Molecular and biochemical studies of fragrance biosynthesis in rose

Sun, P.

Citation for published version (APA):
Chapter 6

General discussion
In this chapter, the research questions are reassessed and the outcomes of the experiments are discussed. Chapter 2 and 3 addressed the identification of NUDX1 genes in rose and the biological function of the corresponding NUDX1 proteins. Chapter 4 provided some information on the transcriptionally regulation of these NUDX1 genes. Finally, the RNA-Seq analysis in Chapter 5 provided new information regarding scent-related genes, which could be used for future studies. The findings in this thesis are summarized in Figure 1. Briefly, rose NUDX1 proteins were found to be responsible for the biosynthesis of both monoterpenoids and sesquiterpenoids. The production of geraniol and E,E-farnesol strongly correlated with the transcript levels of respective NUDX1 genes, which might be regulated by bHLH79 transcription factors (TF). In addition, there are indications that the expression level of the NUDX1 genes is dependent on repeated regions in the promoter. RNA-Seq analysis of four rose lines, two of which produced geraniol, but not E,E-farnesol and the other two produced E,E-farnesol but not geraniol, identified more potential TF candidates that could participate in the regulation of rose NUDX1 genes. With the annotated transcripts, in combination with the scent profile of each rose, more scent-related genes were identified for future study.

**Figure 1.** Summary of the biosynthesis pathways of monoterpenoids and sesquiterpenoids in roses and the regulation thereof. Thick arrows indicate catalytic reactions while the double arrows indicate multiple steps. Purple thin arrow indicates the transport of molecules, the blue thin arrows indicate transcription and translation processes, and the red arrows indicate protein-promoter interactions. The filled boxes (in purple) represent promoters while the red filled boxes represent oligonucleotide repeats within the promoter region. Question marks represent the unknown or unverified steps and/or proteins.

*NUDX1* gene was identified in the scented rose cultivar, *Rosa x hybrida* cv. ‘Papa Meillard’ (PM) by comparing its transcriptome with that of an unscented rose, *R. x hybrida* cv. ‘Rouge Meillard’ (RM).
Subsequently, when analysing the progeny of two rose cultivars, *R. chinensis* cv. ‘Old Blush’ (OB) and *R. x wichurana* (Rw), which have distinct scent profiles, another five NUDX1 genes were found. Two NUDX1 genes were found in Rw, and three were found in OB, although only one was expressed in the petals of each cultivar. The expressed NUDX1 gene in PM is named *RhNUDX1-1*, the one in OB is named *RcNUDX1-1a* and the one in Rw is named *RwNUDX1-2*. *RhNUDX1-1* and *RcNUDX1-1a* are orthologs and *RwNUDX1-2* is a paralog of *RhNUDX1-1*. Both *RhNUDX1-1* and *RcNUDX1-1a* were found to be responsible for the production of monoterprenoids, especially geraniol, while *RwNUDX1-2* was involved in the production of sesquiterpenoids, especially *E,E*-farnesol (Figure 1). *In vitro* enzyme activity assays showed that rose NUDX1 proteins had distinct functions compared to the NUDX1 of *Arabidopsis thaliana* (AtNUDX1). AtNUDX1 not only can hydrolyse farnesyl diphosphate (FPP), geranyl diphosphate (GPP) and isopentenyl diphosphate (IPP), but also hydrolys 8-oxodeoxyguanosine triphosphate (8-oxo-dGTP). However, none of the rose NUDX1 proteins can hydrolyse 8-oxo-dGTP. In other *Rosaceae* plants, whose genomic sequences are available, four to five NUDX1 genes were found. The presence of multiple copies of NUDX1 in the genome is not common across the plant kingdom, most of the plants only have one or two NUDX1 genes, such as *A. thaliana*, *Solanum lycopersicum* and *Nicotiana tabacum*, although in the pepper (*Capsicum annuum*) genome, seven NUDX1 were identified (sequences were obtained from Nucleotide database of National Center for Biotechnology Information). Loss of function of hydrolysing oxidized nucleotides (e.g. 8-oxo-dGTP), or more precisely, in this case, diversification of function, may be a result of gene duplication events. It is known that the duplicated gene can conserve its original function, or be sub-functionalized, evolve a novel function or even become a pseudogene (Magadum et al., 2013, Zhang, 2003). Gene duplication events resulting in changes of gene function have also been found in other plants. For example, four homologous genes of oxidosqualene cyclase (OSC) were found in *A. thaliana*, but they were involved in the biosynthesis of different products: *At1g78950* and *At1g78960* were involved in the production of β-amyrin (Husselstein-Muller et al., 2001, Kushiro et al., 2000, Shibuya et al., 2009), the *At1g78955* encoding protein mainly produced camelliol C (Kolesnikova et al., 2007), and *At1g78970* was involved in the biosynthesis of lupeol (Herrera et al., 1998). They share about 85% similarity and the diversification of their function is likely a result of tandem duplication (Xue et al., 2012). As multiple NUDX1 genes are also found in several *Rosaceae* plants, it would be interesting to know when the diversification of NUDX1 genes started. First, it would be important to know whether the function of geraniol/*E,E*-farnesol biosynthesis is well conserved within the *Rosa* genus. The correlation of expression of NUDX1-1 and NUDX1-2 gene with the production of geraniol and *E,E*-farnesol should be examined among the wild roses as well as modern roses. In addition, a complete rose genome could definitively provide very useful information on the total number of NUDX1 genes in rose, as well as their chromosomal location. Besides, when comparing the rose NUDX1 amino acid sequences with those of other *Rosaceae* plants, two NUDX1 in *Fragaria vesca* subsp. *vesca*, were found to be highly similar to rose NUDX1. FvNUDX1-1 (XP_004304338) is 79% similar to rose NUDX1-2 and FvNUDX1-3 (XP_004297155) is 87%-88% similar to rose NUDX1-1 (Chapter 3). It is important to know the expression of these three *F. vesca* NUDX1 genes as well as their functions to provide further information on the evolution of NUDX1 proteins. Besides, neither a geraniol synthase nor an *E,E*-farnesol synthase has been identified in *F. vesca*, which makes the characterization of FvNUDX1 becoming more appealing. In addition, systematic phylogenetic studies in combination with experimental characterization of NUDX1 proteins across plant kingdom would definitively provide valuable insights into NUDX1 evolution.

