE. *Biol. Pharm. Bull.* **2003**, *26*, 845–848. [CrossRef]
22. Xu, H.; Liu, Q.; Yao, T.; Fu, X. Shedding light on integrative GA signaling. *Curr. Opin. Plant Biol.* **2014**, *21*, 89–95. [CrossRef] [PubMed]
23. Gupta, R.; Chakrabarty, S.K. Gibberellic acid in plant: Still a mystery unresolved. *Plant Signal. Behav.* **2013**, *8*, 1–5. [CrossRef]
24. Wu, K.; Xu, H.; Gao, X.; Fu, X. New insights into gibberellin signaling in regulating plant growth-metabolic coordination. *Curr. Opin. Plant Biol.* **2021**, *63*, 102074. [CrossRef]
25. Li, W.; Liu, C.; Liu, J.; Bai, Z.; Liang, Z. Transcriptomic analysis reveals the GRAS family genes respond to gibberellin in Salvia miltiorrhiza hairy roots. *BMC Genom.* **2020**, *21*, 727. [CrossRef]
26. Li, W.; Bai, Z.; Pei, T.; Yang, D.; Mao, R.; Zhang, B.; Liu, C.; Liang, Z. SmGRAS1 and SmGRAS2 Regulate the Biosynthesis of Tanshinones and Phenolic Acids in Salvia miltiorrhiza. *Front. Plant Sci.* **2019**, *10*, 1–13. [CrossRef] [PubMed]
27. Ma, Q.; Grones, P.; Robert, S. Auxin signaling: A big question to be addressed by small molecules. *J. Exp. Bot.* **2018**, *69*, 313–328. [CrossRef] [PubMed]
28. Leyser, O. Auxin signaling. *Plant Physiol.* **2018**, *176*, 465–479. [CrossRef]
29. Freire-Rios, A.; Tanaka, K.; Crespo, I.; van der Wijk, E.; Sizentsova, Y.; Levitsky, V.; Lindhoud, S.; Fontana, M.; Hohlbein, J.; Boer, D.R.; et al. Architecture of DNA elements mediating ARF transcription factor binding and auxin-responsive gene expression in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 24557–24566. [CrossRef]
30. Xu, Z.; Ji, A.; Song, J.; Chen, S. Genome-wide analysis of auxin response factor gene family members in medicinal model plant Salvia miltiorrhiza. *Biol. Open* **2016**, *5*, 848–857. [CrossRef]
31. Kuai, X.; MacLeod, B.J.; Després, C. Integrating data on the Arabidopsis NPR1/NPR3/NPR4 salicylic acid receptors; a differentiating argument. *Front. Plant Sci.* **2015**, *6*, 1–5. [CrossRef]
32. Wu, Y.; Zhang, D.; Chu, J.Y.; Boyle, P.; Wang, Y.; Brindle, I.D.; De Luca, V.; Després, C. The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Rep.* **2012**, *1*, 639–647. [CrossRef] [PubMed]
33. Zhou, J.M.; Trifa, Y.; Silva, H.; Pontier, D.; Lam, E.; Shah, J.; Klessig, D.F. NPR1 differentially interacts with members of the TGA/OBF family of transcription factors that bind an element of the PR-1 gene required for induction by salicylic acid. *Mol. Plant Microbe Interact.* **2000**, *13*, 191–202. [CrossRef] [PubMed]
34. Fonseca, J.P.; Menossi, M.; Thibaud-Nissen, F.; Town, C.D. Functional analysis of a TGA factor-binding site located in the promoter region controlling salicylic acid-induced NIMIN-1 expression in Arabidopsis. *Genet. Mol. Res.* **2010**, *9*, 167–175. [CrossRef] [PubMed]
35. Sun, T.; Busta, L.; Zhang, Q.; Ding, P.; Jetter, R.; Zhang, Y. TGACG-BINDING FACTOR 1 (TGA1) and TGA4 regulate salicylic acid and pipecolic acid biosynthesis by modulating the expression of SYSTEMIC ACQUIRED RESISTANCE DEFICIENT 1 (SARD1) and CALMODULIN-BINDING PROTEIN 60g (CBP60g). *New Phytol.* **2018**, *217*, 344–354. [CrossRef]
36. Blanco, F.; Garretón, V.; Frey, N.; Dominguez, C.; Pérez-Acle, T.; Van der Straeten, D.; Jordana, X.; Holuigue, L. Identification of NPR1-dependent and independent genes early induced by salicylic acid treatment in Arabidopsis. *Plant Mol. Biol.* **2005**, *59*, 927–944. [CrossRef]
37. Cao, W.; Wang, Y.; Shi, M.; Hao, X.; Zhao, W.; Wang, Y.; Ren, J.; Kai, G. Transcription factor SmWRKY1 positively promotes the biosynthesis of tanshinones in Salvia miltiorrhiza. *Front. Plant Sci.* **2018**, *9*, 1–10. [CrossRef]
38. Hernandez-Garcia, C.M.; Finer, J.J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Sci.* **2014**, *217–218*, 109–119. [CrossRef]
39. Yu, C.P.; Lin, J.J.; Li, W.H. Positional distribution of transcription factor binding sites in Arabidopsis thaliana. *Sci. Rep.* **2016**, *6*, 25164. [CrossRef]
40. Xu, Z.; Peters, R.J.; Weirather, J.; Luo, H.; Liao, B.; Zhang, X.; Zhu, Y.; Ji, A.; Zhang, B.; Hu, S.; et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. *Plant J.* **2015**, *82*, 951–961. [CrossRef]
41. Szymczyk, P.; Grąbkowska, R.; Skała, E.; Żebrowska, M.; Balcerczak, E.; Jeleń, A. Isolation and characterization of a 3-hydroxy-3-methylglutaryl coenzyme A reductase 2 promoter from Salvia miltiorrhiza. *J. Plant Biochem. Biotechnol.* **2018**, *27*, 223–236. [CrossRef]
42. Lv, D.; Zhang, Y. Isolation and functional analysis of apple MdHMGR1 and MdHMGR4 gene promoters in transgenic Arabidopsis thaliana. *Plant Cell Tissue Organ Cult.* **2017**, *129*, 133–143. [CrossRef]
43. Kasahara, H.; Hanada, A.; Kuzuyama, T.; Takagi, M.; Kamiya, Y.; Yamaguchi, S. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in Arabidopsis. *J. Biol. Chem.* **2002**, *277*, 45188–45194. [CrossRef] [PubMed]
44. Rao, S.; Meng, X.; Liao, Y.; Yu, T.; Cao, J.; Tan, J.; Xu, F.; Cheng, S. Characterization and functional analysis of two novel 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes (GbHMGR2 and GbHMGR3) from Ginkgo biloba. *Sci. Rep.* **2019**, *9*, 14109. [CrossRef] [PubMed]

