Cloning and characterization of *CroMATE1,* a novel MATE-type transporter from the medicinal plant *Catharanthus roseus* (L.) G. Don

Francisco Anastácio de Abreu e Lima
Mestrado em Biologia Celular e Molecular
Departamento de Biologia
2013

Orientador
Mariana Sottomayor, Auxiliar Professor, Science Faculty, University of Porto

Coorientador
Patrícia Duarte, Post-Doctoral Researcher, IBMC

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

La familia es todo.

- Hector “Tio” Salamanca

Resumo

Catharanthus roseus (L.) G. Don sintetiza e acumula os alcaloides diméricos terpenoides indólicos (TIAs) vincristina e vinblastina, dois dos agentes naturais com maior atividade anticancerígena descobertos até à data. A escassez destes compostos na planta, a par da sua importância farmacológica, tornou *C. roseus* numa das plantas medicinais mais estudadas. Até à data, as folhas desta planta constituem a única fonte dos TIAs anticancerígenos, uma vez que os esforços conducentes à síntese química ou produção *in vitro*, visando uma maior produtividade do processo, se têm revelado infrutíferos. A complexidade inerente ao processo de biossíntese dos TIAs tem vindo a ser progressivamente desvendada. Contudo, o envolvimento de inúmeros passos enzimáticos e a elevada compartimentalização do processo, envolvendo diferentes tipos de células/organelos e portanto vários passos de transporte transmembranar, constituem um entrave face à tão desejada manipulação desta via biossintética. Recentemente, tem-se compreendido que o estudo dos mecanismos e mediadores dos passos de transporte transmembranar das vias metabólicas fornece um complemento essencial às estratégias convencionais da engenharia metabólica com vista ao melhoramento da produtividade de produtos naturais das plantas (Yazaki, Sugiyama et al. 2007).

Os chamados transportadores multi-fármaco (do inglês *multidrug*) têm vindo a surgir como candidatos por excelência aos passos de transporte transmembranar de metabolitos secundários a nível intra- e inter-celular nas plantas, um papel que poderia colmatar e clarificar os mecanismos de fluxo dos TIAs e respetivos precursores em *C. roseus*. Dentro do grupo dos transportadores multi-fármaco encontra-se a família dos transportadores secundários tipo-MATE (*multidrug and toxic compound extrusion*), especialmente numerosa nas plantas, e que foi já implicada no transporte de vários compostos secundários de plantas. Nomeadamente, foram caraterizados vários transportadores tipo-MATE em tabaco, implicados no transporte do alcaloide nicotina para os vacúolos através de um mecanismo de antiporte protónico. Anteriormente a este trabalho, foi observado que a acumulação vacuolar dos TIAs em *C. roseus* ocorre também por um mecanismo de antiporte protónico, o que sustenta o potencial envolvimento de um transportador tipo-MATE no processo (Carqueijeiro, Noronha et al. 2013).

O presente trabalho reúne o isolamento e a caraterização de base computacional e experimental de um gene candidato tipo-MATE de *C. roseus*,

CroMATE1, que codifica uma proteína anteriormente detetada em elevada abundância no proteoma de vacúolos e do tonoplasto de folhas de *C. roseus*, e que constitui um candidato óbvio ao transporte de TIAs para o vacúolo. Com base na sequência dos péptidos identificados no estudo proteómico, foi possível recuperar a sequência codificante completa de *CroMATE1* a partir de bases de dados de transcriptómica de *C. roseus*. Utilizando essa informação, foi efetuado com sucesso o isolamento, clonagem e sequenciação de *CroMATE1.* De forma consistente, as análises preditivas computacionais desvendaram diversas caraterísticas inerentes à sequência codificante de *CroMATE1*, assim como da respetiva proteína, desvendando a identidade deste candidato, a ocorrência de co-regulação com genes envolvidos na biossíntese dos TIAs, e a potencial localização vacuolar. Esta localização foi confirmada experimentalmente, fazendo uso da transformação transiente de protoplastos de mesófilo de *C. roseus* com fusões de CroMATE1 com a proteína verde fluorescente (GFP), que mostraram um direcionamento inequívoco do transportador para o tonoplasto. Este resultado valida as previsões informáticas anteriormente realizadas e confirma o estudo proteómico dos vacúolos, reforçando o papel de CroMATE1 como candidato ao transporte vacuolar dos TIAs em *C. roseus*.

Com as ferramentas desenvolvidas neste trabalho será possível no futuro próximo determinar experimentalmente a função de CroMATE1. A confirmar-se o seu envolvimento na acumulação intravacuolar dos alcaloides de *C. roseus*, terá sido dado um passo importante para a possibilidade de manipulação da via dos TIAs em *C. roseus* para obtenção de níveis mais elevados dos alcaloides anticancerígenos.

Palavras-Chave: *Catharanthus roseus*, alcalóides terpenóides indólicos, transporte transmembranar, transportadores MATE, estratégias de clonagem, localização subcelular, HCL, reconstrução filogenética.

Abstract

Catharanthus roseus (L.) G. Don accumulates the dimeric terpenoid indole alkaloids (TIAs) vincristine and vinblastine, two of the most potent anticancer natural agents discovered to date. The low abundance of these compounds in the plant, together with their pharmacological importance, has made *C. roseus* one of the most studied medicinal plants. To date, the leaves of the plant remain the only source of the anticancer TIAs, since efforts to achieve the chemical synthesis or *in vitro* production, in attempt to enhance the productivity of the process, have been unsuccessful. The inherent complexity of the TIA biosynthesis pathway has been gradually unveiled. However, the existence of many enzymatic steps and the high compartmentalization of the process, involving different types of cells/organelles and therefore several transmembrane transport events, have hindered the possibility for manipulation of this biosynthetic pathway. Recently, it has become clear that the study of the mechanisms and mediators of the transmembrane transport steps of metabolic pathways provides an essential complement to the conventional metabolic engineering strategies aiming at enhancing the productivity of plant natural products (Yazaki, Sugiyama et al. 2007).

The so-called multidrug transporters have been emerging as valuable candidates for the intra- and inter-cellular transmembrane transport steps of secondary metabolites in plants, a role that could fill the gaps and clarify the flux mechanisms of TIAs and precursors in *C. roseus.* Within the multidrug transporters group, the MATE (multidrug and toxic compound extrusion) secondary transporters family is particularly abundant in plants and has already been shown to be implicated in the transport of several secondary compounds in plants. Namely, tobacco MATE-type transporters involved in the vacuolar transport of the alkaloid nicotine by a mechanism of proton-antiport were already characterized. Prior to this work, the vacuolar accumulation of TIAs in *C. roseus* was shown to be also mediated by a proton-antiport mechanism, which supports the potential involvement of a MATE-type transporter in the process (Carqueijeiro, Noronha et al. 2013).

The work hereby described includes the isolation and the computational and experimental-based characterization of a *C. roseus* MATE-type candidate gene, *CroMATE1*, which codifies a protein previously detected with high abundance in the vacuole and tonoplast proteomes from *C. roseus* leaves, thus constituting an obvious candidate to the transport of TIAs into the vacuole. Based upon the sequences of the peptides identified in the proteomic study, it was possible to retrieve the full-length *CroMATE1* coding sequence from *C. roseus* transcriptomic databases. By using such

information, the isolation, cloning and sequencing of *CroMATE1* were successfully achieved. In a consistent fashion, the computational predictive analyses resulted in the identification of a diverse array of characteristics inherent to the coding sequence of *CroMATE1*, as well as to the respective protein, unveiling the identity of this candidate, the occurring co-regulation with genes involved in the TIA biosynthesis, and a potential vacuolar localization. This localization was experimentally confirmed via the transient transformation of *C. roseus* mesophyll protoplasts with fusions of CroMATE1 and a green fluorescent protein (GFP), which unequivocally showed the sorting of the transporter to the tonoplast. This result validates the informatics predictions previously performed and confirms the proteomic study of vacuoles, reinforcing the role of CroMATE1 as candidate for the vacuolar transport of TIAs in *C. roseus*.

With the tools developed in this work it should be possible to determine experimentally the function of CroMATE1 in the near future. If the CroMATE1 involvement in the intra-vacuolar accumulation of *C. roseus* alkaloids is confirmed, this will be an important contribution towards the possibility of manipulating the *C. roseus* TIA pathway to obtain higher levels of the anticancer alkaloids.

Keywords: *Catharanthus roseus*, terpenoid indole alkaloids, transmembrane transport, MATE transporters, cloning strategies, subcellular localization, HCL, phylogenetic reconstruction.

Index

List of Figures

I. Introduction

List of Tables

Abbreviations

aa - amino acid
AGP58 - Arabinogalactan protein 58
BLAST - Basic local alignment search tool
bp - base-pair
CaMV 35S - Cauliflower Mosaic virus promoter
cDNA - Complementary DNA
CDS - Conding sequence (nucleotide)
CFP - Cyan fluorescent protein
CiAP - Calf intestine alkaline phosphatase
D4H - Deacetoxyvindoline-4-hydroxylase
DAT - Deacetylvindoline-4-O-acetyltransferase
ddH2O - Bi-distillated H2O
DNA - Deoxyribonucleic acid
dNTPs - Deoxyribonucleotide triphosphates
EDTA - Ethylenediaminetetraacetic acid
ER - Endoplasmic reticulum
$EtBr_2$ - Ethidium bromide
G10H - Geraniol 10-hydroxilase
GC% - Guanine/Cytosine content
GFP - Green Fluorescent Protein
HCL - Hierarchical co-expression clustering
JA - Jasmonate
kb/kbp - kilo-base-pair
LB - Luria Bertrani
MATE - Multidrug and toxic compound extrusion
MeJA - Methyl jasmonate
MEP - Methylerythritol Phosphate pathway
MPGR - Medicinal Plant Genomics Resource
Mw - Molecular weight
nanoLC-MS/MS - nano liquid chromatography-mass spectrometry/mass spectrometry
nos ter - Terminator sequence of *nopaline synthase* gene from *A.tumefaciens*
ON - Over-night
ORF - Open reading frame
PCR - Polymerase chain reaction

pDNA - plasmid DNA

PEG - Polyethylene glycol

pI - Isoelectric point

PM - Plasma membrane

PRX - Peroxidase

PT - Post-transformation

RNA - Ribonucleic acid

rpm - Revolutions per minute

SGD - strictosidine β-D-glucosidade

sGFP - Green Fluorescent Protein with a S65T mutation

SLS - Secologanin synthase

SNP - Single nucleotide polymorphism

SOTA - Self-organizing tree algorithm

STR - Strictosidine synthase

TDC - Tryptophan decarboxylase

TIA – Terpenoid indole alkaloid

T_m - Melting temperature

TMD - Transmembrane domain

UV - Ultra-violet

v/v – volume/volume

VCR - Vincristine

VLB - Vinblastine

w/v – weight/volume

I. Introduction

I.1. The medicinal plant *Catharanthus roseus* (L.) G. Don

Catharanthus roseus (L.) G. Don is a perennial semi-shrub commonly known as the Madagascar periwinkle and used worldwide as an ornamental species (Fig. I.1). This medicinal plant accumulates in the leaves, among others, the dimeric terpenoid indole alkaloids (TIAs) vinblastine (VLB) and vincristine (VCR), the first natural anticancer products to be clinically used that are still among the most valuable agents used in cancer chemotherapy. The great pharmacological importance, low *in planta* occurrence (around 0.0005% DW), unavailability of synthetic substitutes and high market cost of these alkaloids gave rise to intense research on the TIA pathway, and *C. roseus* has become one of the most studied medicinal plants (van der Heijden, Jacobs et al. 2004, Verpoorte, Lata et al. 2007). In spite of this, much effort is still ongoing to fully understand the basic architecture and regulation of the complex biosynthesis of these TIAs in the *C. roseus* plant and its cultured tissues (Verma, Mathur et al. 2012).

Fig. I.1 - *Catharanthus roseus* (L.) G. Don cv. Little Bright Eyes flowering plant.

I.2. The *C. roseus* terpenoid indole alkaloids

For many years, plant secondary metabolites were thought to be unnecessary waste products of plants primary metabolism, as little was known in regard to their ecological role and their diversity. After gradually acknowledging the underlying diversity, communication and defense properties, and bioactive effects (physiological, deterrent and anti-microbial) on general plant predators, such as herbivorous mammals, insects and pathogenic microorganisms, secondary metabolites were fairly recognized as being a critical component of plant survival strategy. Most importantly, plant secondary metabolism revealed to be an invaluable and diverse source of compounds with pharmacological and biotechnological application, unraveling the so-called phytochemistry field (Hartmann 2007). Within the secondary metabolites, alkaloids constitute a large group of low molecular-weight, nitrogen-containing compounds. Many of the 12,000 alkaloids for which structures have been described function in the defense of plants against herbivores and pathogens. Moreover, some alkaloids exhibit potent biological activity that can be clinically and effectively used for a wide spectrum of disorders and conditions (Facchini 2001).

The medicinal plant *C. roseus* presents a particularly prolific alkaloid metabolism, producing more than 150 different terpenoid indole alkaloids (TIAs), including the anticancer VLB and VCR, the antihypertensive ajmalicine and the sedative serpentine (van der Heijden, Jacobs et al. 2004, Verpoorte, Lata et al. 2007). VLB biosynthesis was shown to be highly complex, involving at least 30 steps from the amino acid tryptophan and the terpenoid geraniol, of which 13 steps have been characterized at the enzyme and/or gene level (Loyola-Vargas, Galaz-Ávalos et al. 2007, Salim, Yu et al. 2013). All the TIAs of *C. roseus* derive from the common precursor strictosidine (Fig. I.2), after which the TIA pathway splits into several branches leading namely to ajmalicine/serpentine, and to vindoline and catharanthine, the leaf abundant monomeric precursors of VLB and VCR (Fig. I.3).

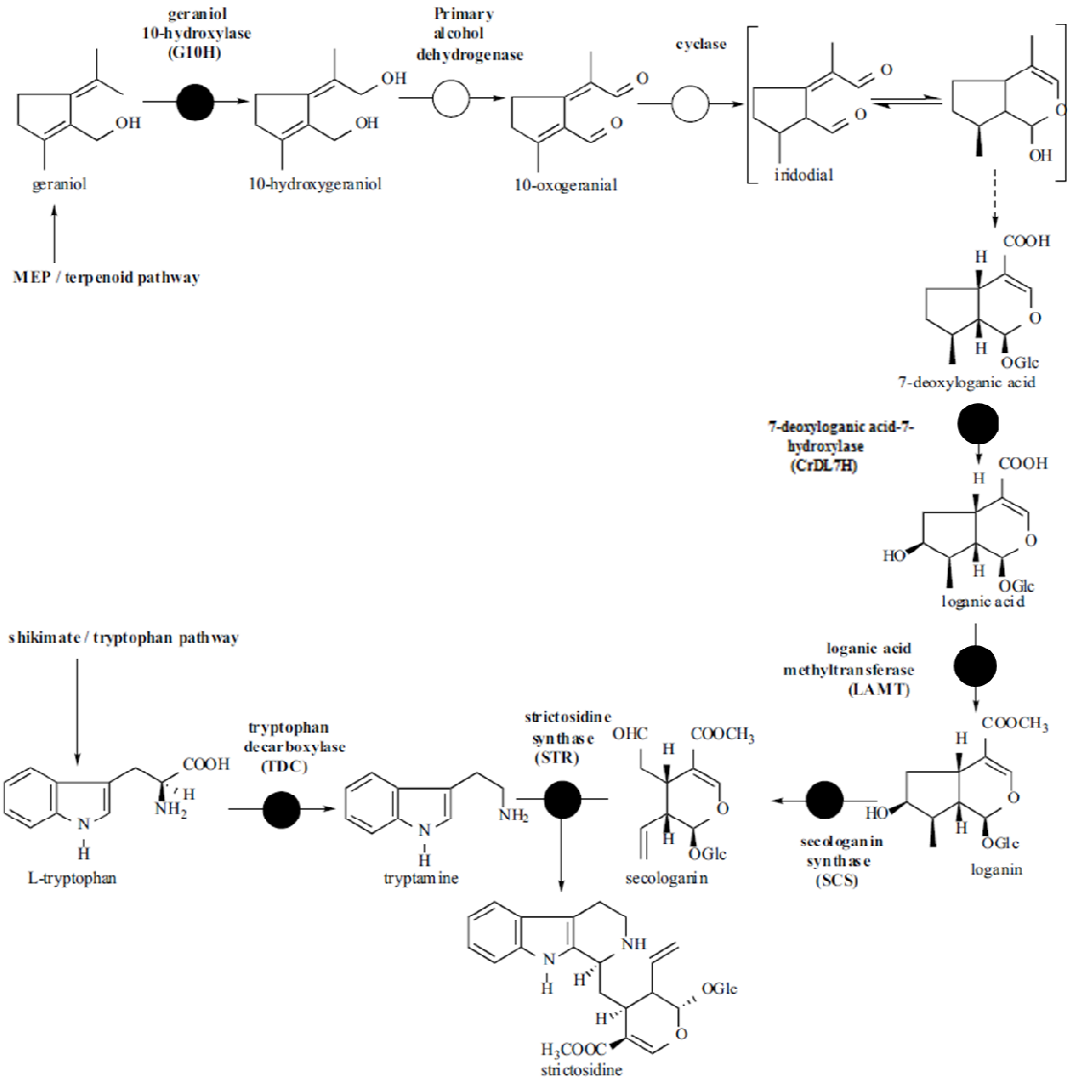


Fig. I.2 - Early steps in the TIA biosynthesis - both the MEP/terpenoid and shikimate/tryptophan pathways contribute to strictosidine synthesis, the precursor of all *C. roseus* TIAs. Closed circles - enzymes already cloned and characterized; dashed lines - conversion by multiple steps; solid lines - conversion by single steps. Adapted from van der Heijden, Jacobs et al. (2004).

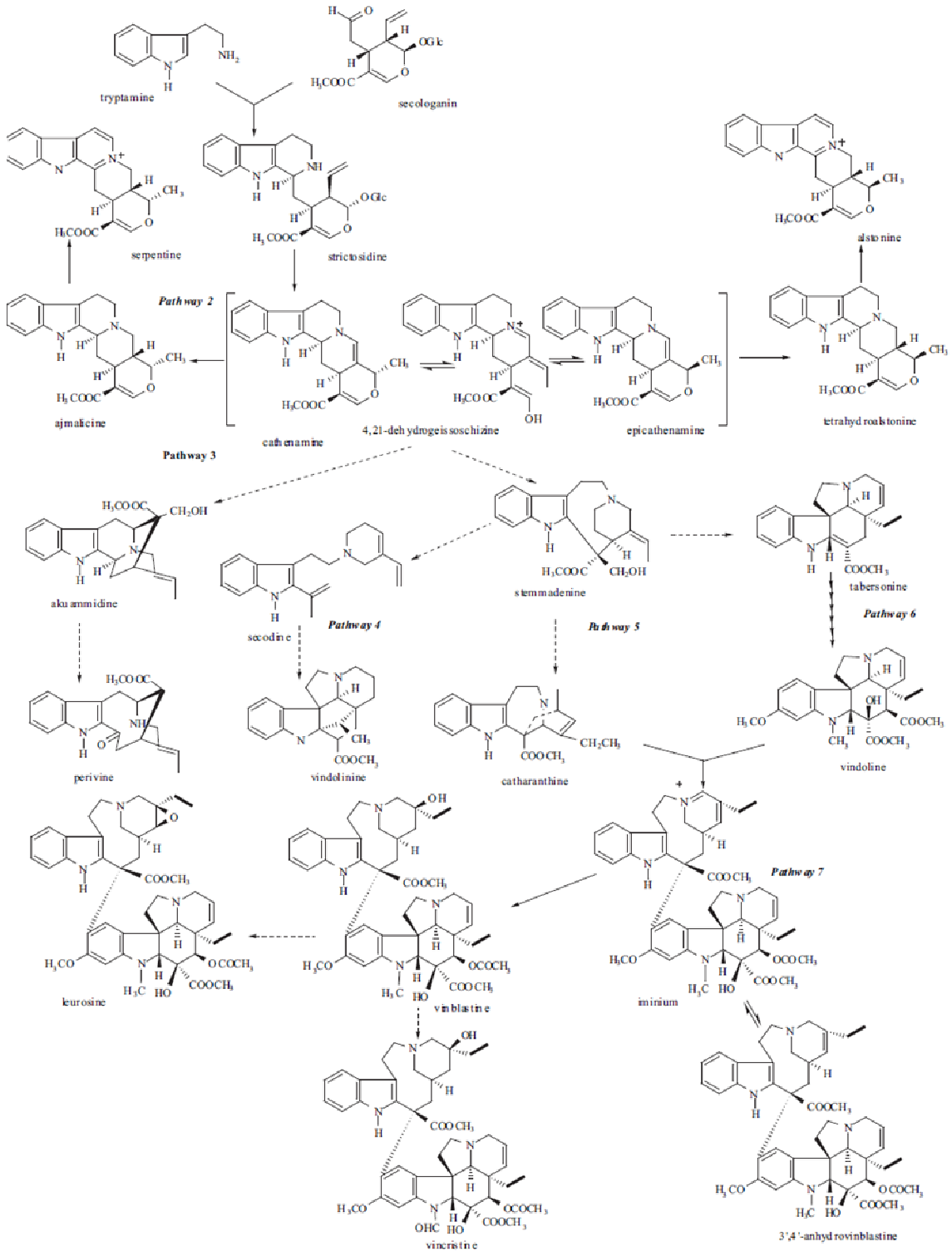


Fig. I.3 – The TIA pathway in *C. roseus*. Solid lines - single enzymatic steps; dashed lines – uncharacterized multiple steps. From van der Heijden, Jacobs et al. (2004).