*In vitro* enzymatic assays showed that all rose NUDX1 proteins could accept GPP and FPP as substrates, but only RwNUDX1-2 showed preference towards FPP. This substrate preference can be partially explained by the difference in amino acid sequence, although the substrate affinity also depends on the other factors, such as pH, presence or absence of certain cofactors, which can lead to changes in protein structure. Besides, the protein conformation in the form of a crystal may not
be a very good representation of the enzyme in solution. Proteins normally function in aqueous solutions and can be influence by many other molecules in the environment (Ladbury, 1996). Furthermore, the protein crystal structure was obtained without presence of the cofactors (e.g. Mg$^{2+}$), which is crucial to the activation of the NUDX1 protein (Mildvan et al., 2005), in order to slow down the reaction and stabilize the ligand-bound protein. Therefore, protein crystallization with the presence of co-factors in combination with the structural refinement using protein dynamic modelling would be very useful to obtain a more precise and “realistic” protein structure for the subsequent analysis of substrate-recognition residues. Besides, nuclear magnetic resonance (NMR) would be an alternative solution to determine the protein structure in solution (Berg et al., 2002) and to study the protein-ligand interaction (Calà et al., 2014). Moreover, mass spectrometry-based methods are also developed to provide additional information to assist the protein structure determination (Walzthoeni et al., 2013). In order to determine whether some amino acids in the active site are important in determining substrate affinity of NUDX1, site-directed mutagenesis experiments were conducted and the mutated enzymes were successfully purified and tested. However, the mutated enzymes did not show any changes in affinity. Probably this was because the target residues were selected based on the difference in amino acid sequences before the protein crystal structure came available. Thanks to the crystallization of RhNUDX1-1 and structural prediction of RwNUDX1-2, more reliable mutation points, such as substitution of Ser135 to Trp135 (from RhNUDX1-1) in RwNUDX1-2, are proposed for the future mutagenesis experiment to see whether this could change the affinity for GPP or FPP of the mutant proteins.