45. Li, J.; Li, B.; Luo, L.; Cao, F.; Yang, B.; Gao, J.; Yan, Y.; Zhang, G.; Peng, L.; Hu, B. Increased phenolic acid and tanshinone production and transcriptional responses of biosynthetic genes in hairy root cultures of Salvia przewalskii Maxim. treated with methyl jasmonate and salicylic acid. *Mol. Biol. Rep.* **2020**, *47*, 8565–8578. [CrossRef]
46. Kai, G.; Xu, H.; Zhou, C.; Liao, P.; Xiao, J.; Luo, X.; You, L.; Zhang, L. Metabolic engineering tanshinone biosynthetic pathway in Salvia miltiorrhiza hairy root cultures. *Metab. Eng.* **2011**, *13*, 319–327. [CrossRef]
47. Li, S.; Zhu, N.; Tang, C.; Duan, H.; Wang, Y.; Zhao, G.; Liu, J.; Ye, Y. Differential distribution of characteristic constituents in root, stem and leaf tissues of Salvia miltiorrhiza using MALDI mass spectrometry imaging. *Fitoterapia* **2020**, *146*. [CrossRef]
48. Hang, L.; Wang, J.; Yang, D.; Shu, Z.; Liang, Z. Distribution traits of bioactives in different parts of Salvia miltiorrhiza Bunge. and Salvia miltiorrhiza Bunge. f. alba. *J. Northwest A F Univ.* **2008**, *36*, 217–222.
49. Leivar, P.; Antolín-Llovera, M.; Ferrero, S.; Closa, M.; Arró, M.; Ferrer, A.; Boronat, A.; Campos, N. Multilevel control of Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase by protein phosphatase 2A. *Plant Cell* **2011**, *23*, 1494–1511. [CrossRef]
50. Huang, W.; Long, C.; Lam, E. Roles of plant-associated microbiota in traditional herbal medicine. *Trends Plant Sci.* **2018**, *23*, 559–562. [CrossRef]
51. Jeliazkov (Zheljazkov), V.D.; Cantrell, C.L. *Medicinal and Aromatic Crops: Production, Phytochemistry, and Utilization*; American Chemical Society: Washington, DC, USA, 2016; pp. 75–90.
52. Yan, Y.; Zhang, S.; Zhang, J.; Ma, P.; Duan, J.; Liang, Z. Effect and mechanism of endophytic bacteria on growth and secondary metabolite synthesis in Salvia miltiorrhiza hairy roots. *Acta Physiol. Plant.* **2014**, *36*, 1095–1105. [CrossRef]
53. Zhao, J.; Zhou, L.; Wu, J. Promotion of Salvia miltiorrhiza hairy root growth and tanshinone production by polysaccharide–protein fractions of plant growth-promoting rhizobacterium Bacillus cereus. *Process Biochem.* **2010**, *45*, 1517–1522. [CrossRef]
54. George, E.F.; Hall, M.A.; De Klerk, G. *Plant Propagation by Tissue Culture*, 3rd ed.; Springer: Dordrecht, The Netherlands, 2008; pp. 465–477.
55. Debergh, P.C.; Zimmerman, R.H. *Micropropagation. Technology and Application*, 1st ed.; Kluwer Academic Publishers: London, UK, 1991; pp. 71–93.
56. Li, W.; Xing, B.; Mao, R.; Bai, Z.; Yang, D.; Xu, J.; Liang, Z. SmGRAS3 negatively responds to GA signaling while promotes tanshinones biosynthesis in Salvia miltiorrhiza. *Ind. Crops Prod.* **2020**, *144*. [CrossRef]
57. Gupta, S.K.; Liu, R.; Liaw, S.; Chan, H.; Tsay, H. Enhanced tanshinone production in hairy roots of 'Salvia miltiorrhiza Bunge' under the influence of plant growth regulators in liquid culture. *Bot. Stud.* **2011**, *52*, 435–443.
58. Murashige, T.; Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **1962**, *15*, 473–497. [CrossRef]
59. Gamborg, O.L.; Miller, R.A.; Ojima, K. Nutrient requirements of suspension cultures of Soybean root cells. *Exp. Cell Res.* **1968**, *50*, 151–158. [CrossRef]
60. Chow, C.N.; Zheng, H.Q.; Wu, N.Y.; Chien, C.H.; Huang, H.D.; Lee, T.Y.; Chiang-Hsieh, Y.F.; Hou, P.F.; Yang, T.Y.; Chang, W.C. PlantPAN 2.0: An update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* **2016**, *44*, D1154–D1160. [CrossRef]
61. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]
62. Wang, K. *Agrobacterium Protocols*, 2nd ed.; Humana Press: Totowa, NJ, USA, 2006; pp. 43–53.
63. Ausubel, F.M.; Brent, R.; Kingston, R.E.; Moore, D.D.; Seidman, J.G.; Smith, J.A.; Struhl, K. *Current Protocols in Molecular Biology*; John Wiley & Sons: New York, NY, USA, 2003; pp. 1.6.1–1.6.10, 12.11.11–12.11.10.
64. Peña, L. *Transgenic Plants. Methods and Protocols*; Humana Press: Totowa, NJ, USA, 2005.
65. Yang, Y.; Hou, S.; Cui, G.; Chen, S.; Wei, J.; Huang, L. Characterization of reference genes for quantitative real-time PCR analysis in various tissues of Salvia miltiorrhiza. *Mol. Biol. Rep.* **2010**, *37*, 507–513. [CrossRef]
66. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**, *29*, 2002–2007. [CrossRef]
67. Wan, X.; Wang, Y.; Row, K.H. Separation of tanshinone I, tanshinone IIA, and cryptotanshinone from Salvia miltiorrhiza Bunge by normal phase HPLC. *J. Liq. Chromatogr. Relat. Technol.* **2009**, *32*, 544–552. [CrossRef]
68. Yuan, L.; Perry, S.E. *Plant Transcription Factors*, 1st ed.; Humana Press: Totowa, NJ, USA, 2011; pp. 45–66.
69. Luján, M.A.; Soria-García, Á.; Claver, A.; Lorente, P.; Rubio, M.C.; Picorel, R.; Alfonso, M. Different Cis-Regulatory Elements Control the Tissue-Specific Contribution of Plastid ω-3 Desaturases to Wounding and Hormone Responses. *Front. Plant Sci.* **2021**, *12*, 727292. [CrossRef] [PubMed]
70. Chen, H.; Yang, Q.; Fu, H.; Chen, K.; Zhao, S.; Zhang, C.; Cai, T.; Wang, L.; Lu, W.; Dang, H.; et al. Identification of Key Gene Networks and Deciphering Transcriptional Regulators Associated With Peanut Embryo Abortion Mediated by Calcium Deficiency. *Front. Plant Sci.* **2022**, *13*, 814015. [CrossRef] [PubMed]
71. Lei, X.; Liu, Z.; Xie, Q.; Fang, J.; Wang, C.; Li, J.; Wang, C.; Gao, C. Construction of two regulatory networks related to salt stress and lignocellulosic synthesis under salt stress based on a Populus davidiana × P. bolleana transcriptome analysis. *Plant Mol. Biol.* **2022**. [CrossRef] [PubMed]