The TIA pathway is under tight molecular regulation and one of the hallmarks of TIA biosynthesis and accumulation is its general induction by methyl-jasmonate (MeJA). MeJA and jasmonates (JAs) in general, are conserved elicitors of plant secondary metabolism that trigger a change in the transcriptional machinery by activating JA-modulated transcription factors (De Geyter, Gholami et al. 2012). In *C. roseus* particularly, *TDC* and *STR* from the early TIA pathway (see Fig. I.2) were found to be often co-regulated, specifically after auxin starvation, exposure to fungal elicitors and treatment with MeJA, in cell suspension cultures (Goddijn, Dekam et al. 1992, Pasquali, Goddijn et al. 1992, Roewer, Cloutier et al. 1992, Menke, Champion et al. 1999). On the other hand, *D4H* and *DAT* expression was shown to be induced by MeJA in cell cultures, suggesting that this elicitor triggers both the early and the late TIA pathway stages (van der Fits and Memelink 2000). In addition, it has been shown that the enhanced expression of *TDC* and *STR* in response to yeast elicitors depends partially on JAs as secondary signals (Menke, Champion et al. 1999).

The TIA pathway involves a complex subcellular organization, with different parts of the pathway being localized in the plastids, the vacuole, the cytosol and the ER, predicting a number of transport events of the TIA intermediates (Fig. I.4) (Mahroug, Burlat et al. 2007). Being so, some sort of transmembrane transport machinery must exist to facilitate the trafficking of these precursors to the interior and the exterior of the mentioned compartments and organelles. As may be seen in Fig. I.4, the TIA end products are thought to be ultimately accumulated inside the vacuole. In fact, it is now recognized that the central vacuole of plant cells, which occupies 40-90% of the inner cell volume, is critical in the accumulation of secondary metabolites. Vacuoles play at least two positive roles: *i)* the sequestration of biologically active endogenous metabolites inside the cells and *ii)* the protection of such metabolites from catabolism (Yazaki 2005). Moreover, this vacuolar sequestration will also protect cytosolic metabolism from the toxic effect of plant defense compounds such as the alkaloids (Sirikantaramas, Yamazaki et al. 2008). This accumulation in such appropriate compartments should then be regulated in a highly sophisticated manner by specific transporters (Yazaki, Sugiyama et al. 2007).

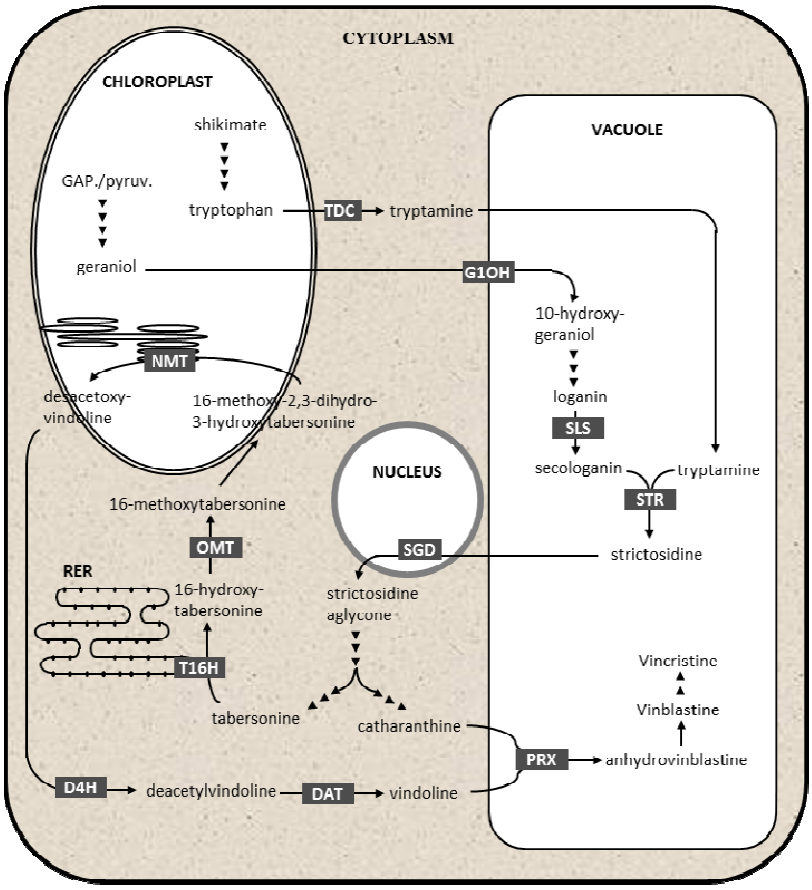


Fig. I.4 – Spatial distribution of the TIA pathway in the *C. roseus* subcellular landscape. G10H: geraniol 10-hydroxylase; SLS: secologanin synthase; TDC: tryptophan decarboxylase; STR: strictosidine synthase; SGD: strictosidine β-D-glucosidase; T16H: tabersonine 16-hydroxylase; OMT: S-adenosyl - L-methionine : 16-hydroxytabersonine - 16-*O*-methyltransferase; NMT: S-adenosyl - L-methionine : 16-methoxy - 2,3-dihydro-3-hydroxytabersonine - N-methyltransferase; D4H: desacetoxy vindoline 4-hydroxylase; DAT: acetylcoenzyme A : 4-*O*-deacetylvindoline 4-*O*-acetyltransferase; PRX: peroxidase. Adapted from Sottomayor and Ros-Barceló (2006).

Adding to the subcellular complexity of the TIA pathway, is has also been shown that the TIA biosynthesis presents multi-cellular compartmentation in *C. roseus* leaves, with the initial biosynthesis of the terpenoid precursor occurring in the internal phloem associated parenchyma, the early and central part of the TIA pathway occurring in epidermal cells, and late steps occurring in laticifer cells and idioblasts, where likely the anticancer TIAs are accumulated (Fig. I.5) (Murata, Roepke et al. 2008, Guirimand, Guihur et al. 2011). This organization predicts inter-cellular translocation of TIA intermediates, adding more transmembrane transport events to the TIA pathway.

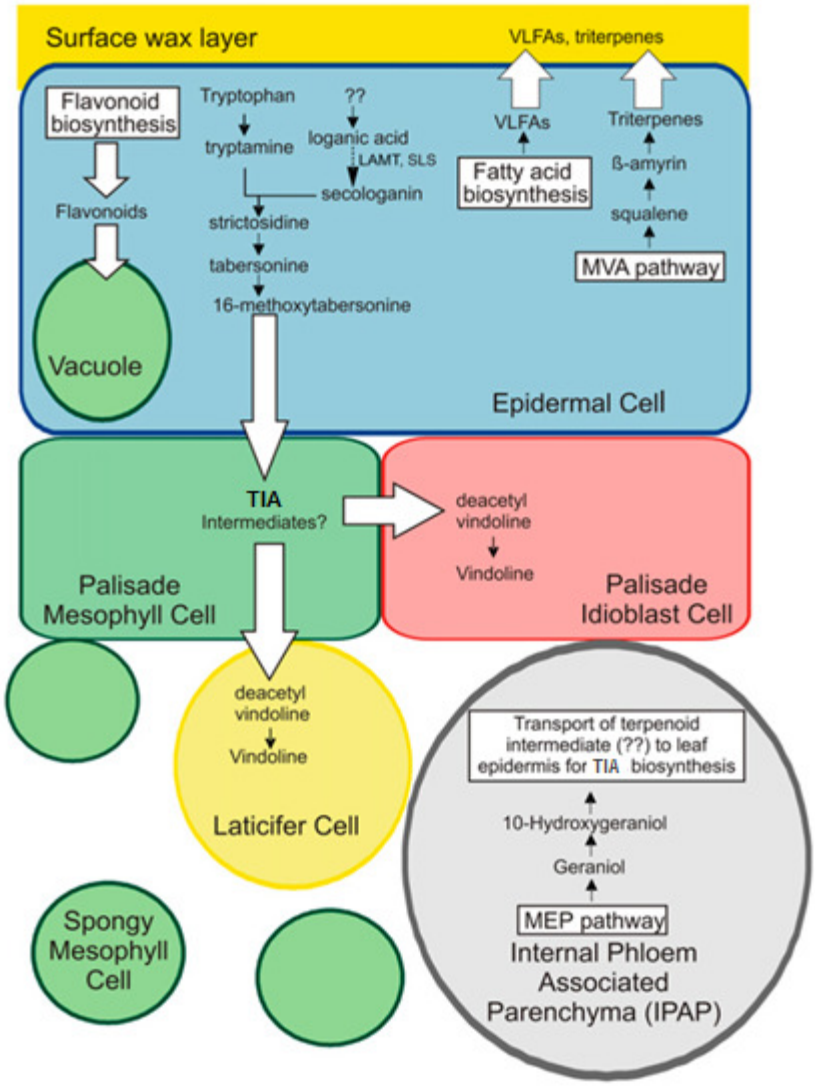


Fig. I.5 – Spatial organization of the TIA pathway in the leaves of *C. roseus*. Adapted from Murata, Roepke et al. (2008).

C. roseus laticifers and idioblasts, specialized parenchyma cells, have been known for long to play a pivotal role in alkaloid accumulation, due to their strong staining by alkaloid indicators (Yoder and Mahlberg 1976). The blue autofluorescence detected in idioblasts and laticifers by UV light incidence is also credited to the vacuolar accumulation of the alkaloid serpentine, further reinforcing this concept (Fig. I.6) (Brown, Renaudin et al. 1984). Furthermore, it was shown that the co-expression of *DAT* and *D4H*, the enzymes that catalyze the two last steps of vindoline biosynthesis, was specifically detected in these cells (St-Pierre, Vazquez-Flota et al. 1999), and several models have now been including idioblasts and lactifers as a central target of the complicated TIA inter-cellular trafficking. Likewise, these cells are most likely the single site in the plant where the anticancer TIAs are synthesized and accumulated.

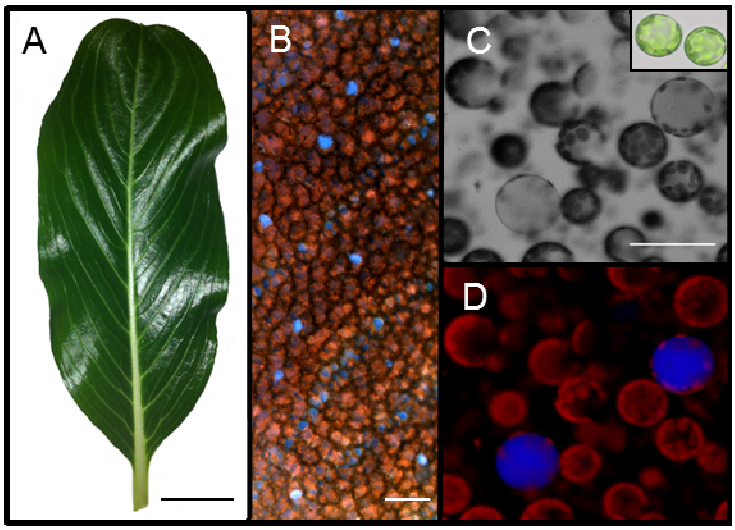


Fig. I.6 – *C.roseus* leaves and their idioblasts. A, Fully developed leaf. B Epifluorescence microscopy image of the adaxial face of a whole mounted leaf. C, Bright field image of mesophyll protoplasts, in set - image in colour. D, Confocal microscopy image of C. The blue fluorescence is due to the alkaloid serpentine and reveals the idioblasts. Red fluorescence is due to chlorophyll and reveals chloroplasts. Bars = 1 cm (A), 50 µm (B) and 30 µm (D). From Carqueijeiro (2013).

This inter-cellular trafficking of the TIAs intermediates reinforces the need for specialized transmembrane transport events mediated by highly regulated, specific transporters. However, little is known about such transporters in plants, particularly in *C. roseus* (Yazaki 2005, Ziegler and Facchini 2008, Verma, Mathur et al. 2012). Recently, several multidrug transporters have been implicated in alkaloid transport in plants, such as the nicotine-transporting MATE-type transporters *NtMATE1*, *NtMATE2* and *Nt-JAT1*, and the berberine-transporting ABC-type transporter *CjMDR1* (Shitan, Bazin et al. 2003, Morita, Shitan et al. 2009, Shoji, Inai et al. 2009). In tobacco, the identified nicotine MATE transporters were shown to mediate vacuolar accumulation of this alkaloid through a proton antiport system (Morita, Shitan et al. 2009, Shoji, Inai et al. 2009). Meaningfully, the BioNatPro lab at IBMC has recently shown that vacuolar accumulation of TIAs in *C. roseus* leaves is also mediated by a proton antiport, strongly suggesting the involvement of a MATE transporter (Carqueijeiro, Noronha et al. 2013).

I.3. MATE-type transporters

Multidrug transporters, in a classical and universal definition, confer multidrug resistance in both prokaryotes and eukaryotes by extruding xenobiotics and toxic metabolites from cells. The early identification of several representative multidrug transporters with different structures and action mechanisms led to the classification of these into five main superfamilies: the major facilitator superfamily (MFS), the small multidrug resistance (SMR) family, the resistance nodulation cell division (RND) family, the ATP-binding cassette (ABC) family and the multidrug and toxic compound extrusion (MATE) family (Fig. I.7). MATE, MFS, SMR and RND members are secondary transporters energetically coupled with an electrochemical gradient of H^+ or Na^+ across membranes, whereas the ABC members are membranous ATPases directly coupled with drug transport and ATP hydrolysis (Moriyama, Hiasa et al. 2008).

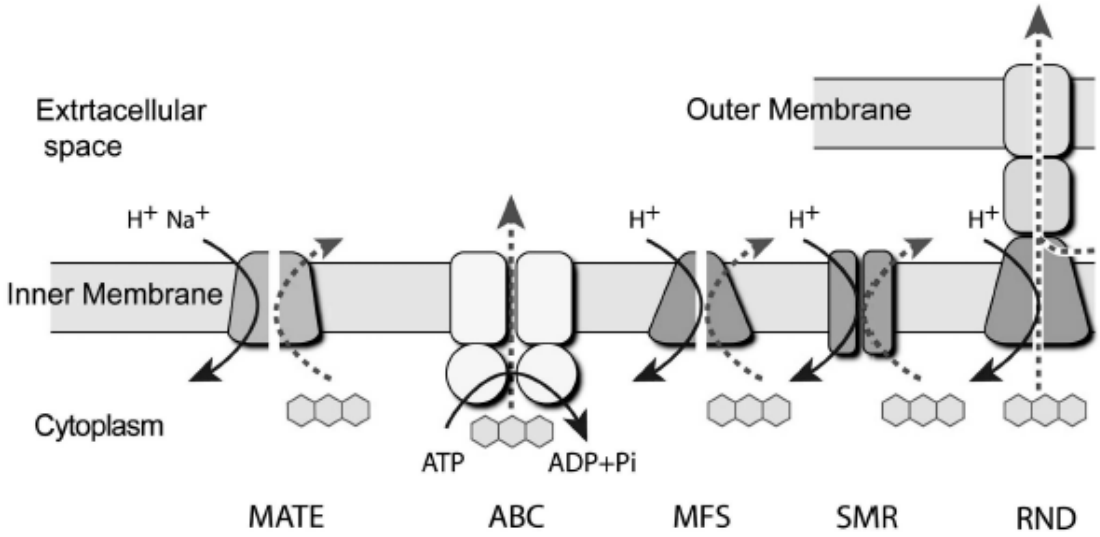


Fig. I.7 - Schematic representation of the five main multidrug transporter superfamilies. From Moriyama, Hiasa et al. (2008).

The first MATE-type transporters were only discovered in 1998, when Morita and co-workers discovered a novel multidrug transporter in *Vibrio parahaemolyticus* (*NorM*) and the respective homolog in *Escherichia coli* (*YdhE*). These transporters were found to possess 12 transmembrane domains (TMDs) and were initially assigned to the MFS family since, at that time, such structure was a unique feature of the MFS proteins (Morita, Kodama et al. 1998). However, shortly after, some authors argued that none of these two proteins showed sequence identity with any of the multidrug transporters known at that time, and should therefore be assigned to a new multidrug superfamily, termed the multidrug and toxic compound extrusion (MATE) family (Brown, Paulsen et al. 1999).

Over time, MATE-type transporters revealed to be basic constituents of all kingdoms of living organisms - several different MATE-type transporters were identified and soon it was observed that these compose one of the mostly conserved transporter families in nature and widespread in the three main kingdoms, *i.e.* Eukarya, Archaea and Eubacteria (Fig. I.8) (Moriyama, Hiasa et al. 2008); so conserved that, in fact, most MATE family members consist of 400–550 residues with 12 transmembrane helices and share about 40% sequence similarity (Omote, Hiasa et al. 2006).

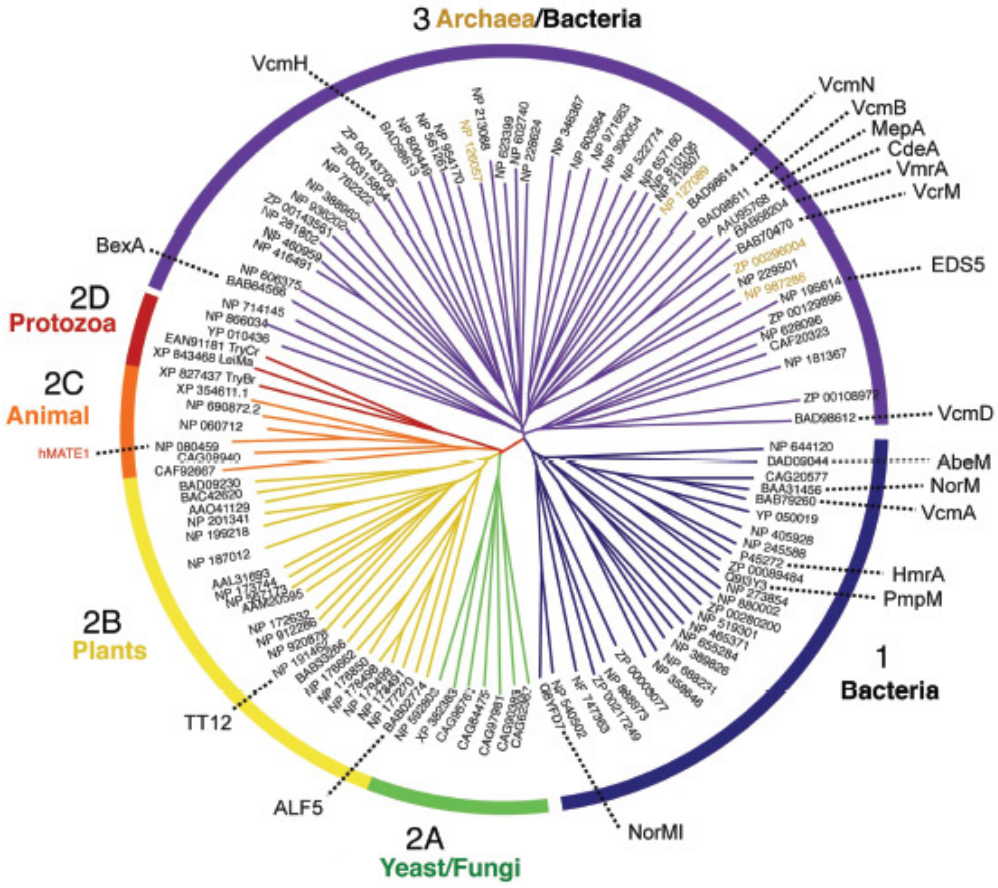


Fig. I.8 - A phylogenetic tree of selective MATE-type transporters from all kingdoms. From Moriyama, Hiasa et al. (2008).