In vivo assays presented a different result compared to the in vitro assays. In vitro enzymatic assays showed that all rose NUDX1 proteins could hydrolyse GPP and FPP. However, in vivo, the expression of RhNUDX1-1 in both Rose and N. benthamiana only led to the increase of geraniol or geraniol glycosides only, while the expression of RwNUDX1-2 in N. benthamiana resulted in the production of both geranyl- and farnesyl-glycosides. This observation could be explained by the fact that introduction of rose proteins into N. benthamiana is an artificial system and GPP and FPP levels may differ in the different tissues. More interestingly, it indicated that not only a pool of FPP, but also a pool of GPP is available in both Rose petals and N. benthamiana. It is known that RhNUDX1-1 is located in the cytosol when transiently expressed in N. benthamiana leaves. Although the location of RwNUDX1-2 remains unclear, it is very likely to be located in cytosol as well, considering the similarity between RwNUDX1-2 and RhNUDX1-1. GPP is conventionally considered to be synthesized in plastid, but there is evidence showing that GPP can be transported from plastid to cytosol and provide the resources for cytosolic monoterpene synthases (see Chapter 1), although no GPP transporter has been discovered so far (see above in Figure 1). Another hypothesis is that there may be cytosolic GPP synthases in rose and in N. benthamiana or GPP could be generated as a by-product during the biosynthesis of FPP. A recent study showed that there could be a cytosolic GPP pool available in N. benthamiana. The truncated geraniol synthase of Valeriana officinalis (VoGES), which was targeted to cytosol, was introduced into N. benthamiana leaves using the 35S cauliflower mosaic virus (CaMV) promoter resulting in the production of geraniol (Dong et al., 2016). However, the co-infiltration of plastid-targeted GPP synthase of Picea abies (PaGDP) with cytosol-targeted VoGES did not increase the production of geraniol (Dong et al., 2016). These data suggest that the plastidial GPP was not transported to the cytosol in N. benthamiana or the transported GPP amount was so low that no significant change in geraniol production was observed in the presence of cytosolic VoGE$S$. This also indicates that there is in situ production of cytosolic GPP in N. benthamiana, which might be the case for roses as well. Due to the difficulty of transforming roses, isotope-labelled feeding experiments would be a useful alternative approach for studying the source of GPP in rose. A similar study has been applied in the fruits of F. x ananassa and F. vesca in order to trace the biosynthesis pathways of (S)-linalool, trans-(S)-nerolidol and (−)-α-pinene (Hampel et al., 2006).
In order to gain more insight into the mechanism that affects the production of rose volatiles, I started to investigate the transcriptional regulation of the rose NUDX1 genes. The promoters of \( \text{RcNUDX1-1a} \), \( \text{RcNUDX1-1b} \) and \( \text{RwNUDX1-2} \) were cloned and used to drive a reporter gene (GUS, \( \beta \)-glucuronidase) in rose petals using transient expression. The results indicated that all three promoters could activate the GUS reporter in rose petals, suggesting the regulation of NUDX1 genes was not the result of inactivation of their promoters. The level of promoter activation remains a question because the quantification failed due to the difficulty in getting protoplasts from rose petals. However, an interesting feature was found in the promoter of \( \text{RcNUDX1-1a} \): several repetitive sequences are present, suggesting that these repeats could serve as enhancers. These repeats were not present in the promoter of \( \text{RwNUDX1-2} \) or \( \text{RcNUDX1-1b} \) and the gene expression data showed that \( \text{RwNUDX1-2} \) had a lower expression level than \( \text{RcNUDX1-1a} \) (Chapter 3 Figure 2). These repeats could have a role in enhancing the expression of the downstream gene. In order to study the role of these repeat sequences in expression of \( \text{RcNUDX1-1a} \), generating a series of promoter-reporter constructs by removing the repeats one by one would be a way to test its functions as well as the role of the repeat numbers.