Table S1. Potential transcription factor binding sites (TFBSs) and interacting transcription factors (TFs) that respond to gibberellic acid (GA_3) signals found in *Salvia miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 tool.

TFBS motif and localisation[a,b,c]		TF family name	TF gene name and locus	processes in which TF is involved
AATTA 1268 **AATAT 1263**	AATAA 309 TTATT 475 ATATT 345; 631	AT-Hook	*AHL25;* At4g35390	negative regulation of GA_3-mediated signaling pathway; vegetative to reproductive phase transition of meristem
TATCT 186		MYB-related	*CCA1;* At2g46830	responded e.g. to GA_3; negative regulation of circadian rhythm; phytochrome regulation; long-day photoperiodism and flowering; regulation of protein homodimerization activity
ATATC 1166	**GATAT 1215**		*RVE8;* At3g09600	responded e.g. to GA_3; histone H3 acetylation; photoperiodism and flowering; positive regulation of circadian rhythm
			RVE4; At5g02840	responded e.g. to GA_3; regulation of circadian rhythm
ATAAT 309		Homeodomain; HD-ZIP	*ATHB-23;* At1g26960	induced by GA_3; lateral root development
CACAT 644	ACATG 644	bHLH	*PIF3;* At1g09530	regulation of GA_3-mediated, phytochrome and red/far-red light signaling pathways; de-etiolation; positive regulation of anthocyanin metabolic process
GGATC 220 TATCC 377 GGATA 392 GATAC 393 GATAA 439 **CGATG 1227** **AGATT 1405** CGATT 424; 1143 TATCG 448; 1169 CATCT 753; 872 **AATCA** 11; **1315** AATCG 75; 130 GATCG 101; 221	GGATG 962; 1178 ATATC 189; 1168 GTATC 376; 761 CTATC 447; 891 **GATAT** 769; **1217** AATCT 15; 197; 919 **AGATG** 120; 662; **1427** CATCC 141; 778; 999 TATCT 162; 190; 326 AATCC 179; 204; 548 **TGATA** 438; 768; **1216** **TTATC** 161; 325; 689; 860; **1226** **TATCA** 690; 762; 861; 892; **1213** TGATT 648; 1076; 1101; 1131; 1198	GATA	*GATA22;* At4g26150 *GATA21;* At5g56860	induced e.g. by GA_3; modulation of chloroplast development, growth and division in cytokinin-dependent manner; chlorophyll biosynthesis; regulation of sugar-sensing genes; circadian rhythm; regulation of germination, senescence, elongation growth and flowering time; influence leaf starch content
TTTTT 77; 78; 79; 80; 81; 82; 83; 84		MADS box;	*SOC1;*	up-regulated by GA_3; integration of signals from photoperiod,