In plants, we have been slowly acknowledging the multifunctional essence of MATE-type transporters, which seem to play several different physiological roles, in contrast with other groups of organisms, and despite maintaining the typical MATE protein architecture (Fig. I.9). Remarkably, *Arabidopsis thaliana* possesses 58 different MATE transporter orthologs, whereas the human genome contains only 2 MATE transporters (Omote, Hiasa et al. 2006, Moriyama, Hiasa et al. 2008). This so-called multifunctional essence of plant MATE-type transporters is credited not only to basic

developmental processes but also to the translocation and compartmentation of secondary metabolites (Fig. I.10).

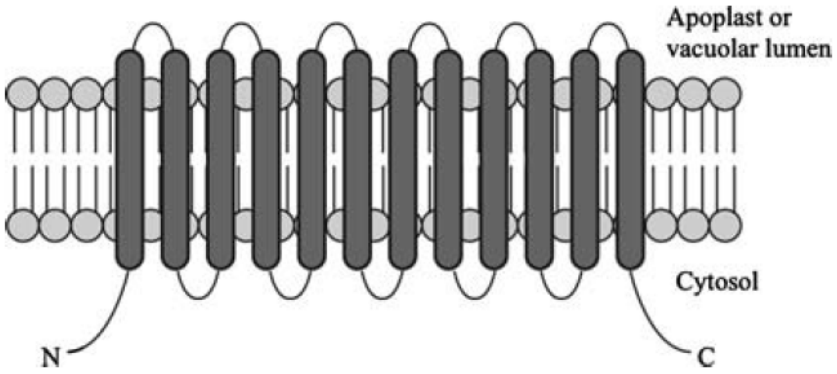


Fig. I.9 - Putative secondary structure of plant MATE-type transporters, exhibiting the 12 transmembrane helices/domains and the loops that allow these domains to bend over and immerge in the lipid bilayer. N - N-terminal or amino-terminal region; C - C-terminal or carboxyl-terminal region. From Yazaki, Sugiyama et al. (2007).

Various reports have been presenting the multitude of MATE-type transporters functions, some identifying the respective substrates and the subcellular localization, and a list of those results is represented in Table I.1. *Arabidopsis AtALF5* was identified from a mutant defective in lateral root formation, suggesting this MATE-type transporter functions as a distinct requirement for plant development, but this transporter was also shown to act as an efflux transporter of xenobiotics (Diener, Gaxiola et al. 2001). Similarly, *Arabidopsis AtDTX1* was shown to be an H^+ antiport efflux carrier for multidrug resistance, to localize at the plasma membrane (PM) and to extrude berberine (alkaloid), cadmium (Cd^{2+}, heavy metal), and norfloxacin (antibiotic) (Li, He et al. 2002). *Arabidopsis AtFRD3* was shown to extrude citrate into the root vasculature, a function that is necessary for iron distribution throughout the plant (Green and Rogers 2004). Some MATE-type transporters were found to confer aluminum tolerance by mediating the efflux of citrate into the rhizosphere, including the barley *HvMATE*, the *AtFRD3* homolog lupin *LaMATE* also upregulated in phosphorous deficiency conditions and, very recently, the *Eucalyptus camaldulensis EcMATE1-4* (Uhde-Stone, Liu et al. 2005, Wang, Raman et al. 2007, Sawaki, Kihara-Doi et al. 2013). So far, whenever the subcellular fate of the MATE-type transporters with the above referred functions was determined, it indicated a plasma membrane localization, as expected.

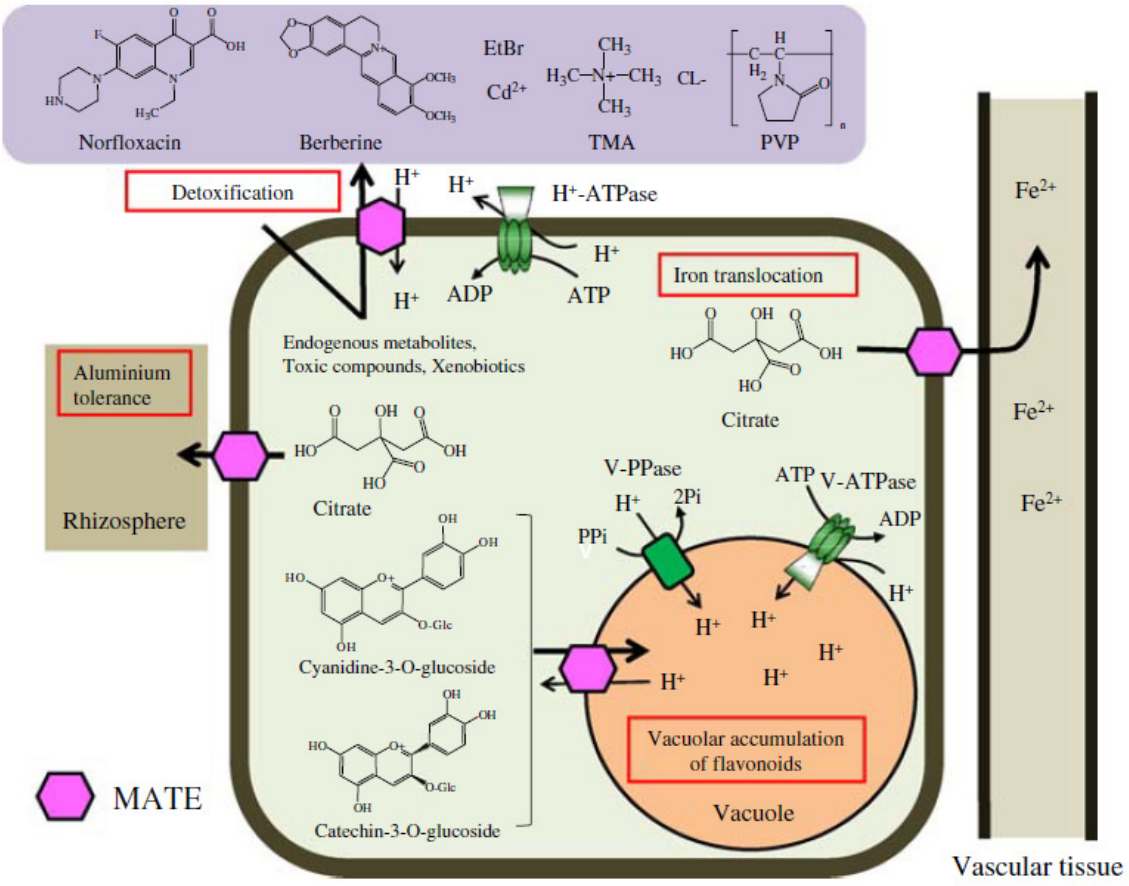


Fig. I.10 - A model of the physiological functions and putative substrates of MATE-type transporters in plants. From Yazaki, Sugiyama et al. (2007).

In regard to vacuolar MATE-type transporters, all members identified so far are responsible for sequestering secondary metabolites inside the vacuole, at the expense of the electrochemical H^+ proton gradient across the tonoplast. *AtTT12* (TRANSPARENT TESTA12) was identified from a T-DNA tag *Arabidopsis* mutant screen, displaying pale brown seeds and much was speculated regarding its function, presumed to be a vacuolar transporter for flavonoids in the seed coat (Debeaujon, Peeters et al. 2001). Only in 2007 it turned out that this gene encodes a vacuolar MATE-type tonoplast H^+-antiporter expressed in the proanthocyanidins-synthesizing cells of the seed coat, that mediates the transport of anthocyanins (Marinova, Pourcel et al. 2007). This gene was considered a reference among other TRANSPARENT TESTA-like transporters that afterwards have progressively become characterized, not only in *Arabidopsis* but in other species as well (Zhao and Dixon 2009, Frank, Keck et al. 2011, Pang, Cheng et al. 2013, Yu 2013).

Table I.1 - Properties of plant MATE-type transporters. From Yazaki, Sugiyama et al. (2007).

Plant	Name	Driving force	Subcellular localization	Tissue	Substrate	Proposed physiological function
Arabidopsis	*ALF5*	N.D.	N.D.	Root (epidermis, cortex)	(TMA, PVP)	Protection of the root from toxic compounds
	DTX1	H^+	PM	Flower, leaf, stem, root	(Norfloxacin, EtBr, berberine, heavy metals)	Efflux of endogenous metabolites and xenobiotics
	EDS5	N.D.	N.D.	N.D.	N.D.	Salicylic acid-dependent signaling cascade for disease resistance
	FRD3	N.D.	N.D.	Root (pericycle, vascular cylinder)	Citrate	Efflux of citrate into the root vasculature for iron translocation
	TT12	H^+	Vac	Seed coat endothelium	Flavonoids	Vacuolar flavonoid/H^+ antiporter in the seed coat
Tomato	*MTP77*	N.D.	N.D.	N.D.	N.D.	Vacuolar transporter of anthocyanins (similar to TT12)
Lupin	*LaMATE*	N.D.	PM	Root	N.D.	N.D.
Barley	*HvMATE*	N.D.	N.D.	N.D.	(Citrate)	Aluminum tolerance
	HvAACT1	N.D.	PM	Root, shoot	Citrate	Al-activated efflux carrier of citrate, aluminum tolerance
Sorghum	*SbMATE*	N.D.	PM	Root	(Citrate)	Al-activated efflux carrier of citrate, aluminum tolerance

N.D.—Not determined; PM—Plasma membrane; Vac—Vacuole; TMA—Tetramethylammonium; PVP—Polyvinylpyrrolidone; EtBr—Ethidium bromide

Only very recently, the first vacuolar transporter for an alkaloid was recently characterized for nicotine in tobacco as being a MATE-type transporter functioning as a H^+-antiporter, *Nt-JAT1*. This transporter presented enhanced expression upon the application of methyl jasmonate (MeJA) and was suggested to be involved in the vacuolar compartmentation of nicotine in the aerial parts of *Nicotiana* species (Morita, Shitan et al. 2009). Similarly, two more MATE transporters were implicated in the vacuolar sequestration of alkaloids in tobacco roots, *NtMATE1* and *NtMATE2*, which were also induced by MeJA (Shoji, Inai et al. 2009). The combination of these tobacco MATE-type transporters, *NtJAT-1*, *NtMATE1* and *NtMATE2* is consistent with the suggested spatial organization of the entire nicotine biosynthesis pathway (Fig. I.11). Nicotine is produced in tobacco roots, after which it can be retained *in situ*, in the root-cells vacuoles (a potential role for *NtMATE1/2*) or be translocated to the sink organs, tobacco leaves, by deposition into the leaf-cells vacuoles (a potential role for *Nt-JAT1*) (Hashimoto and Yamada 2003). Accordingly, expression of *NtJAT-1*, *NtMATE1* and *NtMATE2* in yeast cells, where the transporters localized at the plasma membrane, compromised the accumulation of exogenously supplied nicotine into the cells, indicating that they were all capable of extruding the intracellular nicotine at high rates. Nevertheless, other transporters may be involved in nicotine partioning and in fact it

was suggested that other tobacco multidrug transporters might further contribute to the vacuolar accumulation of nicotine in the root cells (Shoji, Inai et al. 2009).

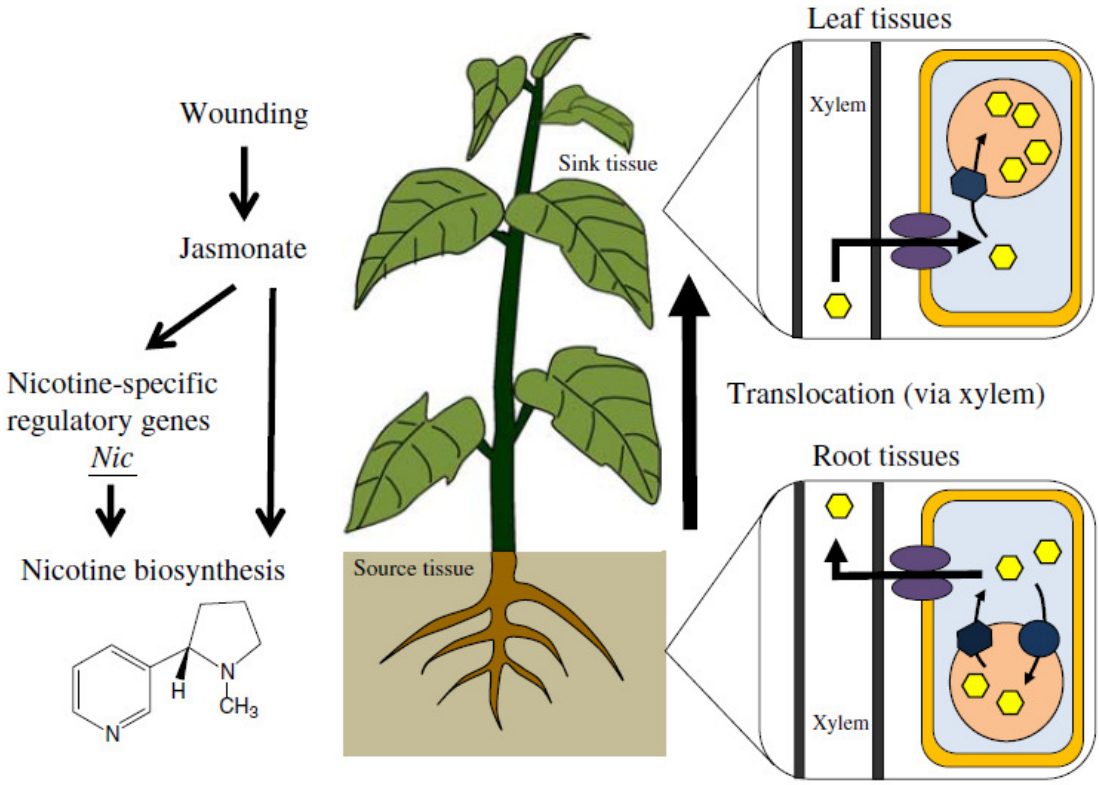


Fig. I.11 - Model of nicotine translocation and cellular accumulation in *Nicotiana* species. Nicotine biosynthesis is induced in root cells, via insect attack, wounding, and jasmonate treatment to the leaves. After biosynthesis, nicotine is translocated to the leaves via the xylem, and is accumulated in the vacuoles of leaves to function as a defensive toxin against insects and herbivores. From Yazaki, Sugiyama et al. (2007).

The case of nicotine in tobacco clearly demonstrates the potential of multidrug transporters studies in medicinal plants. On the one hand, knowledge on the transporters further clarifies the regulation mechanisms and spatial organization of the alkaloid biosynthetic pathways *in planta*, possibly providing new insights and paving the road to innovative approaches for efficient metabolic engineering strategies. On the other hand, the ectopic expression of these transporters *per se* in yeast cells will clearly contribute and facilitate the production of large scale yeast bioreactors for nicotine synthesis or derivatives. The findings in the multidrug transporters field will substantially improve the versatility of the available transport modules that may be combined with biosynthetic enzyme modules to establish effective metabolic engineering strategies to provide higher levels or novel plant products (Nour-Eldin and Halkier 2013).

I.4. Objectives

Previous work with *C. roseus* has shown that the major TIAs present in the leaves of *C. roseus* are accumulated in mesophyll vacuoles by a proton-driven antiport (Carqueijeiro, Noronha et al. 2013), strongly suggesting that vacuolar accumulation of *C. roseus* TIAs may be mediated by a MATE-type transporter. Significantly, when a nanoLC-MS/MS proteomic analysis was recently performed on *C. roseus* leaf-cells vacuolar and tonoplast fractions, a MATE-type transporter protein was abundantly detected and rapidly identified, CroMATE1. Thus, the goal of this work was to isolate, clone and perform the molecular characterization of this candidate MATE-type transporter, potentially involved in *C. roseus* TIA transmembrane transport. For this purpose, it was established and organized a step-by-step framework including *i)* a deep *in silico* characterization and prediction in regard to *CroMATE1* gene and protein topology, specific domains and function, *ii)* a hierarchical co-expression clustering (HCL) analysis using up-to-date *C. roseus* transcriptomic data and pinpointing *CroMATE1* expression profile among characterized TIA biosynthetic genes, *iii)* the isolation of *CroMATE1* full-length CDS by PCR followed by cloning in an entry vector, *iv)* the production of GFP-fusions for investigation of CroMATE1 subcellular localization, and *v)* a phylogenetic reconstruction of CroMATE1 protein similarities with already characterized plant MATE-type transporter proteins.

II. Materials and Methods

II.1. Biological material

II.1.1. Plant material

Catharanthus roseus (L.) G. Don cv. Little Bright Eyes plants were grown at 25 ºC in a growth chamber, under a 16h photoperiod with a light intensity of 70 µmol m^{-2} s^{-1}. Seeds were acquired from B & T World Seeds (Aigues-Vives, France). For the isolation of protoplasts, the 2nd and 3rd leaf pairs of adult *C. roseus* plants were used.

II.1.2. Bacterial strains and cloning vectors

Chemically competent *Escherichia coli* str. TOP10 cells were used for the bacterial cloning procedures. In this work, both pTH-2 (Fig. II.1, A) (Niwa, Hirano et al. 1999) and pTH-2BN (Fig. II.1, B) (Kuijt, Lamers et al. 2004) plasmids were used as expression vectors. Both plasmids possess about 4,100bp, and share several common features, such as: *i)* the presence of a strong, constitutive promoter CaMV 35S (~400 bp), *ii)* a MCS that allows insertion of sequences of interest upstream of GFP, *iii)* a GFP reporter gene with a S65T mutation, resulting in an increased chromophore stability (~750 bp); *iv)* a nopaline synthase terminator from *Agrobacterium tumefaciens* (*nos*T; ~250 bp), *v)* an ampicillin/carbenicillin resistance marker, Amp^r; *vi)* M13 FWD and M13 REV primer-specific sequences that facilitate sequencing analyses. pTH-2BN is a derivative from pTH-2 that lacks the GFP stop-codon. In pTH-2BN, a BN linker was introduced instead (~40bp), allowing cloning of sequences of interest downstream the GFP coding sequence. Furthermore, these expression vectors enable cloning and expression in bacterial and plant systems, respectively.

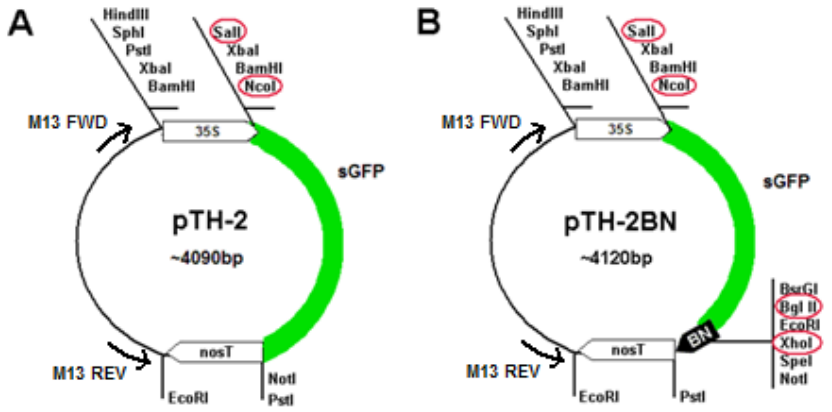


Fig. II.1 - Schematic representation of A) pTH-2 and B) pTH-2BN cloning vectors and related features. Red circles indicate the restriction sites used for the directional cloning of *CroMATE1*.

II.2. Identification and *in silico* characterization of *CroMATE1*

The *CroMATE1* fullest cDNA sequence was obtained from Medicinal Plant Genomics Resource database (MPGR, http://medicinalplantgenomics.msu.edu/). (Gongora-Castillo, Childs et al. 2012). *CroMATE1* CDS open reading frame (ORF) was identified and translated by using Molecular Toolkit translation tool (Bowen 1998) (http://www.vivo.colostate.edu/molkit/translate/index.html). The resulting predicted protein sequence, CroMATE1, was submitted to NCBI-pBLAST® analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi) in order to detect and compare homologous transporters in other species. CroMATE1 theoretical pI and molecular weight were retrieved from the ExPASy Compute pI/Mw tool (http://web.expasy.org/cgi-bin/compute_pi/pi_tool), its putative transmembrane domains (TMDs) were identified using HMMTOP (http://www.enzim.hu/hmmtop/) and PredictProtein (https://www.predictprotein.org/), and its predicted subcellular localization was assessed by using WoLF PSORT (http://wolfpsort.org/) (Horton, Park et al. 2007).