By comparing the promoter sequences, a canonical E-box was found in the promoter of \( \text{RcNUDX1-1a} \), suggesting a possible binding site for bHLH proteins. In addition, two \( \text{bHLH79} \) transcription factors (TF) were found that were co-expressed with \( \text{RcNUDX1-1a} \) and \( \text{RwNUDX1-2} \), respectively, in the RNA-sequencing (RNA-Seq) experiments. The RNA-Seq analysis was conducted on four rose individuals: OB, Rw and two of their \( F_1 \) siblings, OW9035 and OW9047. OB and OW9047 produce geraniol, but no \( E,E \)-farnesol, while Rw and OW9035 produce \( E,E \)-farnesol, but no geraniol. According to the cluster analysis of the RNA-Seq data, a \( \text{bHLH79} \) TF clustered with \( \text{RcNUDX1-1a} \), which was then named \( \text{RcbHLH79} \) (c42493_g1_i2) and the other \( \text{bHLH} \) TF that clustered with \( \text{RwNUDX1-2} \) was called \( \text{RwbHLH79} \) (c42493_g1_i4). These cDNAs shared more than 95% sequence similarity. Their function was subsequently tested by co-expressing them with the different promoter constructs into \( N. \text{benthamiana} \) leaves under the control of the CaMV 35S promoter. The results showed that none of the TFs strongly trans-activated any of the NUDX1 promoter-GUS constructs (Chapter 4). Although no significant effect was found with the tested TFs, an interesting phenomenon was observed. When comparing the transactivation effect between the promoters \( \text{pOB1a} \), \( \text{pOB1b} \) and \( \text{pRw} \), \( \text{pOB1b} \) showed a higher activity compared to \( \text{pOB1a} \) and \( \text{pRw} \), while \( \text{pOB1a} \) and \( \text{pRw} \) showed similarly low activity (see Chapter 4, Figure 5). A hypothesis rises: \( \text{pOB1b} \) may contain a promoter element that can be recognized in \( N. \text{benthamiana} \) leaves, while \( \text{pOB1a} \) and \( \text{pRw} \) do not possess such element, or \( \text{pOB1a} \) and \( \text{pRw} \) have rose petal-specific promoter elements, which could be inhibited or could not be recognized in \( N. \text{benthamiana} \) leaves while \( \text{pOB1b} \) does not have such elements. However, this hypothesis should be verified by additional experiments, such as quantification of the activity of all these promoters in rose petals. Surprisingly, \( \text{RwNUDX1-2} \) was able to trans-activate the promoter of the tomato TPS5 gene, indicating that bHLH can play a role in regulation of the expression of biosynthetic genes.

Expression analysis on these two bHLH TFs in 12 rose \( F_1 \) individuals showed that no correlation could be found between the expression of either of the two bHLH TFs and \( \text{NUDX1} \) expression. In fact, when checking the assembled RNA-Seq transcripts, five transcript variants of \( \text{bHLH79} \), including \( \text{RcbHLH79} \) and \( \text{RwbHLH79} \), were found. Except the two selected TFs that had differential expression across the four individuals, the rest of the homologs were expressed in all four roses and no significant difference was found between individuals. Therefore, it is difficult to conclude whether the insignificant transactivation of GUS activity was due to the plant material used (\( N. \text{benthamiana} \) leaves) or the fact that these two selected TF candidates are not the TFs that are involved in regulating \( \text{NUDX1} \) gene expression. Correlation analysis should be done on the expression of TF candidate genes and the \( \text{NUDX1} \) genes, under a condition that the designed primers should be able to differentiate between the homologs. A newly developed high affinity PCR technique, called Zip
Nuclei acids (ZNAs), could be employed for differentiation of TF transcript variants during RT-qPCR. ZNAs are cationic oligonucleotides that have the ability to discriminate between a perfect match and a single base-pair-mismatched sequence (Moreau et al., 2009). By using high annealing temperature but low primer concentration (Moreau et al., 2009), ZNA gene-specific primers can amplify the target gene accurately even it has low expression levels (Moreau et al., 2009). Alternatively, all TF transcript variants could be cloned and tested via an alternative approach – the Yeast one-hybrid (Y1H). Y1H is a molecular biology technique to screen for protein-DNA interaction, especially suitable for studying the interaction between TF and the promoter, as well as the binding site (Reece-Hoyes and Marian Walhout, 2012).

The RNA-Seq study (in Chapter 5) provides useful leads for finding other scent-related genes as well as other TFs that may play a role in the regulation of NUDX1 expression. Furthermore, it may also provide additional information for searching the downstream enzymes that participate in the biosynthesis of geraniol and \( E,E \)-farnesol. Compared to the conventional terpene synthases, the product of NUDX1 catalytic reaction is GP or FP (Chapter 2 and 3), which need to be further hydrolysed by a phosphatase into geraniol (Chapter 2) or \( E,E \)-farnesol. Phosphatases have a wide range of substrates and they are involved in many biological reactions. Therefore, it is very difficult to find the specific phosphatase that hydrolyses GPP and FPP into geraniol and \( E,E \)-farnesol in rose. By using the RNA-Seq technique and analysing the transcriptome of rose petals that have distinct scent profiles, one may be able to identify candidate phosphatases. For example, sequencing the transcriptome of rose that only produce geraniol, the one that only produce \( E,E \)-farnesol, as well as the one that do not produce either of the compounds and comparing the transcriptomes could result in a shorter list of phosphatase candidates, which could be tested using various \textit{in vitro} and \textit{in vivo} assays. In addition, the resulting database could be also used for searching other downstream genes of NUDX1, such as citronellol synthase.
References