	MIKC	At2g45660		vernalisation and autonomous floral induction pathways; when associated with AGL24, mediate effect of GA_3 on flowering under short-day conditions
AATGG 842 TTTTG 805 **CAAAA 1396**	ATTTCC 804; 801 **GGAAA 1364; 1362** TTTTT 77; 78; 79; 80; 81; 82; 83; 84	*AGL24;* At4g24540		up-regulated by GA_3; floral transition in response to vernalization; effect of GA_3 on flowering under short-day conditions; identity of floral meristem; inflorescence fate in apical meristems
AATGG 842 TTTCC 804; 801	**GGAAA 1364; 1362** TTTTT 77; 78; 79; 80; 81; 82; 83; 84	*AGL71;* At5g51870		act through GA_3-dependent pathway; flowering time; flowering at shoot apical and axillary meristems
TTTTG 805 **CAAAA 1396**	TTTCC 804; 801 **GGAAA 1364; 1362** TTTTT 77; 78; 79; 80; 81; 82; 83; 84	*AGL72;* At5g51860	MADS box	
		AGL42; At5g62165		act through GA_3-dependent pathway; flowering time; flowering at shoot apical and axillary meristems; control of flower organ senescence and abscission; leaf senescence
CTAAT 42 CCCAT 139 GTTGG 389 CCAAC 530 AGTGG 592 TTTGG 811 CCTAT 890 ATGGG959 ATTGC 1039 CGAAT 1115 **ATTAG 1389** CCAAT 195; 202	ATCGG 102; 131 CCAAG 149; 213 AATGG 462; 847 ATTGA 766; 1099 CCATT 169; 266; 701 **CCACT** 144; 380; **1364** **TCAAT** 13; 1149; **1253** **ACAAT** 460; 800; **1247** ATTGT 599; 637; 1145; 1152 **CCAAA** 176; 534; **1261; 1396** C**AAAT** 177; 469; 897; 1107; 1161**;** **1262; 1313**	*NFYC3;* At1g54830	NF-YC	involved in GA_3- and abscisic acid-activated signaling pathway; long-day photoperiodism and flowering; positive regulation of photomorphogenesis; regulation of seed germination

[a] For TFBSs only most conserved positions within a matrix were listed. [b] Binding sites localised in proximal promoter region are in bold. [c] The transcription start site (TSS) is located at 1500 nucleotide of the studied promoter sequence.

Table S2. Potential TFBSs and interacting TFs that respond to indole-3-acetic acid (IAA) signals found in *S. miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 tool.

TFBS motif and localisation[a,b,c]		TF family name	TF gene name and locus	processes in which TF is involved
TATCT 186		MYB-related	*CCA1;* At2g46830	responded e.g. to IAA; see also *CCA1* in GA_3 part
ATATC 1166	**GATAT 1215**		*RVE 8;* At3g09600	responded e.g. to IAA; see also *RVE8* in GA_3 part
TGATT 645; 1195		Homeodomain; HD-ZIP	*HAT2;* At5g47370	induced by IAA; auxin-mediated morphogenesis; negative regulation of lateral root elongation
ATAAT 310	AATAA 308; 307		*ATHB-20;* At3g01220	induced e.g. by IAA; seed germination in micropylar endosperm
CCGCG 229	CGCGT 229	CAMTA	*CAMTA1;* At5g09410	IAA signaling and response to abiotic stresses; binding of calmodulin in calcium-dependent manner; freezing tolerance; drought stress response
AATGG 842 TTTTG 805 **CAAAA 1396**	TTTCC 804; 801 A**GAAA** 293; **1402** **GGAAA 1364; 1362** TTTTT 77; 78; 79; 80; 81; 82; 83; 84	MADS box; MIKC	*AGL14;* At4g11880	induced by IAA; regulation of root development by controlling meristem size and patterning of root apical meristem; regulation of auxin transport and gradients in root meristematic cells; regulation of shoot apical meristem cell identities and transitions; flowering transition, flower meristem maintenance and determinacy; plants over-expressing *AGL14* show early flowering phenotype and flowers have vegetative traits

[a] For TFBSs only most conserved positions within a matrix were listed. [b] Binding sites localised in proximal promoter region are in bold. [c] The TSS is located at 1500 nucleotide of the studied promoter sequence.

Table S3. Potential TFBSs and interacting TFs that respond to salicylic acid (SA) signals found in *S. miltiorrhiza HMGR4* promoter sequence using PlantPan 2.0 tool.

TFBS motif and localisation[a,b,c]		TF family name	TF gene name and locus	processes in which TF is involved
CGTA 720		NAC; NAM	*NAC081;* At5g08790	induced e.g. by SA; regulation of host basal defense responses against viral infection; repression of pathogenesis-related proteins; suppression of brassinosteroid-inactivating enzymes CYP734A1/BAS1 and CYP72C1/SOB7; integration of brassinosteroid homeostasis and seedling development; spatial regulation of brassinosteroid homeostasis, regulation of hypocotyl elongation and root growth by suppression of brassinosteroid catabolism; transactivation of auxin biosynthetic gene *NIT2* promoter; abscisic acid-inducible leaf senescence signalling; seed development and morphology
AGTAA 124	AGTAA 565		*NAC062;* At3g49530	induced e.g. by SA; induction of pathogenesis-related genes in response to cold; regulatory role in -mediated drought-resistance response
ACCTA 887	**AACCA** 152; 526; **1485**	MYB	*MYB46;* At5g12870	slightly induced by SA; positive regulation of secondary wall biosynthesis in fibers and vessels; defense response to fungi; activation of mannan synthase *CSLA9* gene
ACCTA 886		Myb/SANT; MYB	*MYB3;* At1g22640	induced e.g. by SA; cell differentiation; cinnamic acid biosynthetic process; negative regulation of metabolic process
TATCT 186		MYB-related	*CCA1;* At2g46830	responded e.g. to SA; see also *CCA1* in GA_3 part
ATATC 1166	**GATAT 1215**		*RVE8;* At3g09600	responded e.g. to SA; see also *RVE8* in GA_3 part
			RVE4; At5g02840	responded e.g. to SA; see also *RVE4* in GA_3 part
GCCTT 79 GCTTT 369 AAGGG 1031 **AAGGC 1290**	ACCTT 158; 184; 605 TCTTT 252; 586; 874 AAAGC 400; 571 **AAAGA** 939; **1413; 1420**	Dof	*DOF1.1;* At1g07640	induced e.g. by SA; regulation of root development and glucosinolate biosynthetic process; promotion of radial growth of protophloem sieve elements and apical dominance