II.3. Hierarchical co-expression clustering (HCL) of *CroMATE1* with TIA biosynthesis genes

The whole *C. roseus* transcriptomic data available was downloaded from the MPGR database (ftp://ftp.plantbiology.msu.edu/pub/data/MPGR/Catharanthus_roseus/, file contained in *cra.matrix.FPKM.vf.20121704.xlsx.zip*), and the gene expression profiles of interest were filtered. These were submitted to Multi ExperimentViewer 4.9.0 (MeV -

TM_4) (Saeed, Sharov et al. 2003). Gene normalization was performed prior to subsequent clustering computation. MeV-HCL was run by defining Pearson Correlation metrics and an Average Linkage clustering method. Despite not shown here, other combinations of different metrics (*e.g.* Euclidean) and clustering methods (*e.g.* Complete linkage; Single linkage) were computed in several simulations with the same data set. The centroid distribution of gene expression profiles found in the resulting cluster was presented using Self-Organizing Tree Algorithm (SOTA) with the same data set, under the same settings.

II.4. Isolation of the *CroMATE1* full-length CDS

II.4.1. Purification of RNA from *C. roseus* leaf mesophyll protoplasts and cDNA synthesis

The 2nd and 3rd leaf pairs of 6 month-old *C. roseus* plants were used for protoplast isolation performed as previously described (Duarte, Ribeiro et al. 2011). The obtained highly pure solution of viable protoplasts was used for total RNA purification with RNeasy® Plant Mini Kit (QIAGEN), with minor adaptations to the included protocol. Before this procedure, the solution was vortexed after mixing a 3:1 volume of RLT buffer with 10 µl β-ME/mL and immediately frozen in liquid nitrogen. Immediately after thawing the material, the kit protocol was carried on at step no. 4. The resulting RNA concentration was measured with a NanoDrop ND-1000 (Thermo Scientific) device and a final total RNA concentration of 50 ng μL^{-1} was used for the RT reactions with iScript™ Select cDNA Synthesis Kit (Bio-Rad), following the kit recommendations with a customized reaction setup (Table II.1). This reaction took place in a sequential order:

Table II.1 – Reverse transcription reaction setup, reagents and respective volumes.

RNA (µL)	Nuclease-free H_2O (µL)	DNaseI (µL)	EDTA (µL)	Oligo-dT (µL)	RT Mix (µL)	RT enzyme (µL)	Total Vol. (µL)
1	10	1	1	2	4	1	20

after mixing RNA, H_2O and DNaseI (Thermo Scientific), the tubes were incubated at 37 ºC for 30 min; after addition of EDTA, were incubated at 65 ºC for 10 min; after addition of Oligo$(dT)_{20}$, were incubated at 65 ºC for 5 min; finally, after addition of RT Mix and

the RT enzyme, the tubes were incubated at 42 ºC for 90 min. In the end of the reaction, the samples were incubated at 85 ºC for 5 min in order to denature the RT enzyme, thus preventing unspecific activity or uncontrolled events. For further use, the cDNA aliquots were stored at -20 ºC.

II.4.2. Primer design

CroMATE1 CDS was submitted to WebCutter 2.0 (http://rna.lundberg.gu.se/cutter2/) to account for restriction enzymes that would not cut its sequence while cutting the multiple coning sites of the vectors. Primers used for the amplification and cloning of *CroMATE1* CDS into the cloning vectors pTH-2 and pTH-2BN were manually designed and were extended to include endonuclease restriction sites to allow directional cloning (Table II.2). The respective biochemical features were evaluated by using OligoAnalyzer 3.1 (http://eu.idtdna.com/analyzer/applications/oligoanalyzer/, IDT®) (Table II.3). Primers were ordered to STABVida.

Table II.2 - Primers used to amplify CroMATE1 for cloning in pTH2 and pTH2BN. Gray – part of *CroMATE1* CDS. Yellow - linker that encodes a SGSGS amino acid motif. Red - extensions with restriction sites, accordingly to the respective cloning purpose: G'TCGAC - *Sal*I restriction site; A'CATGT - *Pci*I restriction site; A'GATCT - *Bgl*II restriction site; C'TCGAG - *Xho*I restriction site.

ID	Sequence (5'→3')	Size (bp)
CroMATE1 FWD	ATGGGTTCCAAACAAAACTATGAAATAAACC	31
CroMATE1 REV	TTATTCATTGGACAAAGATTTTGGCTTG	28
N-terminal FWD	AGCCGTCGACATGGGTTCCAAACAAAACTATGAAATAAACC	41
N-terminal REV	GATGACATGTATCCTGAACCAGATTCATTGGACAAAGATTTTGGCTTG	48
C-terminal FWD	CAAGATCTCTATGGGTTCCAAACAAAACTATGAAATAAACC	41
C-terminal REV	CGCCTCGAGTTATTCATTGGACAAAGATTTTGGCTTGTCC	40

Table II.3 - Biochemical features assigned for each primer, by using OligoAnalyzer 3.1.

ID	Bp	T_m	GC%	Hairpin	Homodim.	Heterodim.
CroMATE1 FWD	31	56.6	32.3	-0.55	-6.36	-8.91
CroMATE1 REV	28	55.0	32.1	-1.63	-5.84	
N-terminal FWD	41	56.6	32.3	-0.55	-6.36	-8.91
N-terminal REV	48	54.8	36.0	-1.63	-8.07	
C-terminal FWD	41	56.6	32.3	-1.63	-7.82	-8.91
C-terminal REV	40	58.2	35.5	-3.5	-9.96	

II.4.3. PCR amplification

PCR amplification settings were individually optimized for *CroMATE1* CDS isolation (Fig. II.2), *CroMATE1* CDS isolation for cloning into pTH-2 (Fig. II.3) and *CroMATE1* CDS isolation for cloning into pTH-2BN (Fig. II.4). The PCRs hereby described, made use of the cDNA template produced in section *4.1.*, a *DreamTaq* DNA polymerase and respective reaction buffer (Thermo Scientific), 10mM dNTP mixture (Thermo Scientific), 10µM forward and reverse primer stocks and bi-distillated H_2O. The reactions were performed in a T100™ thermal cycler (Bio-Rad).

Reagents	Volume (µL)
cDNA (template)	2
10x *DreamTaq* Buffer	2.5
10mM dNTPs	0.5
10µM CroMATE1 FWD	1
10µM CroMATE1 VER	1
DreamTaq	0.2
ddH_2O	17.8
Total	25

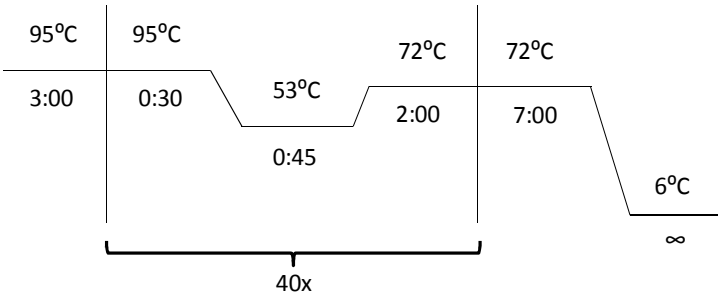


Fig. II.2 - PCR optimized setup for amplification of the *CroMATE1* CDS (left) and the respective PCR settings defined in T100™ thermal cycler (right).

Reagents	Volume (µL)
cDNA (template)	2
10x *DreamTaq* Buffer	2.5
10mM dNTPs	0.5
10µM N-terminal FWD	1
10µM N-terminal REV	1
DreamTaq	0.2
ddH_2O	17.8
Total	25

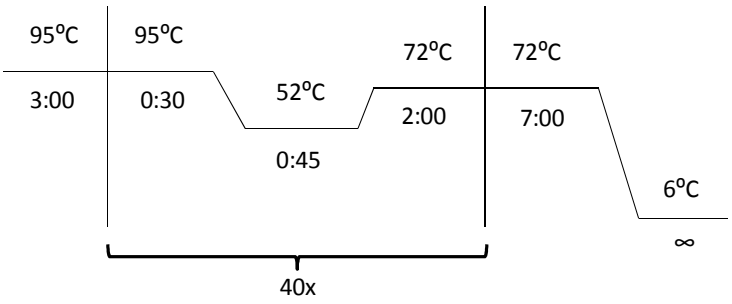


Fig. II.3 - PCR optimized setup for amplification of the *CroMATE1* CDS to be used for cloning into pTH-2 (left) and the respective PCR settings defined in T100™ thermal cycler (right).

Reagents	Volume (µL)
cDNA (template)	2
10x *DreamTaq* Buffer	2.5
10mM dNTPs	0.5
10µM C-terminal FWD	1
10µM C-terminal REV	1
DreamTaq	0.2
ddH_2O	17.8
Total	25

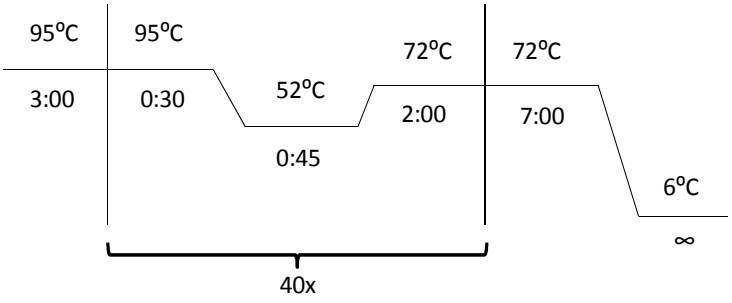


Fig. II.4 - PCR optimized setup for amplification of the *CroMATE1* CDS to be used for cloning into pTH-2BN (left) and the respective PCR settings defined in T100™ thermal cycler (right).

For the cloning-specific products, 5 to 6 replicates were prepared and submitted to PCR amplification. Once finished, these replicates were pooled together in a single PCR tube, mixed with loading dye, and loaded into 1%(w/v) agarose gel prepared with 1x TAE buffer (40 mM Tris base, 10% (v/v) acetic acid and 10 mM EDTA), supplemented with 0.5 µg mL^{-1} $EtBr_2$ (Bio-Rad), and ran at 80 V for 25 min using PowerPac Basic Electrophoresis Kit (Bio-Rad). GeneRuler™ DNA Ladder Mix (Thermo Scientific) was used as a molecular ruler in every electrophoretic analysis. The DNA of interest was purified from the agarose gel using the GeneJET™ Gel extraction kit (Thermo Scientific), according to the manufacturer's instructions.

II.5. Cloning of CroMATE1

II.5.1. Digestion of PCR products and cloning vectors

The DoubleDigest web tool was consulted to help defining the digestion reactions (http://www.thermoscientificbio.com/webtools/doubledigest/). The purified *CroMATE1* CDS products were subject to digestion with the appropriate enzymes to trim the primer extensions incorporated in the products. In parallel, the corresponding cloning vectors, pTH-2 and pTH-2BN, already available in the laboratory in the form of purified plasmid minipreps, were also digested. For cloning of the *CroMATE*1 CDS into pTH-2, the double digestion of *Sal*I and *Pci*I was performed by firstly using 1x Tango™ buffer with *Pci*I, incubate at 37 ºC for 1 h 30 min, and when the first digestion was complete, adding 10x concentrated Tango™ buffer to a final 2x concentration and 2-fold excess of *Sal*I, again incubating at 37 ºC for 1 h 30 min. For pTH-2, the double digestion was performed using *Nco*I, a slightly greater amount of *Sal*I and 2x Tango™ buffer, following an incubation at 37 ºC for 2h (Table II.4). Then, 1µL CiAP (Thermo Scientific) was added to the mixture and it was further incubated at 37 ºC for 1h. Finally, 3.2µL of 500mM EDTA pH 8.0 were added to inhibit further unspecific activity. At the end of the digestions, both samples were incubated at 65 ºC for 20min, to denature the restriction enzymes and prevent interference with the subsequent processes.

Table II.4 - Double digestion reaction setups of pTH-2 cloning vector (left panel) and *CroMATE1* CDS for cloning into pTH-2 (right panel). Asterisks - Volumes added after a first incubation period was complete.

pTH-2 digestion		**N-terminal CroMATE1 digestion**	
Reagents	**Volume (µL)**	**Reagents**	**Volume (µL)**
ddH_2O	46	ddH_2O	45
10x Tango™ Buffer	16	10x Tango™ Buffer	7.5 + 10*
pDNA	10	PCR product	20
*Sal*I	4.5	*Sal*I	5*
*Nco*I	3.5	*Pci*I	2.5
Total	80	**Total**	90

CroMATE1 CDS for cloning into pTH-2BN was digested using *Bgl*II and *Xho*I in equal amounts and 2x Tango™ Buffer, following incubation at 37 ºC for 3h. pTH-2BN cloning vector was digested using equal amounts of the same enzymes, *Bgl*II and *Xho*I and 2x Tango™ Buffer, then incubating at 37 ºC for 2 h (Table II.5). Again, 1 µL CiAP (Thermo Scientific) was added to the mixture and it was further incubated at 37 ºC for 1h. Finally, 3.2 µL of 500mM EDTA pH 8.0 were added to inhibit further unspecific

activity. At the end the digestions, both samples were incubated at 65 ºC for 20 min, to denature the restriction enzymes and prevent interference with the subsequent processes.

Table II.5 - Double digestion reaction setups of pTH-2BN cloning vector (left panel) and *CroMATE1* CDS for cloning into pTH-2BN (right panel).

pTH-2BN digestion		**C-terminal CroMATE1 digestion**	
Reagents	**Volume (µL)**	**Reagents**	**Volume (µL)**
ddH_2O	47	ddH_2O	35
10x Tango™ Buffer	16	10x Tango™ Buffer	15
pDNA	10	PCR product	20
*Bgl*II	3.5	*Bgl*II	2.5
*Xho*I	3.5	*Xho*I	2.5
Total	80	**Total**	75

All samples were individually mixed with loading dye, and loaded into 1%(w/v) agarose gel prepared with 1x TAE buffer containing 40 mM Tris base, 10 % (v/v) acetic acid and 10 mM EDTA, supplemented with 0.5 µg mL^{-1} $EtBr_2$ (Bio-Rad), and ran at 80 V for 25 min using PowerPac Basic Electrophoresis Kit (Bio-Rad). The digested products were purified from the agarose gels using the GeneJET™ Gel Extraction Kit (Thermo Scientific), according to the manufacturer's instructions.

II.5.2. Ligation of PCR products and cloning vectors

All four samples previously collected were submitted to a NanoDrop ND-1000 (Thermo Scientific) device. Making use of the knowledge and experience of previous molecular cloning approaches in the laboratory, the molar ratio of 4:1 (insert:vector) was selected as a starting point. The insert:molar ratio was calculated using a formula described in a Promega molar ratio calculator (Fig. II.5).

$$\frac{length\ of\ insert\ (kb)}{length\ of\ vector\ (kb)} \times [vector] \times \frac{4}{1} = [insert]$$

Fig. II.5 - Equation used to determine vector and insert amounts for a 4:1 molar ratio ligation preparation. [vector] - either pTH-2 or pTH-2BN concentration, in ng μL^{-1}; [insert] - either N- or C-terminal CroMATE1 products concentration, in ng μL^{-1}. From Promega (http://www.promega.com/techserv/tools/biomath/calc06.htm).

The ligation reaction setups were defined and prepared by using T4 DNA ligase (Thermo Scientific) along with the respective reaction buffer for pTH-2 and *CroMATE*1 CDS for cloning into pTH-2 (Table II.6) to yield a *CroMATE1*-sGFP construct; pTH-2BN and *CroMATE1* CDS for cloning into pTH-2BN (Table II.7) to yield a sGFP-*CroMATE1* construct. Once prepared, the ligation reactions were kept incubating at 6 ºC overnight (ON).

Table II.6 - Ligation reaction setup for pTH-2 and *CroMATE*1 CDS for cloning into pTH-2 digested products, using a 4:1 molar ratio.

Reagents	Volume (µL)
10x T4 DNA ligase Buffer	2
N-terminal CroMATE1	9
T4 DNA ligase	2
ddH_2O	6
pTH-2	1
Total	20

Table II.7 - Ligation reaction setup for pTH-2BN and *CroMATE1* CDS for cloning into pTH-2BN digested products, using a 4:1 molar ratio.

Reagents	Volume (µL)
10x T4 DNA ligase Buffer	2
C-terminal CroMATE1	6.14
T4 DNA ligase	2
ddH_2O	8.86
pTH-2BN	1
Total	20

II.5.3. Heat-shock mediated transformation of *E. coli* str. TOP10

Chemically competent *Escherichia coli* str. TOP10 bacteria were subjected to transformation using the heat shock method. For each of the two constructs, 50-100 µL of TOP10 chemically competent cells were mixed with 5 µL of the ligation reaction. After a 30 min incubation on ice, the mixture of competent bacteria and pDNA was heat-shocked at 42 ºC for 2 min, and placed back on ice for 2 min. 1 mL of Luria-

Bertrani (LB) medium was added to the mixture and the cells were left to recover for 1 h at 37 ºC. LB medium (Difco) is composed of tryptone (10 g L^{-1}), yeast extract (5 g L^{-1}), and NaCl (10 g L^{-1}). The cell suspension was then centrifuged (Centrifuge 5415 R, Eppendorf) at 4000 rpm for 4 min at RT, and 900 µL of supernatant were removed. The cells were resuspended in the remainder volume, were plated onto LB-agar (LB with 1.5 % agar; Liofilchem) supplemented with 100 µg mL^{-1} of ampicillin (Sigma) to select the transformants, and the plates were incubated overnight (ON) at 37 ºC.

Nine colonies were randomly selected throughout the 4 replicate plates prepared for each of both constructs, *CroMATE1*-sGFP and sGFP-*CroMATE1*. The colonies were pinched with a white micropipette tip which was ejected into individual sterile 50 mL Falcon tubes with 5 mL of LB medium supplemented with 150 µg mL^{-1} of ampicillin (Sigma). The inoculated tubes were then incubated ON at 37 ºC with vigorous shaking (200 rpm).

II.5.4. Miniprep-based screening of positive clones

After collecting nine transformed *E. coli* str. TOP10 liquid clonal cultures for each transformation, GeneJET™ Plasmid Miniprep kit (Thermo Scientific) was used in order to collect and purify the pDNA, according to manufacturer instructions. Restriction analysis to confirm transformation was performed with *Sal*I for *CroMATE1*-sGFP clones (Table II.8), and with *Xho*I and with *Bgl*II + *Xho*I for sGFP-*CroMATE1* clones (Table II.9). The DoubleDigestion webtool was again consulted to check for the appropriate digestion buffers to be employed. All these reactions were promptly incubated at 37 ºC for 1h, following to a brief incubation at 65 ºC for 20 min.

Table II.8 - Restriction analysis reaction setup for *CroMATE1*-sGFP clones.

Reagents	**Volume (µL)**
*Sal*I	1
10x O buffer	2
pDNA (Mp)	2
H_2O	15
Total	20

Table II.9 - Restriction analysis reaction setups for sGFP-*CroMATE1* clones.

Single digestion		**Double digestion**	
Reagents	**Volume (µL)**	**Reagents**	**Volume (µL)**
*Xho*I	1	*Bgl*II	1
10x R buffer	2	*Xho*I	1
pDNA (Mp)	2	10x Tango™ Buffer	4
H_2O	15	pDNA (Mp)	2
Total	20	H_2O	12
		Total	20

The digested products were individually mixed with loading dye, and loaded into 1 % (w/v) agarose gel prepared with 1x TAE buffer containing 40 mM Tris base, 10% (v/v) acetic acid and 10 mM EDTA, supplemented with 0.5 µg mL^{-1} $EtBr_2$ (Bio-Rad), and ran at 80 V for 25 min using PowerPac Basic Electrophoresis Kit (Bio-Rad). After the electrophoresis, the gels were observed under UV light.

II.5.5. Sequencing of positive clones

Primers were manually designed to sequence *CroMATE1*-sGFP (Table II.10) and sGFP-*CroMATE1* constructs to validate cloning (Table II.11). M13 universal primers were also used. All the primers designed for sequencing purposes were submitted to OligoAnalyzer 3.1 to review the respective biochemical features (Table II.12). The primer synthesis and sequencing services were ordered to STABVida. The results were consulted by aligning the predicted sequences of both fusions with the respective sequenced clones, using MultiAlin (Corpet 1988) (http://multalin.toulouse.inra.fr/multalin/).

Table II.10 - M13 FWD and primers designed to sequence the *CroMATE1*-sGFP clones.