A**AGGA 1372** CCTTT 80; 807 AAAGT 109; 982	**AAAGG** 1030; **1289; 1371** TCCTT 275; 328; 338; 806; 1156 **ACTTT** 382; 881; 509; 612; 1021; **1297**		*DOF3.4;* At3g50410	induced e.g. by SA; cell wall modification; positive regulation of cell cycle; zinc ion binding
CGTCC 362 GGACG 407 CGTTA 687 CGTAA 722 TGACA 796 GTCAT 944 TCACG 1112 **CCTCA 1458**	TGACT 5; 832 ATGAC 4; 795 **AGTCA** 111; **1240** CGTCA 232; 229 CGTCT 250; 782 TGAAG 565; 789; 969 **TGACG 1210; 1208** CTTCA 237; 607; 955; 1158 **TGTCA** 601; 913; 948; 1147; **1251**	bZIP	*TGA2;* At5g06950	recognition and binding to as-1-like cis elements mediating auxin- and SA-inducible transcription; induction of systemic acquired resistance via regulation of pathogenesis-related genes expression
CCATG 574	GCACA 643	bHLH	*LRL1;* At2g24260	induced e.g. by SA; microgametogenesis; pollen sperm cell differentiation; root hair cell development; root hair cell differentiation and elongation
CCGCG 229	CGCGT 226; 229	CG-1; CAMTA	*CAMTA2;* At5g64220	induced e.g. by SA; regulation of transcriptional activity in response to calcium signals; binding calmodulin in calcium-dependent manner; freezing tolerance; positive regulation of general stress response; tolerance to aluminium
CCGCG 229	CGCGT 229	CAMTA	*CAMTA4;* At1g67310	induced e.g. by SA; regulation of transcriptional activity in response to calcium signals; binding to calmodulin in calcium-dependent manner; positive regulation of general stress response
			CAMTA6; At3g16940	induced e.g. by SA; binding of calmodulin in calcium-dependent manner; regulation of transcriptional activity in response to calcium signals
TTTAT 62; 667		CSD	*CSP2;* At4g38680	induced e.g. by SA; chaperone that unwind nucleic acid duplex; acceleration of seed germination and seedling growth under cold stress; enhancement of freezing tolerance; flowering transition; flower and seed development; seed germination under salt stress regulation of respiratory oxygen uptake
TGACT 5; 832 TTGAC 830; 831	**AGTCA** 111; **1240** **GTCAA** 913; 1147; **1251**	WRKY	*WRKY6;* At1g62300	induced e.g. by SA; control of processes related to senescence and pathogen defense; cellular response to boron-containing substance deprivation; response to chitin; cellular response to

			phosphate starvation; modulation of phosphate homeostasis and Pi translocation
		WRKY40; At1g80840	responded to SA; defense response to bacterium, fungus; response to chitin and wounding
		WRKY4; At1g13960	induced e.g. by SA; positive role in resistance to necrotrophic pathogens (e.g. *Botrytis cinerea*); negative effect on plant resistance to biotrophic pathogens (e.g. *Pseudomonas syringae*)
		WRKY60; At2g25000	responded to SA; regulation of: defense response to bacterium and fungus
		WRKY21; At2g30590	induced by SA; glucosinolate metabolic process; calmodulin binding
		WRKY54; At2g40750	induced e.g. by SA; negative regulation of leaf senescence; prevention of stomatal closure and, consequently, osmotic stress tolerance; promotion of brassinosteroid-regulated plant growth; prevention of defense response to necrotrophic pathogens *Pectobacterium carotovorum* and *B. cinerea*; promotion of defense against biotrophic/hemibiotrophic pathogens *Pseudomonas syringae* pv. tomato
TTGAC 831 TGACT 5; 832	**AGTCA** 111; **1240** **GTCAA** 913; 1147; **1251**	*WRKY70;* At3g56400	responded e.g. to SA; senescence; biotic and abiotic stress responses by modulation of various phytohormones signaling pathways; positive regulation of SA-mediated signal pathway; negative regulation of jasmonic acid-mediated signal pathway; prevention of defense response to necrotrophic pathogens *P. carotovorum* and *B. cinerea*; promotion of defense responses against biotrophic/hemibiotrophic SA-monitored pathogens (e.g. *P. syringae*, *Erwinia carotovora* subsp. *carotovora* and *Erysiphe cichoracearum*); defense response toward insects (e.g. *Plutella xylostella* and *Brevicoryne brassicae*); repression of biosynthesis of phytoalexin, camalexin, indol-3-ylmethyl glucosinolate and SA; promotion of brassinosteroid-regulated plant growth; prevention of stomatal closure and, consequently, osmotic stress tolerance; cellular response to hypoxia
		WRKY53; At4g23810	induced by SA; regulation of early events of leaf senescence; regulation of jasmonic acid-mediated signaling pathway; defense response to bacterium; response to chitin