ID	**Sequence (5'→3')**	**Size (bp)**
M13 FWD	GTAAAACGACGGCCAGT	17
sGFP REV	ACGCTGAACTTGTGGCCGTTT	21
MATE1 FWD	GTCTATGTTCACTCAAATATTTTCCGGGC	29

Table II.11 - M13 REV and primers designed and ordered to sequence the sGFP-*CroMATE1* clones.

ID	Sequence (5'→3')	Size (bp)
M13 REV	AGCGGATAACAATTTCACACAGGA	24
sGFP FWD	CGACCACTACCAGCAGAACA	20
MATE1 FWD	GTCTATGTTCACTCAAATATTTTCCGGGC	29

Table II.12 - Biochemical features assigned for each sequencing primer, by using OligoAnalyzer 3.1.

ID	Bp	T_m	GC%	Hairpin	Homodim.
MATE1 FWD	29	57.8	41.4	1.05	-11.69
sGFP FWD	20	56.9	55.0	0.35	-3.61
sGFP REV	21	60.2	52.4	0.24	-9.28

For each construct, three error-free clones were used for midiprep purification of plasmid DNA using the Plasmid Midi Kit (Qiagen), according to manufacturer's instructions, in order to obtain highly pure and concentrated DNA to be used in transient expression assays with *C. roseus* mesophyll protoplasts.

II.6. Subcellular localization of GFP-tagged CroMATE1 fusions in *C. roseus* mesophyll protoplasts

II.6.1. Isolation of *C. roseus* mesophyll protoplasts

C.roseus mesophyll protoplasts were isolated according to Duarte et al. (Duarte, Ribeiro et al. 2011). Leaves of the second or third pair of adult plants were cut into ~1 mm strips, after excising the central vein, and were immediately transferred to a Petri dish with 10 mL of digestion medium composed of 2 % (w/v) cellulose (Onozuka R-10, Duchefa), 0.3 % (w/v) macerozyme (Onozuka R-10, Serva) and 0.1 % pectinase

(Sigma) dissolved in MM buffer (0.4 M mannitol and 20 mM Mes, pH 5.6-5.8), keeping the abaxial face down. The material was vacuum infiltrated in 30 s intervals during 15 min and then incubated at 25 ºC, in the dark, during 3 h. After this incubation, the Petri dishes were placed on an orbital shaker (~60 rpm) for 15 min in the dark and at RT to help release the protoplasts. The suspension was filtered through a 100 µm nylon mesh and the filtrate was transferred into 15 mL falcon tubes. To pellet the protoplasts, the suspension was centrifuged at 65 g for 5 min at 20 ºC. The supernatant was removed, the protoplasts were washed twice in MM buffer and once in cold W5 solution (154 mM NaCl, 125 mM $CaCl_2.2H_2O$, 5 mM KCl and 2 mM Mes, pH 5.7), and the pellet was resuspended in a minimum volume of W5. Protoplasts were counted using a haemocytometer and were incubated on ice for 30 min. After incubation, the suspension was once again centrifuged and the pellet was resuspended in an appropriate volume of MMg buffer (0.4 M mannitol, 15 mM MgCl2 and 4 mM Mes, pH 5.7) to yield a final protoplast concentration of $5x10^6$ cells mL^{-1}.

II.6.2. PEG-mediated transformation of *C. roseus* mesophyll protoplasts

The integrity of the isolated protoplasts was checked by observation under an optical microscope (Olympus) before the transformation. *C.roseus* mesophyll protoplasts were transformed following the procedure by Duarte et al. (Duarte, Ribeiro et al. 2011). Typically, 20 µg of plasmid DNA were mixed with 100 µL of protoplast suspension in a 2 mL round bottom eppendorf. One volume (110 µL) of PEG solution (40 % w/v PEG, 0.2 M mannitol and 0.1 M $CaCl_2.2H_2O$) was added drop by drop to this mixture, flicking the tube after every drop. The tubes were left to incubate for 15 min at RT and then four volumes of W5 solution (440 µL) were slowly added. The mixture was centrifuged at 600 rpm for 2 min, with acceleration and deceleration set at the minimum. The supernatant was removed, the pellet was gently resuspended in 100 µL of W5 solution, the protoplasts were transferred to 15 mL falcon tubes containing 900 µL of W5, and were incubated in the dark at 25ºC, with the tubes lying in a slight slope, for at least 2 days. For the co-transformation with subcellular markers, 15 µg of each construct were mixed with 120 µL of protoplast suspension. The ratios DNA + protoplasts:PEG solution and DNA + protoplasts:PEG solution:W5, were as described above.

II.6.3. Confocal microscopy

Protoplast fluorescence was examined 48 h and 72 h post-transformation, using a Leica SP2 AOBS SE confocal microscope equipped with a scan head coupled to an argon laser. Visualization of GFP was performed using an excitation wavelength of 488 nm and an emission wavelength window from 506 to 538 nm. Visualization of chloroplast autofluorescence was performed using the same excitation wavelength and an emission wavelength window from 648 to 688 nm. Visualization of CFP was performed using an excitation wavelength of 450 nm and an emission wavelength spectrum from 473 to 505 nm. Image analysis was performed with Fiji and GIMP 2.8.4.

II.7. Evolutionary relationship of CroMATE1 and characterized plant MATE-type transporters

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987), using a bootstrap test (500 replicates). The evolutionary distances were computed using the Jukes-Cantor nucleotide-substitution corrective model (Cantor and Jukes 1966). Analyzed codon positions included were $1^{st}+2^{nd}+3^{rd}$, and the sequence codons were first aligned with Muscle, thus using protein-coding sequence information to the subsequent phylogenetic reconstruction. This evolutionary analysis was performed in MEGA5 (Tamura, Peterson et al. 2011). CroMATE1 peptide sequence was aligned with its closest member to evaluate the identity between the two, using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

III. Results

III.1. Identification and *in silico* characterization of a MATE transporter candidate to alkaloid vacuolar transport in *C. roseus,* CroMATE1

In spite of the current knowledge on the TIA biosynthetic pathway in *C. roseus*, little is known about the predicted intra- and intercellular TIA transport events. Previously, the vacuolar accumulation of TIAs in *C. roseus* leaves was shown to be mediated by a H^+ driven antiport transporter in *C. roseus* (Carqueijeiro, Noronha et al. 2013), similarly to what has been observed for nicotine in tobacco, where MATE-type transporters were implicated in this function (Morita, Shitan et al. 2009, Shoji, Inai et al. 2009). Therefore, the presence of a highly abundant MATE-type transporter (Fig. III.1) in a proteomic analysis performed with *C. roseus* vacuolar and tonoplast fractions at the BioNatPro lab, IBMC, clearly indicated that this MATE protein, identified as CroMATE1, was a strong candidate to TIA vacuolar transport.

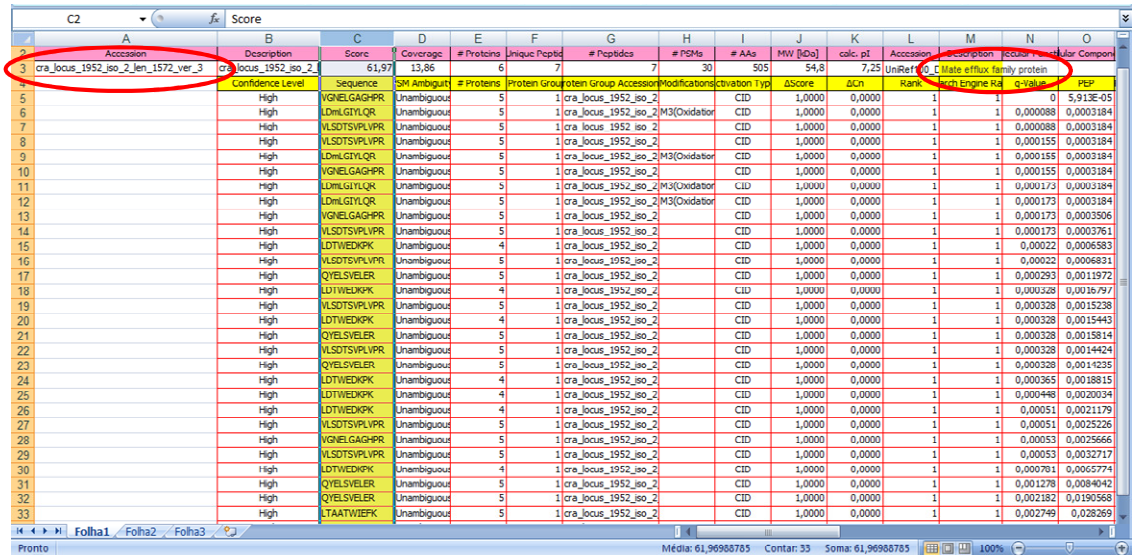


Fig. III.1 - NanoLC-MS/MS Excel-format sheet exhibiting CroMATE1 specific information, namely the MPGR reference (red circle on the left), the functional annotation (red circle on the right), and the respective oligopeptide reads (yellow column).

By using the annotated MPGR accession number, it was possible to retrieve the whole *CroMATE1* CDS-containing cDNA sequence (Appendix1). Immediately after, the ORF detection and translation resulted in the determination of the CDS and the predicted protein sequence, which are represented in Fig.III.2. The putative protein

```
   1  M  G  S  K  Q  N  Y  E  I  N  Q  P  L  L  L  S  N  N  G  G  G  S  S  A  I  T  A  A  E  S
   1  ATGGGTTCCAAACAAAACTATGAAATAAACCAACCACTGTTACTAAGTAACAACGGCGGTGGCAGCAGTGCTATTACAGCTGCAGAGTCG

  31  P  K  A  A  K  T  V  E  E  R  Q  Y  E  L  S  V  E  L  E  R  V  L  S  D  T  S  V  P  L  V
  91  CCAAAGGCGGCTAAGACGGTGGAGGAAAGGCAGTATGAGCTTAGTGTTGAGTTAGAAAGGGTACTTTCCGATACATCTGTACCATTAGTG

  61  P  R  L  T  A  A  T  W  I  E  F  K  L  L  F  R  L  A  A  P  A  A  A  V  Y  L  I  N  Y  V
 181  CCAAGACTCACCGCCGCCACTTGGATTGAGTTTAAGCTTCTCTTTCGGCTGGCGGCTCCAGCAGCCGCCGTTTATTTGATAAACTATGTA

  91  M  S  M  F  T  Q  I  F  S  G  H  L  G  N  L  E  L  A  A  A  S  L  G  N  N  G  I  Q  T  F
 271  ATGTCTATGTTCACTCAAATATTTTCCGGGCATCTTGGGAATCTTGAGCTTGCTGCTGCTTCTCTTGGTAATAATGGTATCCAGACCTTC

 121  A  Y  G  I  M  L  G  M  G  S  A  V  E  T  L  C  G  Q  A  Y  G  A  Q  K  L  D  M  L  G  I
 361  GCCTATGGTATCATGCTTGGAATGGGAAGTGCAGTGGAAACACTATGTGGACAAGCATATGGAGCACAAAAACTAGACATGCTTGGAATA

 151  Y  L  Q  R  S  T  I  L  L  T  L  T  G  I  V  I  M  F  V  Y  I  F  S  K  P  I  L  L  L  L
 451  TATCTCCAAAGATCAACAATTCTTCTAACCCTAACAGGCATTGTTATTATGTTTGTTTACATATTCTCAAAACCAATCCTCCTTTTACTT

 181  G  Q  S  E  A  I  A  A  A  A  A  L  F  T  Y  G  L  I  P  Q  I  F  A  Y  A  A  N  F  P  I
 541  GGCCAATCAGAAGCCATAGCTGCAGCAGCTGCCTTATTTACTTATGGTCTTATCCCACAAATCTTTGCCTACGCAGCCAATTTCCCAATC

 211  Q  K  F  L  Q  A  Q  S  I  V  A  P  S  T  Y  I  S  A  G  A  I  V  F  H  V  L  F  S  W  L
 631  CAAAAGTTCTTACAAGCTCAAAGTATAGTGGCACCTAGTACTTATATATCAGCAGGAGCTATTGTTTTCCATGTTTTGTTTAGTTGGTTA

 241  A  I  Y  K  V  G  L  G  L  F  G  A  S  L  V  L  S  L  S  W  W  V  V  V  V  G  Q  F  I  Y
 721  GCAATTTATAAGGTTGGACTTGGATTGTTCGGGGCATCTTTAGTATTGAGTTTGTCTTGGTGGGTTGTTGTGGTTGGTCAGTTTATTTAT

 271  I  L  Y  S  D  R  T  K  D  T  W  R  G  F  S  V  E  A  F  H  G  L  W  S  F  F  K  L  S  A
 811  ATTCTATATAGTGATAGGACTAAGGACACTTGGCGTGGATTTAGTGTTGAAGCTTTCCATGGACTTTGGAGCTTTTTTAAGTTGTCTGCT

 301  A  S  A  V  M  L  C  L  E  A  W  Y  F  Q  I  L  V  L  L  A  G  M  L  P  D  P  K  I  A  L
 901  GCTTCTGCTGTTATGCTTTGTTTGGAAGCTTGGTATTTTCAGATTCTTGTTCTTTTGGCTGGAATGCTTCCTGATCCTAAAATCGCTTTG

 331  D  S  L  S  I  C  I  T  I  L  G  W  V  F  M  I  A  V  G  F  N  A  A  A  S  V  R  V  G  N
 991  GATTCCCTCTCCATTTGCATTACAATCTTGGGTTGGGTATTCATGATAGCCGTTGGATTCAATGCTGCTGCCAGTGTGAGAGTAGGGAAT

 361  E  L  G  A  G  H  P  R  A  A  A  F  S  V  V  I  V  T  T  M  S  F  I  I  A  V  I  I  S  L
1081  GAACTAGGGGCAGGACATCCAAGGGCAGCTGCATTTTCAGTAGTAATAGTGACAACAATGTCATTCATAATAGCAGTGATAATATCATTA

 391  V  V  L  A  L  R  Y  K  I  S  Y  I  F  T  E  G  E  V  V  S  N  A  V  A  D  M  C  P  L  L
1171  GTGGTACTTGCTTTGCGCTACAAAATTAGCTATATCTTTACCGAAGGTGAAGTTGTAAGCAATGCTGTTGCCGATATGTGTCCCTTGCTC

 421  A  I  T  L  V  L  N  G  I  Q  P  V  L  S  G  V  A  V  G  C  G  W  Q  A  F  V  A  Y  V  N
1261  GCCATCACTCTTGTTCTTAATGGAATTCAACCTGTTTTATCCGGTGTTGCTGTTGGATGTGGATGGCAAGCTTTTGTTGCATATGTGAAC

 451  V  G  C  Y  Y  I  V  G  I  P  T  G  A  L  L  G  F  Y  F  K  L  G  A  K  G  I  W  S  G  M
1351  GTTGGCTGTTATTACATTGTTGGTATCCCAACCGGTGCACTTCTTGGGTTCTACTTCAAACTTGGAGCCAAGGGTATTTGGTCCGGTATG

 481  I  G  G  T  L  M  Q  T  I  I  L  I  W  F  T  Y  R  T  D  W  K  K  E  V  D  I  A  Q  S  R
1441  ATTGGTGGGACATTGATGCAAACTATTATTCTCATCTGGTTTACTTATCGAACGGACTGGAAAAAAGAGGTGGATATAGCTCAAAGTAGG

 511  L  D  T  W  E  D  K  P  K  S  L  S  N  E  -
1531  TTGGATACATGGGAGGACAAGCCAAAATCTTTGTCCAATGAATAA
```

Fig. III.2 - *CroMATE1* full-length CDS and respective peptide sequence. Underlined amino acid residues correspond to the oligopeptide sequences detected by nanoLC-MS/MS.

sequence was aligned with each of the nanoLC-MS/MS detected oligopeptides, revealing a perfect match. Additionally, the protein theoretical pI and molecular weight were predicted (Table III.1).

Table III.1 - Molecular features predicted for *CroMATE1* CDS and the respective protein.

Nucleotide CDS length	1,575 bp
Peptide length	524 aa
Theoretical pI	6.89
Mw	56.82389 kDa

NCBI-pBLAST® analysis of the CroMATE predicted amino acid sequence resulted in the detection of several highly identical MATE-type transporters, namely *TESTA TRANSPARENT*-like transporters whose reference, *AtTT12*, is known to localize in the tonoplast and transport anthocyanins (Marinova, Pourcel et al. 2007). These results have also shown high degrees of identity (~70%) with CroMATE1 and null *E* values, corroborating the MATE-type nature of this candidate transporter (Table III.2).

Table III.2 - NCBI-pBLAST® output obtained for the CroMATE1 predicted amino acid sequence, presenting the top ten best hits and score criteria.

Description	Score	Query cover	*E* value	Ident	Accession
PREDICTED: protein TRANSPARENT TESTA 12-like isoform X2 [Cicer arietinum]	747	92%	0.0	73%	XP_004515071.1
PREDICTED: protein TRANSPARENT TESTA 12-like isoform X1 [Cicer arietinum]	739	92%	0.0	72%	XP_004515070.1
Transparent testa 12 protein [Medicago truncatula] >gb|AES67243.1| Transparent testa 12 protein [Medicago truncatula]	738	99%	0.0	70%	XP_003596992.1
PREDICTED: LOW QUALITY PROTEIN: protein TRANSPARENT TESTA 12-like [Glycine max]	737	99%	0.0	69%	XP_003540839.1
PREDICTED: protein TRANSPARENT TESTA 12-like [Glycine max]	731	98%	0.0	69%	XP_003526327.1
PREDICTED: protein TRANSPARENT TESTA 12-like [Solanum lycopersicum]	729	100%	0.0	71%	XP_004251515.1
PREDICTED: protein TRANSPARENT TESTA 12-like [Glycine max]	727	95%	0.0	71%	XP_003540303.1
MATE efflux family protein isoform 1 [Theobroma cacao]	724	99%	0.0	72%	EOY01465.1
unnamed protein product [Vitis vinifera]	723	99%	0.0	69%	CBI28937.3
PREDICTED: protein TRANSPARENT TESTA 12 [Vitis vinifera]	721	92%	0.0	73%	XP_002279330.1

The transmembrane domain (TMD) prediction analysis of CroMATE1 using both HMMTOP (Fig. III.3, A) and PredictProtein (Fig. III.3, B) uniformly identified the very same 12 transmembrane, hydrophobic α-helices, typical of MATE-like transporter structures, hence supporting the function attributed to this candidate.

Finally, pursuing the *in silico* analysis of the CroMATE1 amino acid sequence, the latter was submitted to WoLF PSORT to provide clues concerning CroMATE1 predicted subcellular localization. Interestingly enough, the higher score was found for vacuolar localization ("*vacu*"), 7.0, rather than plasma membrane (PM) ("*plas*"), 4.0, and Golgi apparatus ("*golg*"), 2.0 (Fig. III.4).

Taken together, all the *in silico* analyses seem to be consistent in confirming the MATE family assignment of CroMATE1 and in predicting this protein to possess 12 TMDs and to localize in the tonoplast.