	WRKY18; At4g31800	induced e.g. by SA; regulation of defense response to bacterium and fungus; response to chitin
	WRKY26; At5g07100	induced e.g. by SA; positive regulator of plant thermotolerance by participating in ethylene-response signal transduction pathway
	WRKY38; At5g22570	involved in SA mediated signaling pathway; defense response to bacterium
	WRKY30; At5g24110	induced by SA; leaf senescence
	WRKY8; At5g46350	induced e.g. by SA; positive regulator of salt stress response; negative regulation of basal resistance to bacterial pathogen *P. syringae*; positive regulation of resistance to fungal pathogen *B. cinerea*; positive regulation of defense response againt tobamovirus (TMV)

[a] For TFBSs only most conserved positions within a matrix were listed. [b] Binding sites localised in proximal promoter region are in bold. [c] The TSS is located at 1500 nucleotide of the studied promoter sequence.

Dorobek naukowy

Publikacje będące podstawą rozprawy doktorskiej

Majewska M., Szymczyk P., Gomulski J., Jeleń A., Grąbkowska R., Balcerczak E., Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4* gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354

IF_{2021} 4,927; $MEiN_{2021}$ 140

Majewska M., Kuźma Ł., Szymczyk P. „Isolation and comprehensive *in silico* characterisation of a new *3-hydroxy-3-methylglutaryl-coenzyme A reductase 4* (*HMGR4*) gene promoter from *Salvia miltiorrhiza*: comparative analyses of plant *HMGR* promoters" Plants, 2022, 11, 1861, doi: 10.3390/plants11141861

IF_{2021} 4,658; $MEiN_{2021}$ 70

Majewska M., Wysokińska H., Kuźma Ł., Szymczyk P. „Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview" Gene, 2018, 644, 38–48, doi: 10.1016/j.gene.2017.10.079

IF_{2018} 2,638; $MNiSW_{2016}$ 20

Pozostały dorobek naukowy

Publikacje naukowe

Szymczyk P., Kuźma Ł., Jeleń A., Balcerczak E., **Majewska M.** „Isolation of *Salvia miltiorrhiza* kaurene synthase-like (KSL) gene promoter and its regulation by ethephon and yeast extract" Genes, 2023, 14, 54, doi: 10.3390/genes14010054

IF_{2021} 4,141; $MEiN_{2021}$ 100

Szymczyk P., Szymańska G., **Majewska M.**, Weremczuk-Jeżyna I., Kołodziejczyk M., Czarnecka K., Szymański P., Kochan E. „Homology modelling and docking studies of strictosidine β-D-glucosidase from Madagascar periwinkle (*Catharanthus roseus* Bunge)" International Journal of Bioinformatics Research and Applications, 2022, 18 (3), 234–269, doi: 10.1504/IJBRA.2022.10033607

$MEiN_{2021}$ 20

Majewska M., Lewandowska U. „The chemopreventive and anticancer potential against colorectal cancer of polyphenol-rich fruit extracts" Food Reviews International, 2018, 34 (4), 390–409, doi: 10.1080/87559129.2017.1307388

IF_{2018} 3,933; $MNiSW_{2016}$ 30

Rosset I., Strapagiel D., Sitek A., **Majewska M.**, Ostrowska-Nawarycz L., Żądzińska E. „Association of *FTO* and *TMEM18* polymorphisms with overweight and obesity in the

population of Polish children" Anthropological Review, 2016, 79 (1), 17–33, doi: 10.1515/anre-2016-0002

$MNiSW_{2016}$ 15

Sitek A., Rosset I., Strapagiel D., **Majewska M.**, Ostrowska-Nawarycz L., Żądzińska E. „Association of *FTO* gene with obesity in Polish schoolchildren" Anthropological Review, 2014, 77 (1), 33–44, doi: 10.2478/anre-2014-0003

$MNiSW_{2014}$ 7

Patenty

Strapagiel D., **Majewska M.**, Słomka M., Janik K., Sobalska M., Bartosz G. „Sposób ustalania płci" Biuletyn Urzędu Patentowego RP, 2015, 13, 1082, 406569

Komunikaty zjazdowe

Majewska M., Kochan E., Szymańska G., Wysokińska H., Szymczyk P. „Struktura i klonowanie promotora genu 4 reduktazy 3-hydroksy-3-metyloglutarylo-koenzymu A z szałwii czerwonokorzeniowej − *Salvia miltiorrhiza* (*SmHMGR4*)", I Biotechnologiczna Konferencja Naukowa, Biotechnologia roślin − perspektywy i wyzwania, 9 czerwca 2017, Warszawa

Majewska M., Podsędek A., Owczarek K., Fichna J., Lewandowska U. „Differentiated impact of digested and non-digested red cabbage extracts on the growth of human colon cancer cell lines", X Parnas Conference, Young Scientist Forum, Molecules in the Living Cell and Innovative Medicine, 10–12 lipca 2016, Wrocław

Majewska M., Strapagiel D., Rosset I., Sitek A., Żądzińska E., Bartosz G. „Wpływ czynników genetycznych na ryzyko wystąpienia nadwagi i otyłości w populacji polskiej", IV Polski Kongres Genetyki, 10–13 września 2013, Poznań

Majewska M., Strapagiel D., Bartosz G. „Single nucleotide polymorphisms of *ABCB1* and *ABCG2* genes in Polish cancer patients", The role of MDR proteins in pharmacokinetics and toxicology, 3–6 września 2013, Ryn

Majewska M., Sobalska M., Siewierska A., Strapagiel D., Bartosz G. „Single nucleotide polymorphisms of selected *ABCC* genes in the Polish population", 4th FEBS Special Meeting, ATP-Binding Cassette (ABC) Proteins: From Genetic Disease to Multidrug Resistance, 3–9 marca 2012, Innsbruck

Majewska M., Soliman A., Leśnikowski Z.J. „Aktywność przeciwwirusowa wybranych pochodnych kumaryny", II Lubelskie Dni Wirusologiczne, 22–24 września 2010, Nałęczów

Majewska M., Paradowska E., Olejniczak A.B., Studzińska M., Wojtczak B.A., Piskała A., Laskowski A., Leśnikowski Z.J. „Screening laboratory virology and bacteriology", BioForum, 19–21 maja 2010, Łódź

Majewska M., Studzińska M., Wojtczak B.A., Andrysiak A., Paradowska E., Olejniczak A.B., Leśnikowski Z.J. „Cytotoksyczność nukleozydów modyfikowanych klasterami boru", 52. Zjazd PTChem i SITPChem, 12–16 września 2009, Łódź

Nagrody i wyróżnienia

Wyróżnienie w konkursie InnoTechAkademia II BioTechScience – droga do własnych marzeń za pracę: Strapagiel D., Janik K., **Majewska M.**, Słomka M., Sobalska M. „Identyfikacja płci z użyciem technik z zakresu biologii molekularnej – masowa analiza krzywych topnienia produktów PCR w niskiej rozdzielczości dla genów *AMELY*, *SRY*, *ABCG2*", 19 grudnia 2013

Oświadczenie autora rozprawy doktorskiej

UNIWERSYTET MEDYCZNY W ŁODZI

Małgorzata Majewska
(imię i nazwisko Kandydata)

Łódź, 09.02.2023

Przewodnicząca Rady Nauk
Farmaceutycznych/~~Prodziekan ds. Nauki*~~
~~Uniwersytetu Medycznego w Łodzi~~

OŚWIADCZENIE AUTORA ROZPRAWY DOKTORSKIEJ

Świadomy/a odpowiedzialności prawnej oświadczam, że niniejsza rozprawa doktorska została przygotowana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że prezentowana przeze mnie rozprawa doktorska nie była wcześniej podmiotem procedur związanych z uzyskaniem stopnia naukowego w żadnej wyższej uczelni, instytucie badawczym i/lub jednostce badawczo-rozwojowej.

09.02.2023 Małgorzata Majewska
(data i podpis Kandydata)

* niewłaściwe skreślić

Oświadczenia współautorów publikacji wchodzących w skład rozprawy doktorskiej

mgr Małgorzata Majewska
Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Wysokińska H., Kuźma Ł., Szymczyk P. „Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview" Gene, 2018, 644, 38-48, doi: 10.1016/j.gene.2017.10.079
polegał na: opracowaniu koncepcji artykułu, zebraniu literatury naukowej, przygotowaniu tabel, napisaniu manuskryptu, korespondencji z czasopismem, przygotowaniu odpowiedzi na recenzje, końcowej korekcie manuskryptu.

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Kuźma Ł., Szymczyk P. „Isolation and comprehensive *in silico* characterisation of a new 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 (*HMGR4*) gene promoter from *Salvia miltiorrhiza*: comparative analyses of plant *HMGR* promoters" Plants, 2022, 1861, doi: 10.3390/plants11141861
polegał na: opracowaniu koncepcji badań, przeprowadzeniu analiz *in silico*, interpretacji wyników, przygotowaniu rycin i tabel, zebraniu literatury naukowej, napisaniu manuskryptu, korespondencji z czasopismem, przygotowaniu odpowiedzi na recenzje, końcowej korekcie manuskryptu.

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Szymczyk P., Gomulski J., Jeleń A., Grąbkowska R., Balcerczak E., Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354

polegał na: opracowaniu koncepcji badań, założeniu kultur *in vitro S. miltiorrhiza*, przeprowadzeniu analiz *in silico*, zaprojektowaniu konstruktu nadekspresyjnego oraz starterów do reakcji PCR i real-time PCR, optymalizacji zastosowanych metod badawczych, wykonaniu części eksperymentalnej, analizie statystycznej wyników, interpretacji wyników, przygotowaniu rycin i tabel, zebraniu literatury naukowej, napisaniu manuskryptu, korespondencji z czasopismem, przygotowaniu odpowiedzi na recenzje, końcowej korekcie manuskryptu.

Małgorzata Majewska

dr hab. n. farm. Łukasz Kuźma, prof. UM
Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Wysokińska H., **Kuźma Ł.**, Szymczyk P. „Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview". Gene, 2018, 644, 38-48, doi: 10.1016/j.gene.2017.10.079
polegał recenzji i korekcie manuskryptu.

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., **Kuźma Ł.**, Szymczyk P. „Isolation and comprehensive *in silico* characterisation of a new 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 (*HMGR4*) gene promoter from *Salvia miltiorrhiza*: comparative analyses of plant *HMGR* promoters". Plants, 2022, 1861, doi: 10.3390/plants11141861
polegał na: nadzorze merytorycznym nad badaniami, recenzji i korekcie manuskryptu.

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Szymczyk P., Gomulski J., Jeleń A., Grąbkowska R., Balcerczak E., **Kuźma Ł.** „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones". Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354
polegał na: opracowaniu koncepcji badań, wsparciu w założeniu kultur *in vitro S. miltiorrhiza*, opracowaniu i optymalizacji metody UHPLC oraz ocenie zawartości tanszinonów w ekstraktach *S. miltiorrhiza*, interpretacji wyników, nadzorze merytorycznym nad badaniami, recenzji i korekcie manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułów jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. prac był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów ich powstawania.