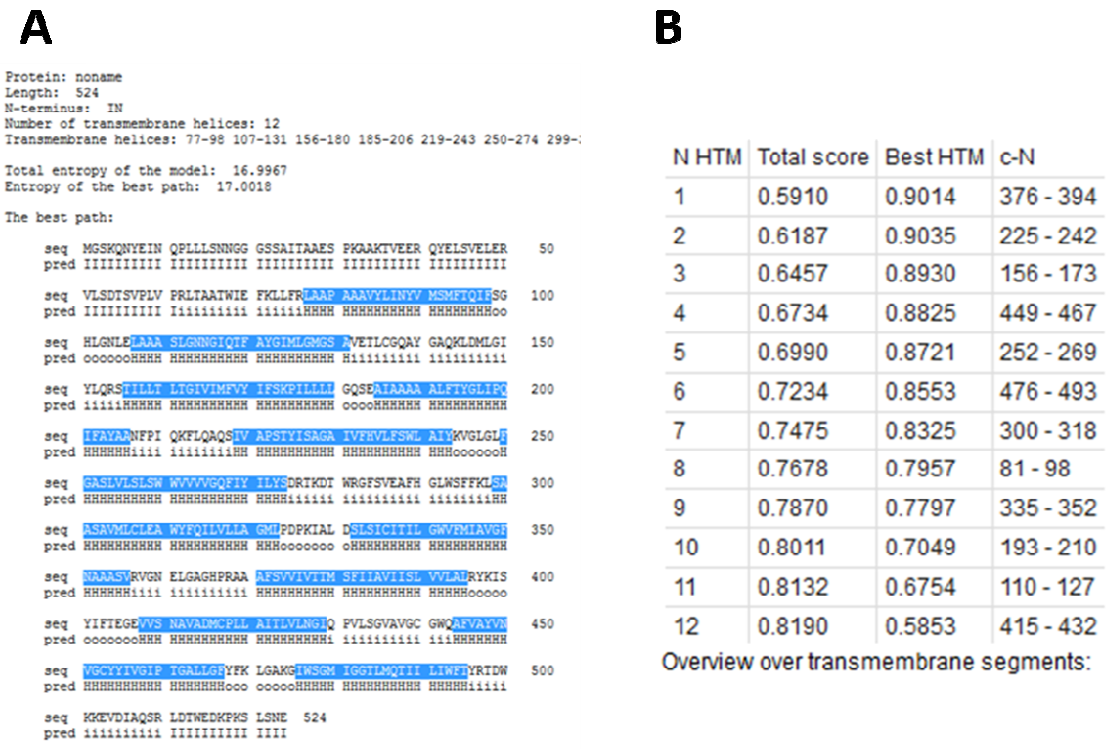


Fig. III.3 - CroMATE1 protein predicted transmembrane domains (TMDs) by using (A) HMMTOP and (B) PredictProtein. Blue - distribution of the predicted TMDs within CroMATE1 protein sequence.

k used for kNN is: 14

queryProtein details vacu: 7.0, plas: 4.0, golg: 2.0

Normalized Feature Values

id	site	iPSORT		PSORT Features													Amino Acid Content							Misc.
		MxHy1 30	Mx-1 20	dna	erl	mlb	m3a	mNt	mip	mit	nuc	pox	psg	rib	rnp	tms	vqr	A	C	Q	H	I	L	S	V	length
queryProtein	vacu?	24	20	47	50	48	97	49	21	14	30	49	96	49	50	98	44	88	42	64	12	98	97	51	75	76
At3g51860.1	vacu	70	6	47	50	48	97	49	21	15	30	49	38	49	50	98	44	68	39	47	30	98	100	62	61	69
At1g16780.1	golg	14	45	47	50	48	97	49	21	28	30	49	38	49	50	100	44	88	40	15	10	95	79	45	88	90
At1g78920.1	golg	9	45	47	50	48	97	49	21	28	30	49	38	49	50	100	44	92	47	12	10	97	75	46	83	90
At2g38290.1	plas	17	6	47	50	48	97	49	21	12	30	49	38	49	50	97	44	95	18	44	44	63	92	6	76	71
CHL1_ARATH	plas	12	1	47	50	48	97	49	21	13	30	49	38	49	50	100	44	66	60	36	39	69	98	19	57	83
At3g13320.1	vacu	82	20	47	50	48	97	49	21	16	30	49	38	49	50	97	44	58	75	75	50	95	98	73	58	67
At2g47600.1	vacu	47	6	47	50	48	97	49	21	24	30	49	38	49	50	96	44	63	58	11	51	96	89	60	90	78
At1g15690.1	vacu	96	6	47	50	48	97	49	21	13	30	49	38	49	50	100	44	91	61	5	13	97	53	50	82	90
CLCD_ARATH	plas	11	6	47	50	48	97	49	21	25	30	49	38	49	50	98	44	36	72	34	53	86	95	57	60	90
FD3E_SOYBN	E.R.	15	45	47	50	48	97	49	21	17	30	49	81	49	50	92	44	8	36	43	100	87	92	55	22	57
VATL_BETVU	vacu	92	6	47	50	48	97	49	21	10	30	49	38	49	50	94	44	98	45	13	9	99	82	55	85	14
At2g36910.1	plas	29	20	47	50	48	47	49	21	14	30	49	38	49	50	97	44	86	39	70	52	76	75	69	65	99
VATL_PHAAU	vacu	92	20	47	50	48	97	49	21	26	30	49	38	49	50	94	44	99	46	14	9	99	83	66	77	14
VATL_AVESA	vacu	92	6	47	50	48	97	49	21	21	30	49	38	49	50	94	44	99	45	13	9	99	82	55	85	14

Fig. III.4 - WoLF PSORT output after submitting the CroMATE1 predicted amino acid sequence. Red square - likelihood attributed to the vacuolar localization of CroMATE1 protein.

III.2. Hierarchical co-expression clustering of *CroMATE1* with TIA biosynthetic genes

In order to investigate the potential co-regulation of *CroMATE1* with characterized genes involved in the biosynthesis of TIAs in *C. roseus*, which could support the involvement of CroMATE1 in this pathway, the hierarchical co-expression clustering (HCL) approach seemed to be appropriate. By using specific *C. roseus* gene expression data available at the MPGR database, HCL was performed for *CroMATE1* together with genes encoding TIA biosynthetic enzymes and isoforms, plus several other candidate MATE-type and ABC-type transporters and transcription factors identified in the BioNatPro lab (Fig. III.5). The HCL analysis resulted in the identification of a robust, pattern-specific cluster across the 23 set of samples. Additionally, the SOTA analysis resulted in the graphical display of the centroid distribution of gene expression profiles of all of those found in the very same cluster of interest (Fig. III.6), thus supporting the results obtained with HCL.

CroMATE1 expression profile seems to be highly correlated with that of a broad set of genes involved in the TIA biosynthesis (e.g. *STR*, *SLS*, *D4H*, *G10H*, *DAT*), and several candidate transcription factors, ABC-type transporters and other MATE-like

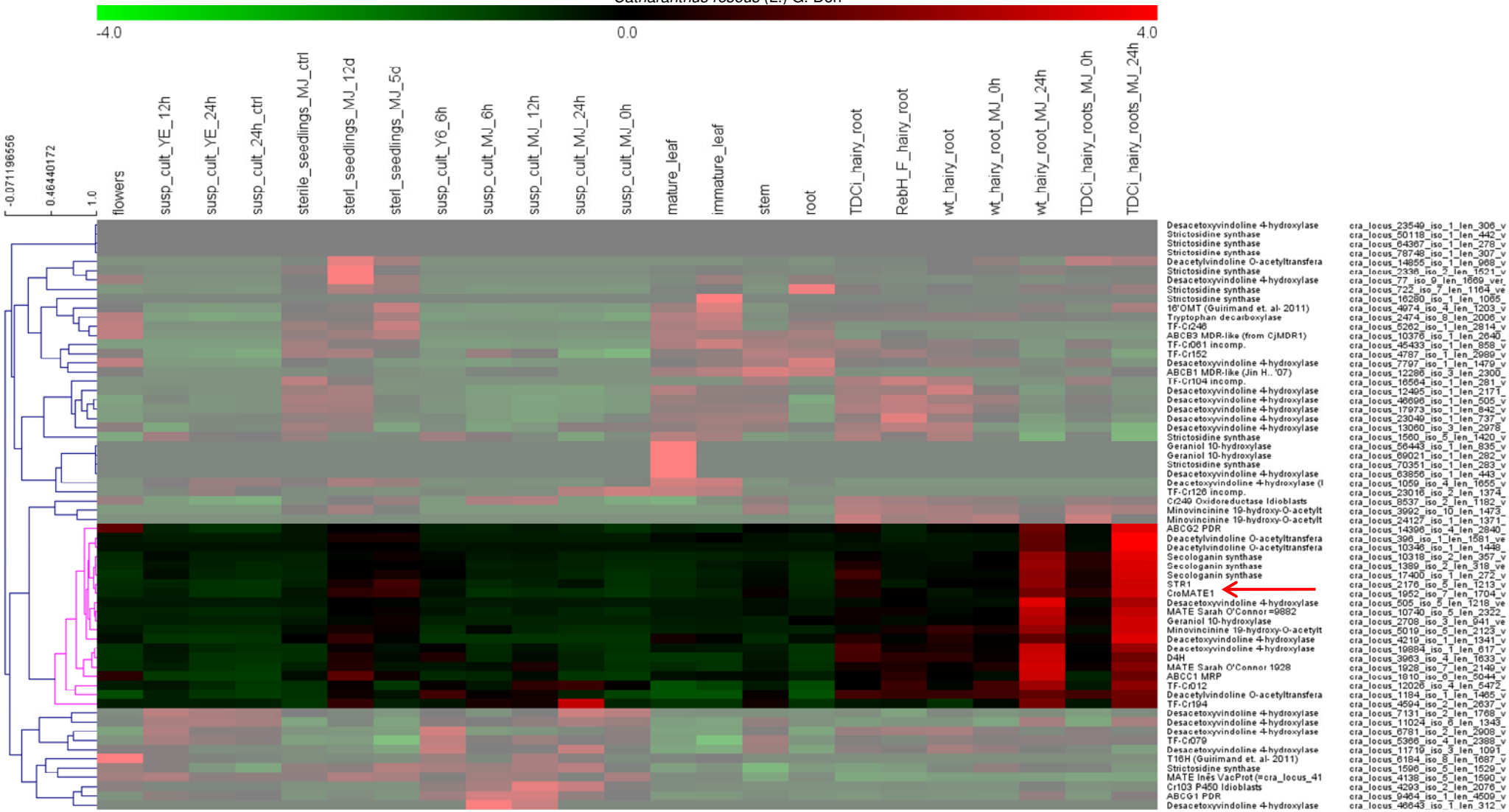


Fig. III.5 - HCL analysis output from MeV 4.9.0 after submitting and normalizing the gene expression data of interest, highlighting the cluster of interest defined by a pink tree-branch with a height of about 0.5 within Pearson correlation value range. Red arrow - *CroMATE1* expression profile.

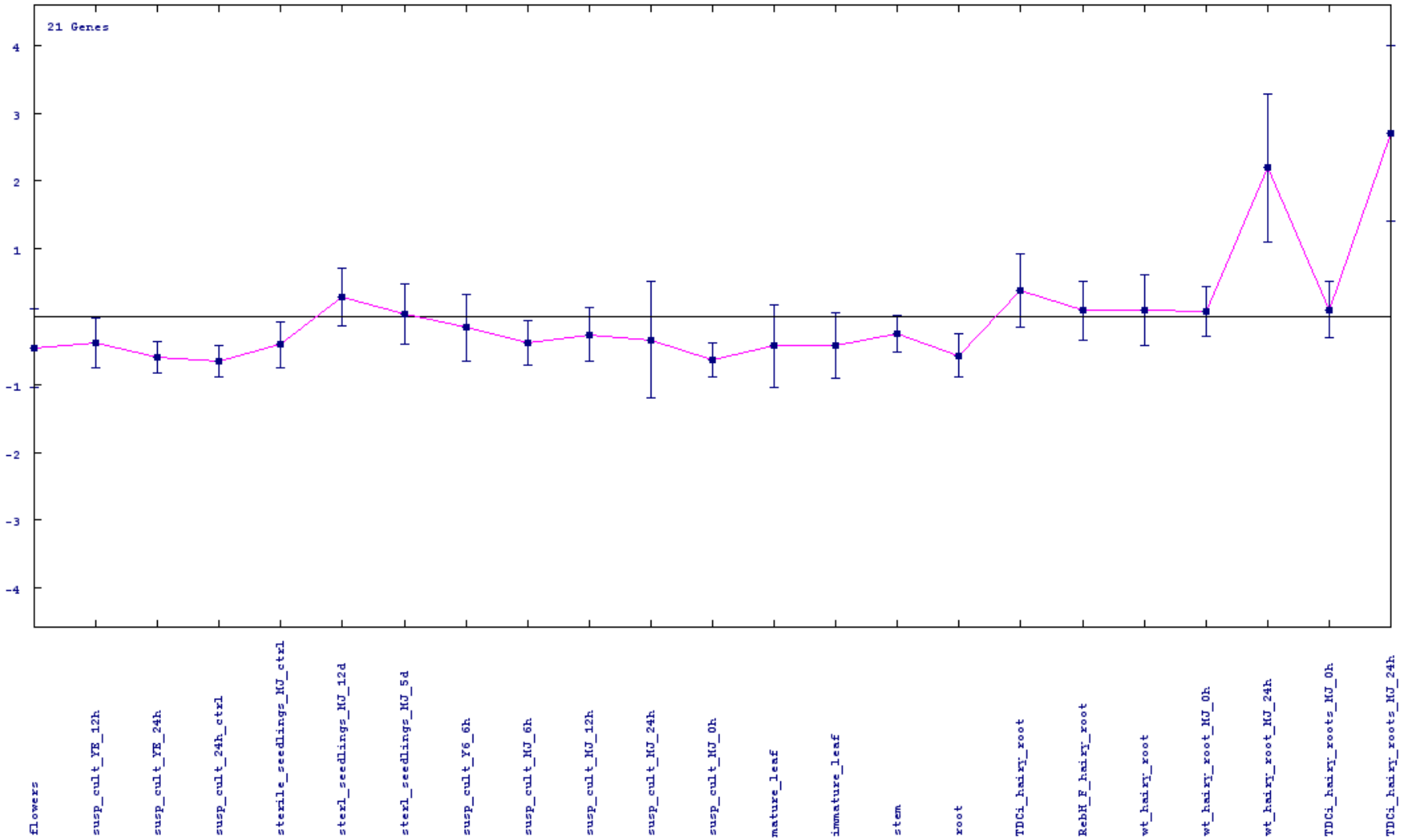


Fig. III.6 - Centroid distribution of the gene expression profiles of genes contained in the cluster of interest, including *CroMATE1*, originated from a SOTA analysis with MeV.

transporters, supported by the cluster tree-branch whose height corresponds to a significant Pearson correlation value of about 0.5. Furthermore, such correlation seems to be supported by the uppermost, higher expression values found in samples *sterl_seedlings_MJ_12d*, *sterl_seedlings_MJ_5d*, *wt_hairy_root_MJ_24h* and *TDCi_hairy_roots_MJ_24h*, curiously the only four samples analysed after MeJA (MJ, methyl jasmonate) was applied. Also, running the algorithm with different combinations of other metrics and clustering methods resulted in the identification of the very same cluster (data not shown), reinforcing the robustness of this cluster.

III.3. Isolation and cloning of the *CroMATE1* full-length CDS

In order to generate the tools for the functional characterization of CroMATE1, the isolation and cloning of the respective CDS was performed. Since the first goal was to investigate the subcellular localization of CroMATE1, highly relevant for its putative function, it was decided to follow a strategy leading directly to the production of GFP-tagged fusions in a vector suitable for plant transformation. Keeping in mind that very little is known about sorting signals and mechanisms of membrane proteins, it was planned to engineer two types of *GFP*-tagged *CroMATE1* constructs, *CroMATE1-sGFP* or N-terminal fusion (Fig. III.7, A), and *sGFP-CroMATE1* or C-terminal fusion (Fig. III.7, B). This approach included the accurate informatic sequence analysis of the resulting fusion constructs. As a result, both predicted complete sequences seemed to fulfill all the requirements necessary for effective protein expression, stability and integrity (Appendix2, Appendix3).

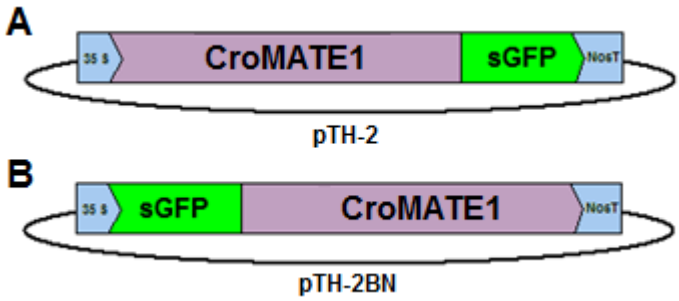


Fig. III.7 - Schematic representation of A) N-terminal fusion contruct, *35S::CroMATE1-sGFP*, harbored by the pTH-2 cloning vector and B) C-terminal fusion construct, *35S::sGFP-CroMATE1*, harbored by the pTH-2BN cloning vector.

In order to perform the PCR amplification of the *CroMATE1* CDS, three pairs of primers were used. One pair was designed to amplify the CDS, the other two pairs were exactly the same plus extensions adding the restriction sites needed for cloning in pTH2 and in pTH2-BN. This was done, since the primers adding extensions are less efficient than the ones fully annealing with the target, and it was predicted that a double round PCR would be needed starting with the primers without extensions. Surprisingly, the PCRs performed with all the three pairs of primer revealed to yield a single, consistent band with an observable length slightly greater than 1,500 bp (1.5 kbp), as expected (Fig. III.8).

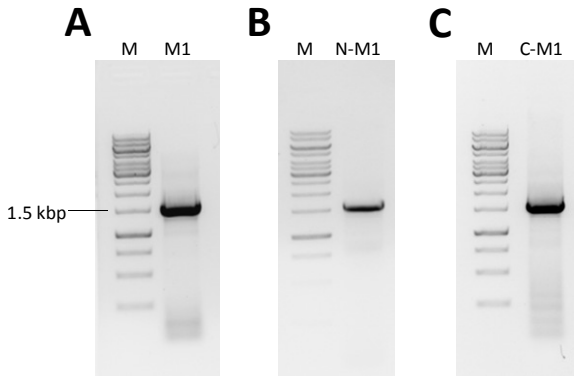


Fig. III.8 - *CroMATE1* CDS PCR-amplification products in electrophoresis agarose gel. A) *CroMATE1* CDS (M1); B) *CroMATE1* CDS for cloning into pTH-2 (N-M1); C) *CroMATE1* CDS for cloning into pTH-2BN (C-M1). M - GeneRuler™ 1 kb DNA Ladder Mix (Thermo Scientific). All observed bands seem to possess slightly over 1.5kbp.

After ligation of the correspondent *CroMATE1* CDS products with pTH-2 (Fig. III.9, A) and pTH-2BN (Fig. III.9, B) and transformation of *E. coli* TOP10, the miniprep-based harvest of pDNA in each set of 9 clones was taken to advance into restriction analysis and hence, to the screening of positive clones in agarose gel.

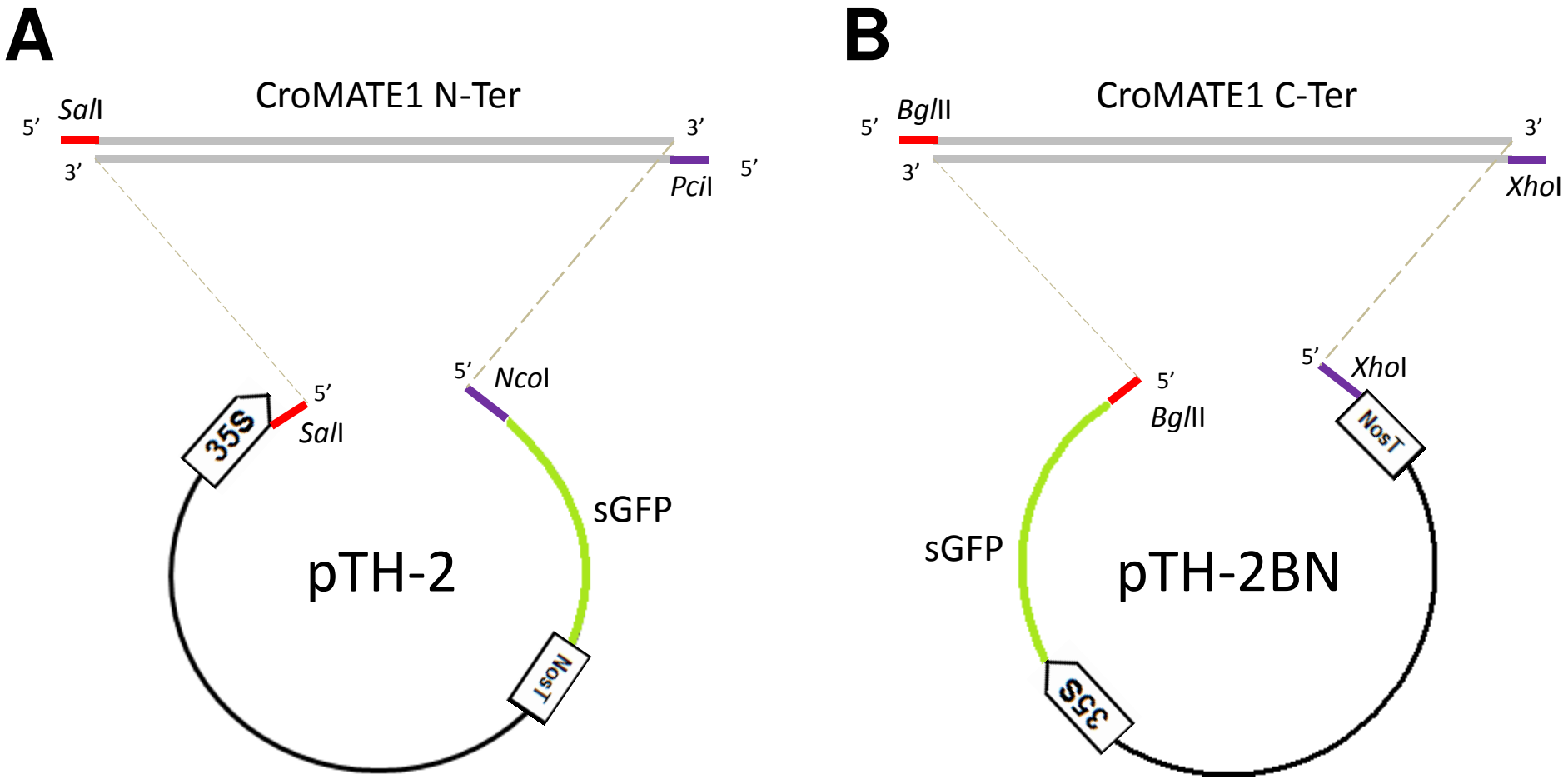


Fig. III.9 - Schematic representation of digested *CroMATE1* CDS PCR products and the respective cleaved cloning vectors, displaying the established single-direction insertion. Grey - *CroMATE1* CDS. Green - *sGFP* nucleotide sequence. Red/Purple - overhangs generated by the action of the restriction enzymes, compatible with overhangs with the same color. For simplicity, cloning vectors were drawn as a single line despite only the red/purple restriction sites have a single-strand nucleotide sequence, with obvious direction. This representation was not drawn to scale.