PROFESOR UCZELNI
Zakład Biologii i Botaniki Farmaceutycznej
Katedra Biologii i Biotechnologii Farmaceutycznej
Uniwersytetu Medycznego w Łodzi

.....
dr hab. n. farm. Łukasz Kuźma

dr n. biol. Piotr Szymczyk
Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Wysokińska H., Kuźma Ł., **Szymczyk P.** „Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview" Gene, 2018, 644, 38-48, doi: 10.1016/j.gene.2017.10.079
polegał na: nadzorze merytorycznym, recenzji i korekcie manuskryptu.

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Kuźma Ł., **Szymczyk P.** „Isolation and comprehensive *in silico* characterisation of a new 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 (*HMGR4*) gene promoter from *Salvia miltiorrhiza*: comparative analyses of plant *HMGR* promoters" Plants, 2022, 1861, doi: 10.3390/plants11141861
polegał na: opracowaniu koncepcji badań, założeniu hodowli *S. miltiorrhiza*, izolacji sekwencji promotora genu *HMGR4 S. miltiorrhiza*, interpretacji wyników, nadzorze merytorycznym nad badaniami, recenzji i korekcie manuskryptu.

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., **Szymczyk P.**, Gomulski J., Jeleń A., Grąbkowska R., Balcerczak E., Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354
polegał na: opracowaniu koncepcji badań, interpretacji wyników, nadzorze merytorycznym nad badaniami, recenzji i korekcie manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułów jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. prac był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów ich powstawania.

ADIUNKT
Zakładu Biologii i Botaniki Farmaceutycznej
Uniwersytetu Medycznego w Łodzi
dr n. biol. Piotr Szymczyk

.....

prof. dr hab. Halina Wysokińska
Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., **Wysokińska H.**, Kuźma Ł., Szymczyk P. „Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview" Gene, 2018, 644, 38-48, doi: 10.1016/j.gene.2017.10.079
polegał na recenzji i korekcie manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułu jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. pracy był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów jej powstawania.

.....Halina Wysokińska.....

mgr Jan Gomulski
Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

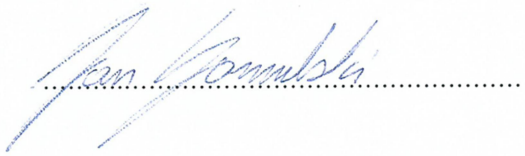
Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Szymczyk P., **Gomulski J.**, Jeleń A., Grąbkowska R., Balcerczak E., Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354
polegał na: optymalizacji metody UHPLC i obsłudze aparatury.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułów jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. prac był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów ich powstawania.



.....

dr n. farm. Renata Grąbkowska
starszy specjalista ds. badań laboratoryjnych
Zakład Biologii i Botaniki Farmaceutycznej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023 r.

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu:
Majewska M., Szymczyk P., Gomulski J., Jeleń A., **Grąbkowska R.**, Balcerczak E., Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354
polegał na wsparciu merytorycznym w zakresie optymalizacji i prowadzenia hodowli *S. miltiorrhiza*.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułu jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. pracy był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów jej powstawania.

10.02.23r. Renata Grąbkowska

dr n. farm. Agnieszka Jeleń
Zakład Biochemii Farmaceutycznej i Diagnostyki Molekularnej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Szymczyk P., Gomulski J., **Jeleń A.**, Grąbkowska R., Balcerczak E., Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354
polegał na: wsparciu merytorycznym w zakresie planowania doświadczeń, optymalizacji, przeprowadzania i analizy wyników reakcji PCR i real-time PCR oraz analizy statystycznej zawartości tanszinonów w ekstraktach *S. miltiorrhiza*, recenzji i korekcie manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułów jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. prac był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów ich powstawania.

Zakład Biochemii Farmaceutycznej i Diagnostyki Molekularnej
Uniwersytetu Medycznego w Łodzi
Agnieszka Jeleń

.....

prof. dr hab. n. farm. Ewa Balcerczak
Zakład Biochemii Farmaceutycznej i Diagnostyki Molekularnej
Wydział Farmaceutyczny
Uniwersytet Medyczny w Łodzi

Łódź, 09.02.2023

OŚWIADCZENIE

Oświadczam, że mój udział w powstaniu artykułu
Majewska M., Szymczyk P., Gomulski J., Jeleń A., Grąbkowska R., **Balcerczak E.**, Kuźma Ł. „The expression profiles of the *Salvia miltiorrhiza* 3-hydroxy-3-methylglutaryl-coenzyme A reductase 4 gene and its influence on the biosynthesis of tanshinones" Molecules, 2022, 27, 4354, doi: 10.3390/molecules27144354
polegał na: wsparciu merytorycznym w zakresie planowania doświadczeń, optymalizacji, przeprowadzania i analizy wyników reakcji PCR i real-time PCR, recenzji i korekcie manuskryptu.

Jednocześnie wyrażam zgodę na przedłożenie przez mgr Małgorzatę Majewską ww. artykułów jako części rozprawy doktorskiej przygotowanej w formie spójnego tematycznie zbioru artykułów opublikowanych w czasopismach naukowych.

Oświadczam, że wkład mgr Małgorzaty Majewskiej w powstanie ww. prac był dominujący, łatwy do wyodrębnienia i dotyczył wszystkich etapów ich powstawania.

KIEROWNIK
Katedry Diagnostyki Laboratoryjnej i Molekularnej
Uniwersytetu Medycznego w Łodzi
.....
Prof. dr hab. n. farm. Ewa Balcerczak