The restriction analysis performed on *CroMATE1-sGFP* clones revealed four potential positive clones, *i.e.* 1, 2, 4 and 9 (Fig. III.10). The digested empty pTH-2 cloning vector, used as the reaction control, seems to be slightly longer than 4,000 bp (4 kbp), which is concordant with the empty plasmid length. Thus, digested clones presenting such length were most likely not-recombinant, empty pTH-2 cloning vectors. In the other hand, digested clones 1, 2, 4 and 9 exhibited a sequence length of approximately 6,000 bp (6 kbp), in agreement with the expected length for pTH-2 harboring *CroMATE1*. Thus, these four clones were pinpointed as candidates for sequencing.

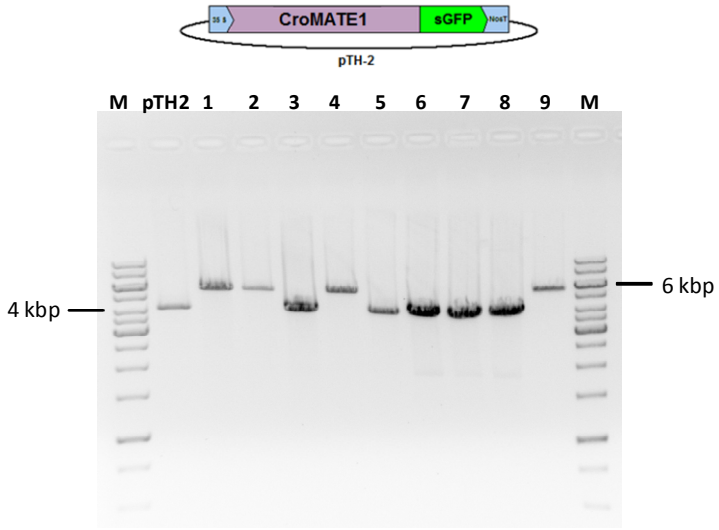


Fig. III.10 - Restriction analysis of *CroMATE1-sGFP* clones with *Sal*I in electrophoresis agarose gel. pTH2 - empty pTH-2 cloning vector; 1-9 - *CroMATE1-sGFP* clones; M - GeneRuler™ 1 kb DNA Ladder Mix (Thermo Scientific).

The restriction analysis performed on *sGFP-CroMATE1* clones revealed positive clones only, exhibiting a sequence length of approximately 6,000 bp (6 kbp), in agreement with the expected length for pTH-2BN harboring CroMATE1 (Fig. III.11, A). The digested empty pTH-2BN cloning vector, used as the reaction control, seems to be slightly longer than 4,000 bp (4 kbp), which is concordant with the empty plasmid length. Clones 2, 3, 7 and 9 were further selected for one more round of restriction analysis that employed double digestion, with the intent of excising CroMATE1 fragment from the plasmid. As expected, all four clones harbored CroMATE1 (Fig. III.11, B), presenting two different bands: *i)* one with the same length of empty pTH-2BN and *ii)* a smaller fragment found to possess about 1,500 bp (1.5 kbp), in agreement with the size of the *CroMATE1* CDS. Digested empty pTH-2BN presented a single band slightly longer than 4,000 bp (4 kbp), as expected. Therefore, these four positive clones were pinpointed as candidates for sequencing.

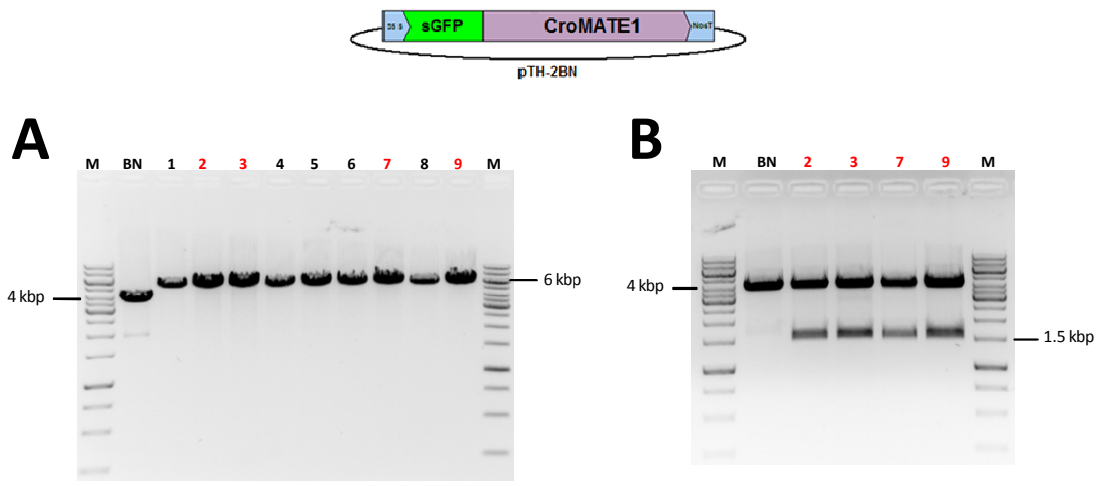


Fig. III.11 - Restriction analysis of *sGFP-CroMATE1* clones in electrophoresis agarose gel. A) Restriction analysis with *Xho*I. BN - empty pTH-2BN cloning vector; 1-9 - *sGFP-CroMATE1* clones; Red - clones selected for the additional restriction analysis with double digestio; B) Restriction analysis of selected clones with *Bgl*II and *Xho*I. BN - empty pTH-2BN cloning vector; 2, 3, 7 and 9 - selected *sGFP-CroMATE1* clones. M - GeneRuler™ 1 kb DNA Ladder Mix (Thermo Scientific).

For the *CroMATE1-sGFP* clones, miniprep samples of clones 1, 4 and 9 were submitted to sequencing, whereas for *CroMATE1-sGFP*, miniprep samples of clones 2, 3 and 9 were also submitted to sequencing. After retrieval, the sequencing files were aligned for the individual clones using the respective predicted fusion sequence, *CroMATE1-sGFP* or N-terminal fusion (Appendix2) and *sGFP-CroMATE1* or C-terminal fusion (Appendix3). The primers designed for this step were strategically placed so that the resulting sequencing for each pair of primers would overlap the following amplicon, to overcome the limitation of quality sequencing being obtained only for about 800bp-long. Despite the sequencing errors found further away from this limit, and furthermore those found in the vicinity of the primers pairing site, these overlaps establish a better, more accurate assessment of the sequence present in each clone. By addition, consulting and combining the chromatograms associated with each individual sequencing result (data not show) was used to rule out sequencing errors.

All the submitted *CroMATE1-sGFP* clones 1, 4 and 9 (Appendix4) had an identical sequence to that of the predicted fusion sequence, with the exception of a single nucleotide, C at position 1,428, found in all the different clones, in contrast to the T at the same position observed in the MPGR *CroMATE1* CDS sequence. Similarly, the submitted *sGFP-CroMATE1* clones 2, 3, and 9 (Appendix5) had an identical sequence to that of the predicted fusion sequence, with the very same exception as above, of a single nucleotide, C at position 1,428, found in all the different clones. In

attempt to clarify what this substitution found in all clones of both constructs would mean, a *CroMATE1* CDS consensus sequence was built from all the alignment results and aligned with the original, predicted *CroMATE1* CDS sequence (Fig. III.12). Effectively, the only difference stands for a C at position 1,428, instead of a T base-pair, as stated before. Once this situation was confirmed for the nucleotide sequences, it was planned to check for the resulting amino acid sequence and to see if this substitution could be a synonymous rather than a non-synonymous substitution. The alignment showed absolutely no differences between the CroMATE1 consensus amino acid sequence and the original CroMATE1 predicted polypeptide, indicating that the mismatch was a synonymous substitution most likely corresponding to a polymorphism (Fig. III.13).

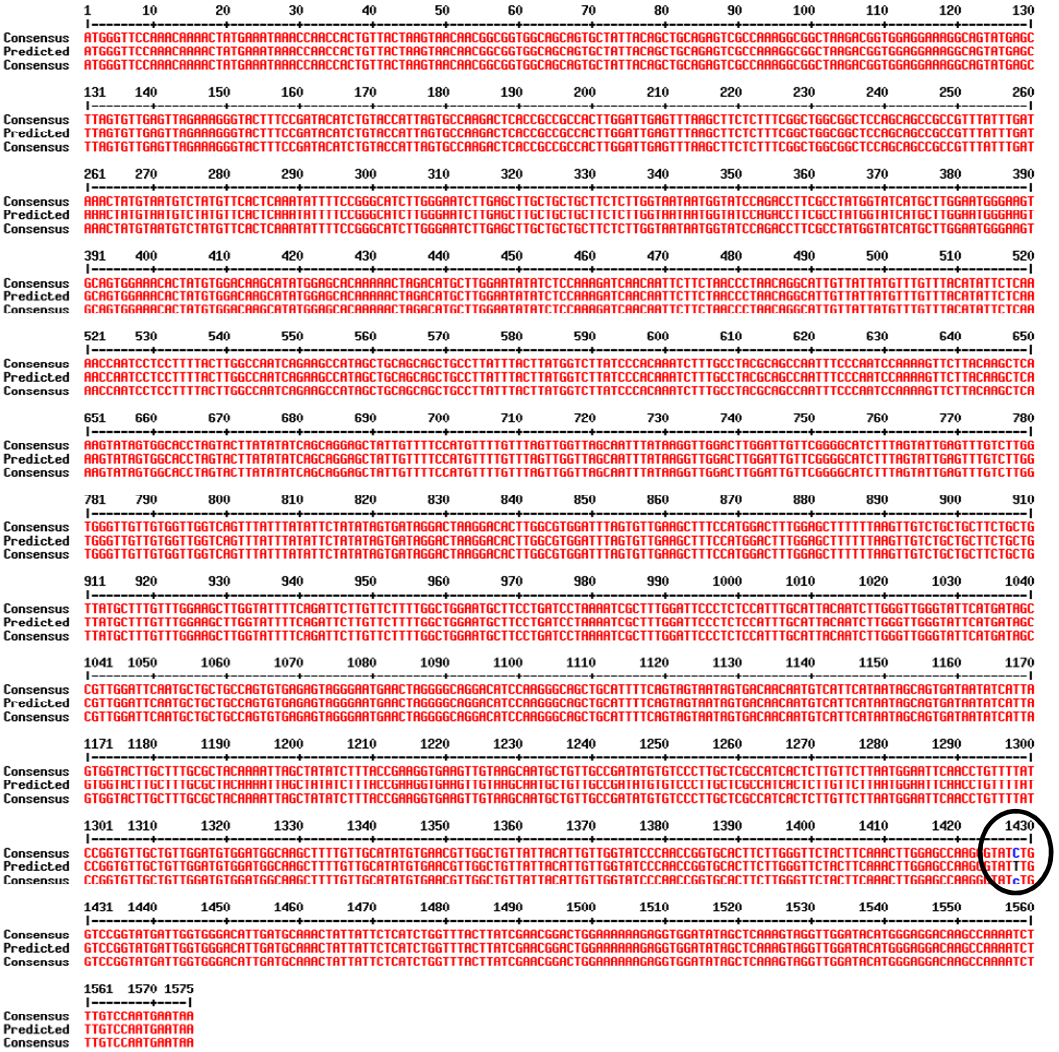


Fig. III.12 - Alignment of the cloned *CroMATE1* CDS consensus sequence and the MPGR *CroMATE1* CDS. Black circle - C/T mismatch at position 1,428.

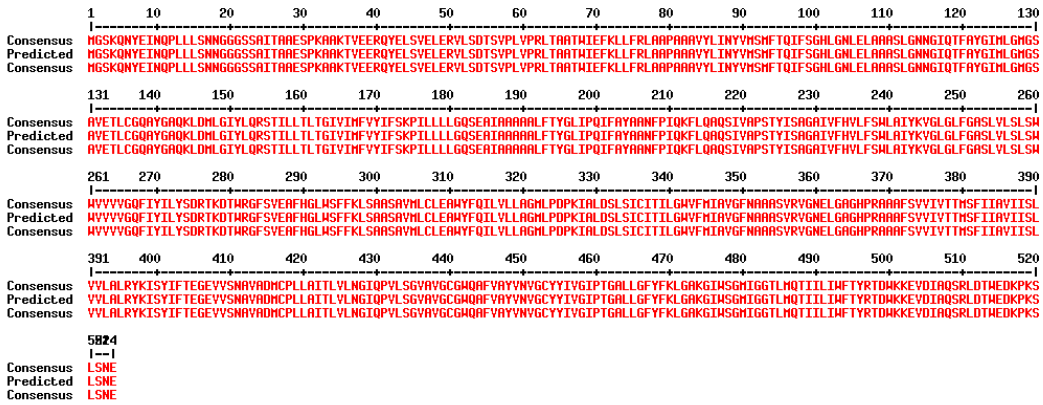


Fig. III.13 - Alignment of the consensus predicted amino acid sequence of the cloned *CroMATE1* and the MPGR amino acid sequence for CroMATE1, here demonstrated to be identical.

III.4. Subcellular localization of CroMATE1 in *C. roseus* mesophyl protoplasts

In spite of all the clues obtained with the computational analysis, the experimental determination of CroMATE1 subcellular fate using the *GFP* fusion constructs previously prepared was critical to confirm the subcellular localization of CroMATE1. Thus, the *GFP* fusions and the marker constructs represented in Fig. III.14 were used for transient expression in *C. roseus* mesophyll protoplasts. The transformed protoplasts were observed under a confocal microscope to investigate the subcellular localization of the protein fusions by detection of their green fluorescence.

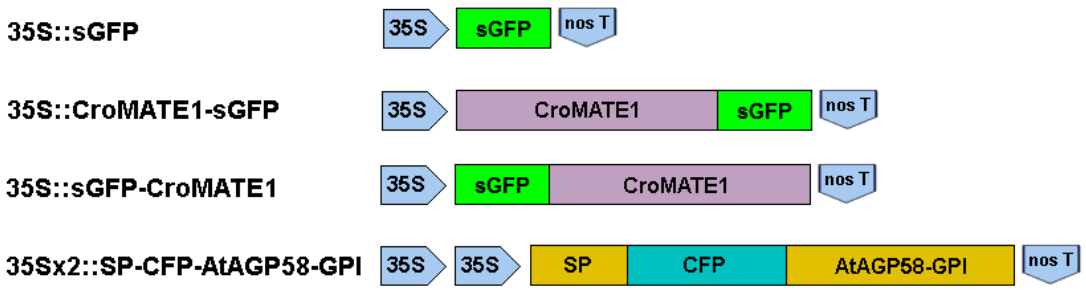


Fig. III.14 - Schematic representation of the constructs used for PEG-mediated transient expression in *C. roseus* mesophyll protoplasts to investigate CroMATE1 subcellular localization. The last construct encodes a plasma membrane marker.

The first protoplasts observed at 48 hours post-transformation (PT) were the ones transformed with the empty pTH-2 cloning vector, which expresses sGFP alone, expected to be found throughout the cytosol. As can be seen in Fig. III.15, sGFP was found fulfilling the cytosol and the nucleoplasm of protoplasts, not marking the chloroplasts, or the vacuolar lumen.

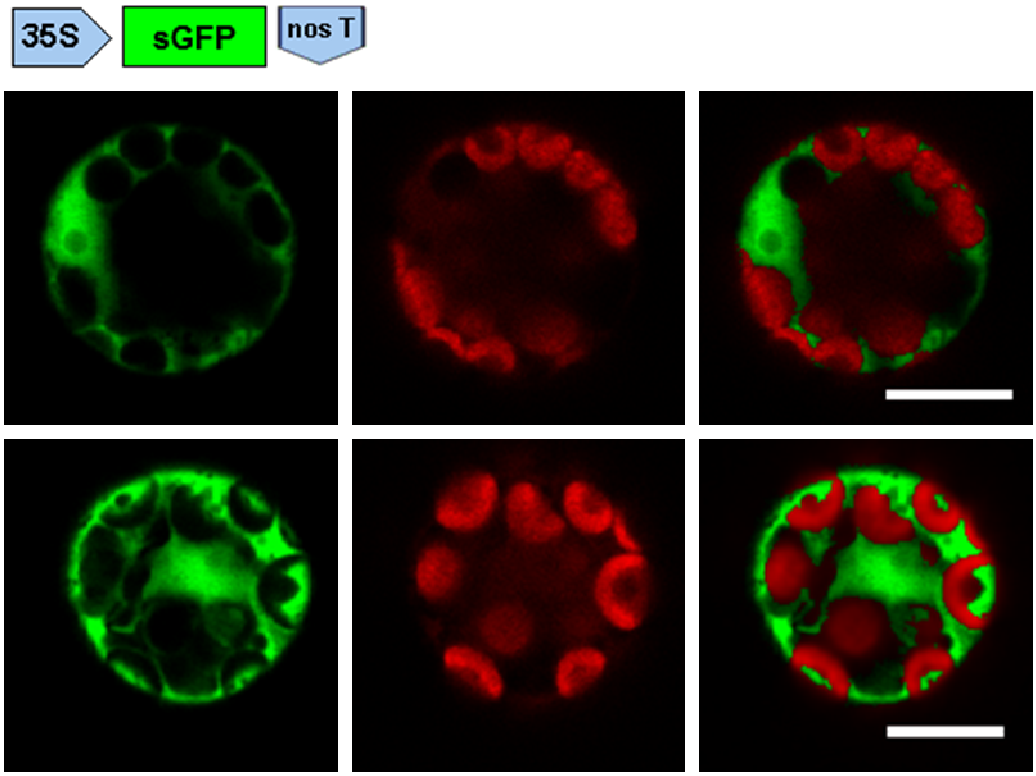


Fig. III.15 - Transient expression of sGFP in *C. roseus* mesophyll protoplasts observed under the confocal microscope 48h after transformation. A schematic representation of the construct used is shown on top of the set of images. Left – GFP channel; middle – red channel showing chloroplast autofluorescence; right – merged images. Bars = 10 µm.

The results obtained 48 hours after transformation, for the transient expression of the N-terminal and C-terminal fusions of CroMATE1 with GFP are shown in Fig. III.16 and III.17. To allow a better conclusion about the exact subcellular localization of the CroMATE1 fusions, the construct *sGFP-CroMATE1* was co-expressed with the construct *SP-CFP-AtAGP58* which encodes an arabinogalactan protein localized at the plasma membrane. The results obtained 48 hours after co-transformation are shown in Fig. III.18.

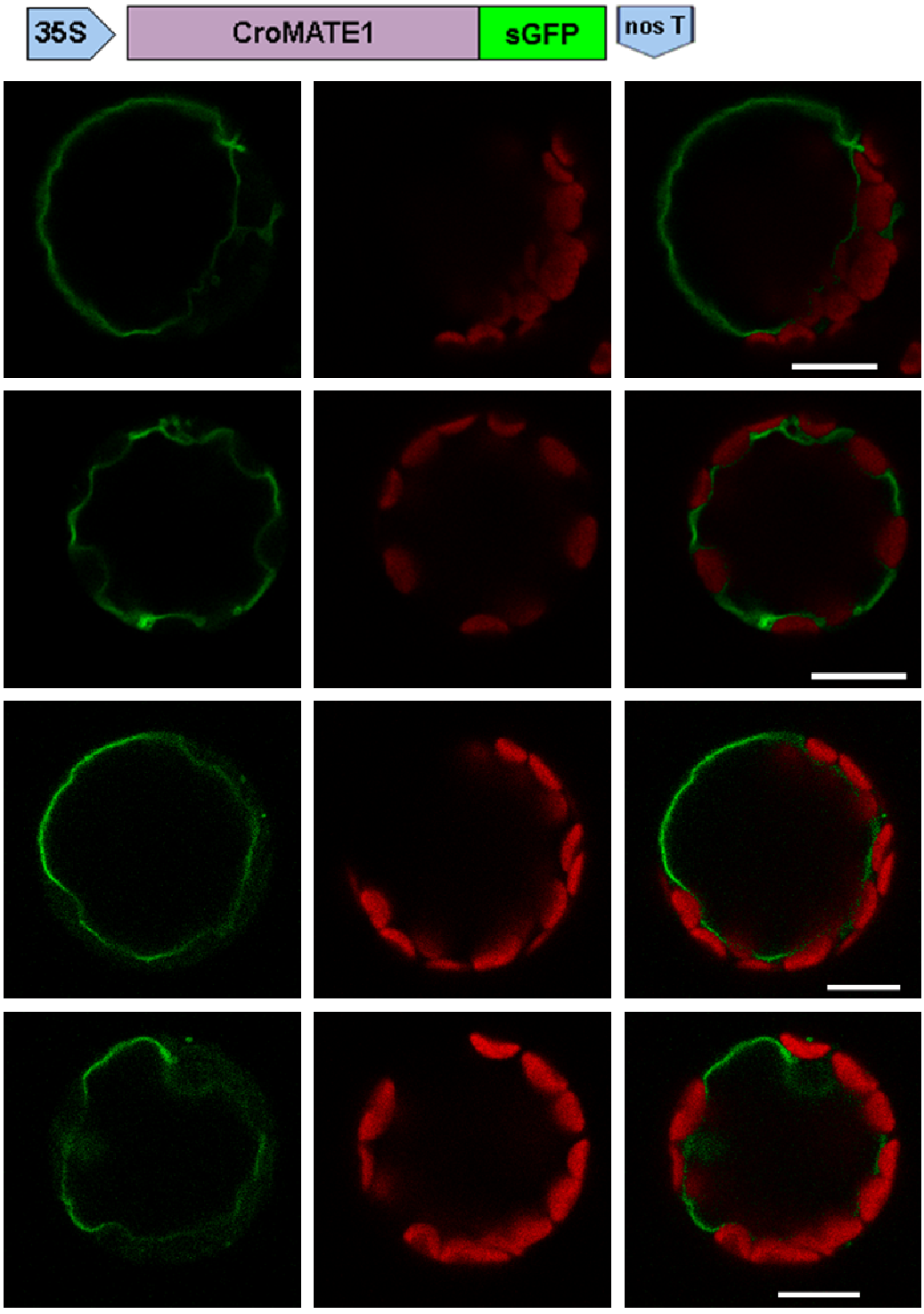


Fig. III.16 - Transient expression of CroMATE1-sGFP in *C. roseus* mesophyll protoplasts observed under the confocal microscope 48 h after transformation. A schematic representation of the construct used is shown on top of the set of images. Left – GFP channel; middle – red channel showing chloroplast autofluorescence; right – merged images. Bars = 10 µm.

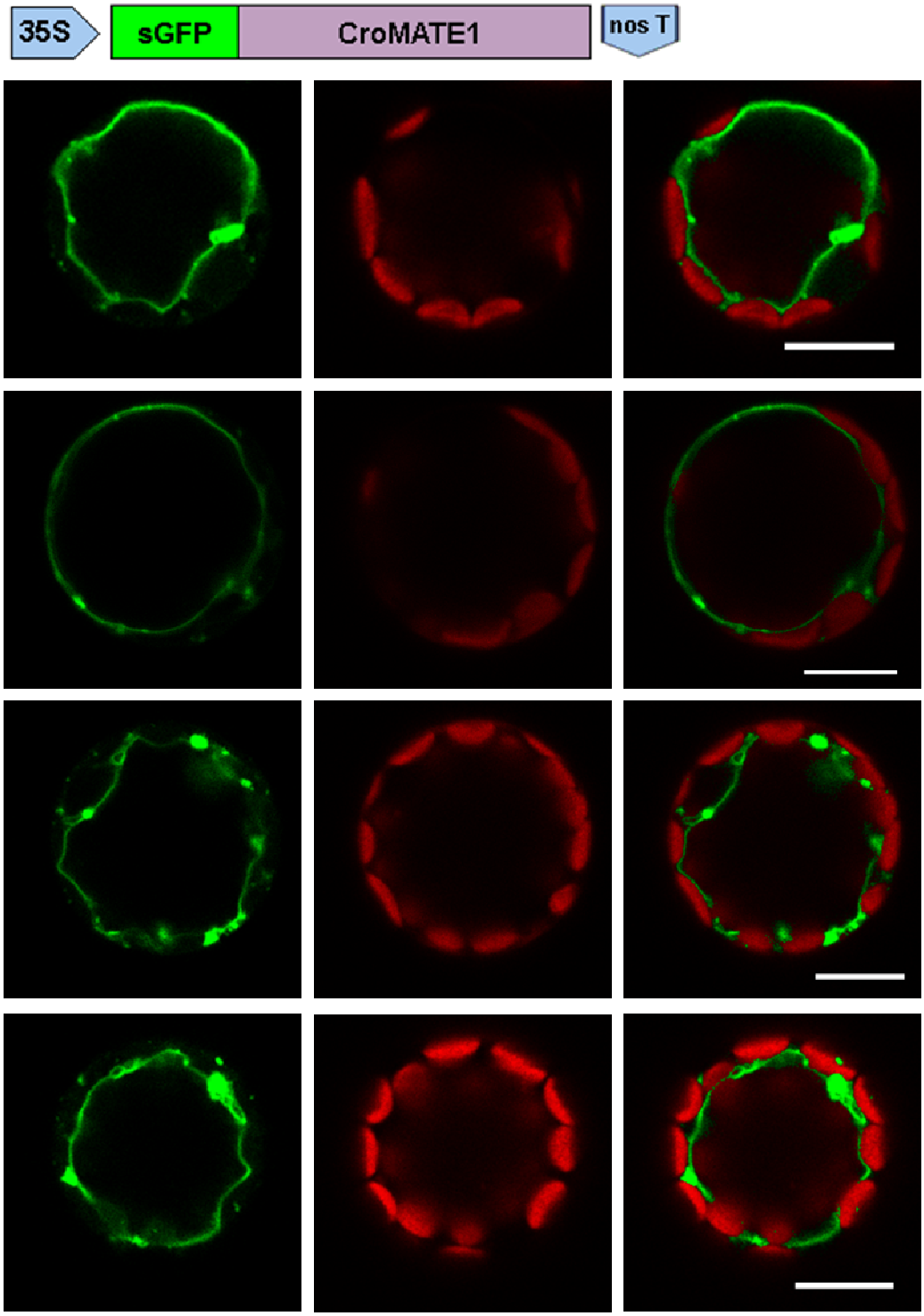


Fig. III.17 - Transient expression of sGFP-CroMATE1 in *C. roseus* mesophyll protoplasts observed under the confocal microscope 48 h after transformation. A schematic representation of the construct used is shown on top of the set of images. Left – GFP channel; middle – red channel showing chloroplast autofluorescence; right – merged images. Bars = 10 µm.

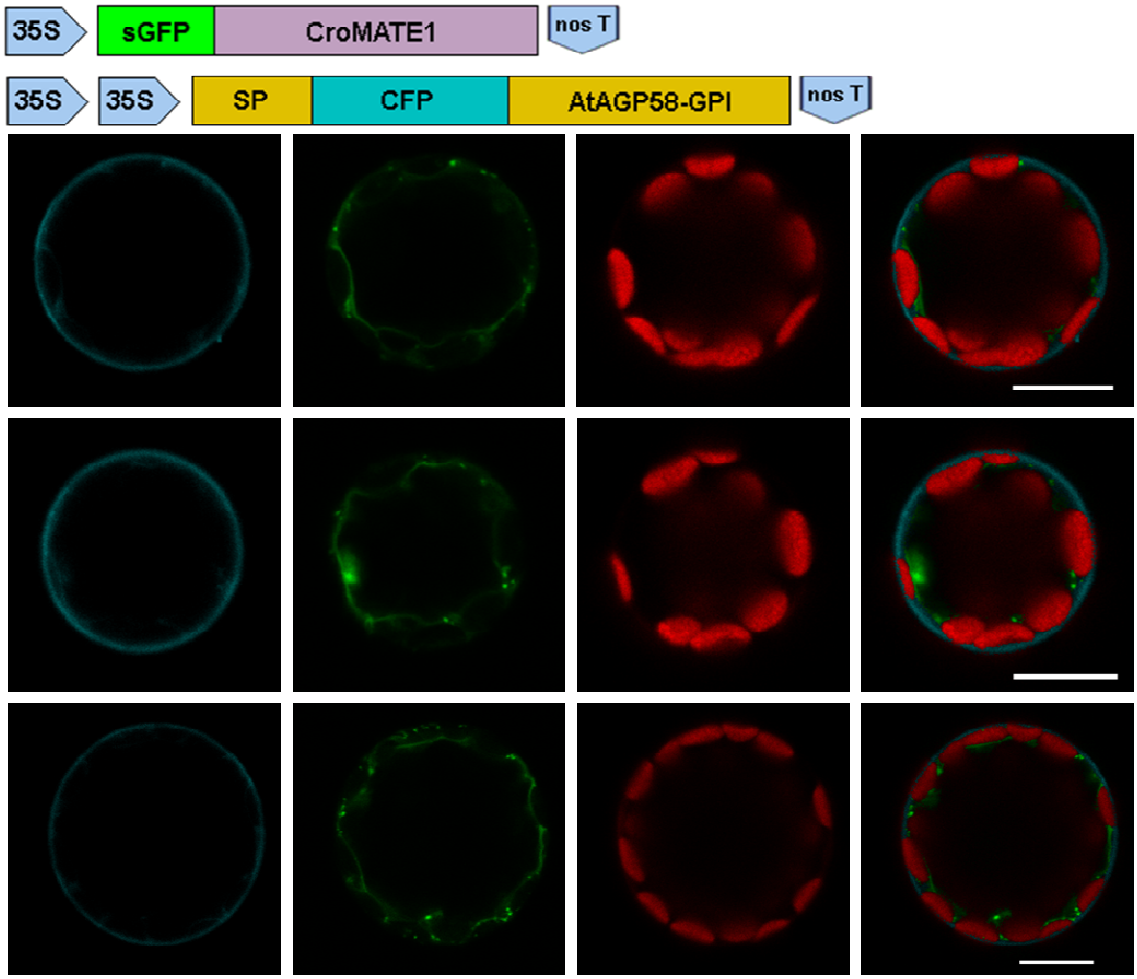


Fig. III.18 - Transient co-expression of sGFP-CroMATE1 and the plasma membrane marker CFP-AtAGP58 in *C. roseus* mesophyll protoplasts observed under the confocal microscope 48 h after transformation. A schematic representation of the construct used is shown on top of the set of images. First column – CFP channel; second column – GFP channel; third column - red channel showing chloroplast autofluorescence; fourth column – merged images. Bars = 10 µm.

As can be observed in Fig. III.16 to III.18, 48 hours after transformation, both the N-terminal and C-terminal CroMATE1 fusion proteins exhibited, remarkably and unequivocally, a green fluorescence at the tonoplast. Such fact is strongly supported by the localization of labeling which clearly surrounds the chloroplasts always from the inside, as expected for the tonoplast. Furthermore, the co-expression assay clearly discriminated the plasma membrane from the membranous structure marked by the CroMATE1 fusion protein, which can only be the tonoplast. It is interesting to note that there were no significant differences found between the expression patterns observed for the two different CroMATE1 fusion proteins, with the exception that the sGFP-CroMATE1 fusion often generated small, spherical, densely green bodies, suggesting a possible aggregation of this type of fusion protein.

The subcellular localization of the CroMATE1-sGFP and the sGFP-CroMATE1 fusion proteins were also observed 72 h after transformation (Fig. III.19 and III.20). Not surprisingly, both protein fusions presented the same fluorescence patterns as in the 48 hours observations, hence further supporting the reliability and the conclusion that both fusion proteins localize, in a stable fashion, to the tonoplast. Moreover, in some images it was clear that the protoplasts had suffered plasma membrane disruption with release of what was clearly identified as intact vacuoles, carrying chloroplasts still bound to the outside of the tonoplast. It was possible to find the tonoplast of such isolated vacuoles presenting a conspicuous green fluorescence, as can be seen in the lower panel of Fig. III.19.

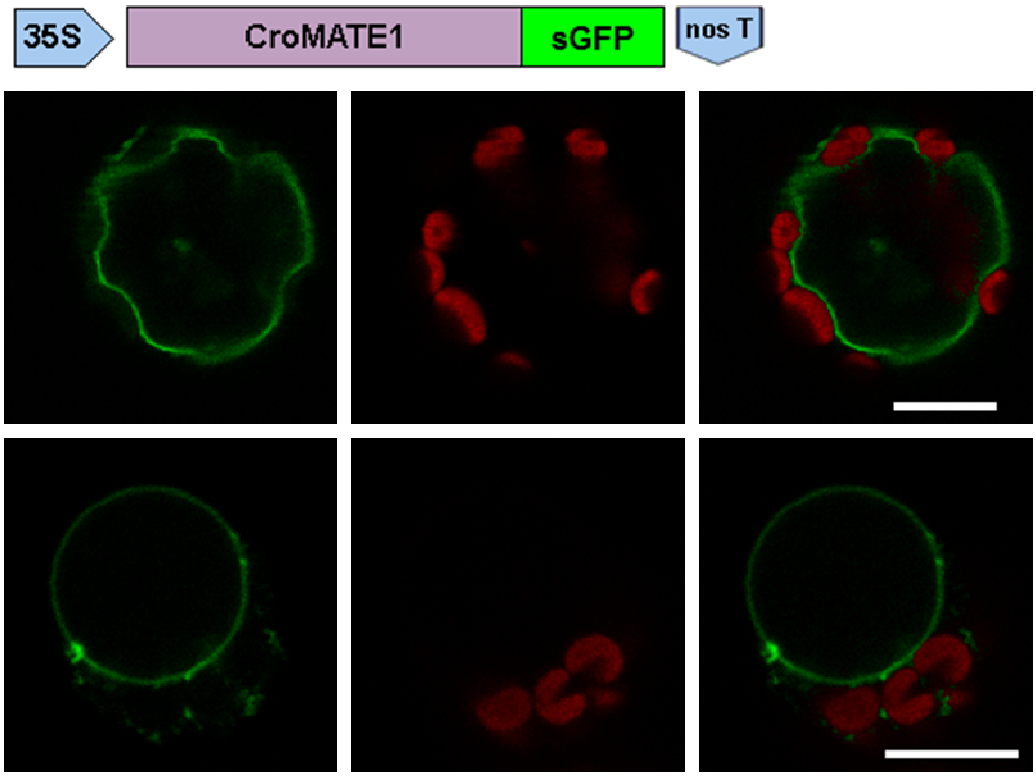


Fig. III.19 - Transient expression of CroMATE1-sGFP in *C. roseus* mesophyll protoplasts observed under the confocal microscope 72 h after transformation. A schematic representation of the construct used is shown on top of the set of images. Left – GFP channel; middle – red channel showing chloroplast autofluorescence; right – merged images. Bars = 10 µm.

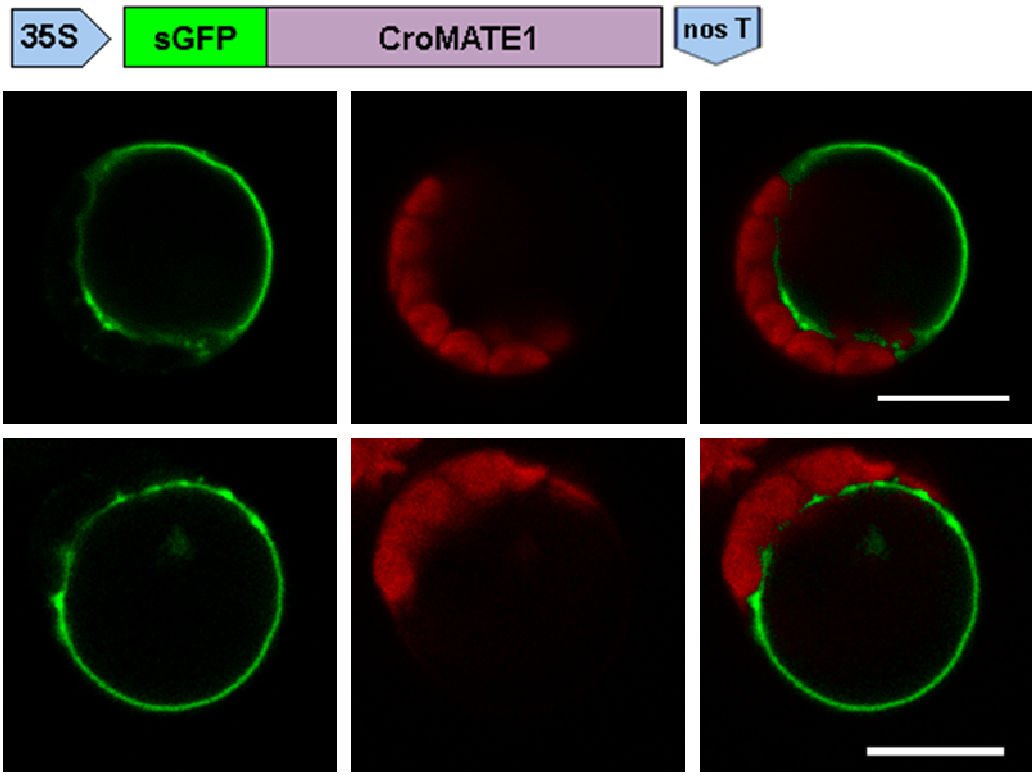


Fig. III.20 - Transient expression of sGFP-CroMATE1 in *C. roseus* mesophyll protoplasts observed under the confocal microscope 72 h after transformation. A schematic representation of the construct used is shown on top of the set of images. Left – GFP channel; middle – red channel showing chloroplast autofluorescence; right – merged images. Bars = 10 µm.

III.5. Evolutionary relationship of CroMATE1 and characterized plant MATE-type transporter proteins

After acknowledging the CroMATE1 subcellular localization - the tonoplast, it was decided to take advantage of this information and follow up to a phylogenetic reconstruction analysis, using other plant MATE-type transporters characterized to date. The purpose was to identify which plant MATE-type transporters most relate to CroMATE1 and recall which function they play *in planta* and where they locate in the subcellular landscape, perhaps resulting in new evidences regarding CroMATE1 function. Being conserved as MATE-type proteins seem to be (Moriyama, Hiasa et al. 2008), but also keeping in mind that very little is still known in regard to plant MATE-type transporters, it was expected to find some type of assortment based in the function of the MATE transporters and/or their subcellular localization, in a species-independent manner. For this analysis, we included the amino acid sequence of most of the characterized plant MATE-type transporters with clear evidence about the

respective subcellular localization and function (Table III.3). Additionally, two other *C. roseus* MATE-type transporter candidates identified by our group but still uncharacterized were used, CroMATE4 (MPGR - cra_locus_4138_iso_4_len_1354_ver_3) and CroMATE5 (MPGR - cra_locus_6853_iso_2_len_2018_ver_3), to analyze the respective assortment among these characterized sequences. The bacterial *Vibrio* VpNorM (AB010463.1) was added to be used as an outgroup member.

Table III.3 - Set of plant MATE-type transporters used in the phylogenetic reconstruction analysis indicating the respective ID, accession number, subcellular localization, function, and the respective reference.

ID	Accession no.	Subcellular Localization	Function	Reference
AtTT12	NM_115765.3	Tonoplast	Vacuolar uptake of flavonoids	Marinova, Pourcel et al. (2007)
VvAM1	NM_001281108.1	Tonoplast	Vacuolar uptake of flavonoids	Gomez, Terrier et al. (2009)
VvAM3	FJ264202.1	Tonoplast	Vacuolar uptake of flavonoids	Gomez, Terrier et al. (2009)
MtMATE1	FJ858726.1	Tonoplast	Vacuolar uptake of flavonoids	Zhao and Dixon (2009)
AtDTX35	NM_118696.4	Tonoplast	Vacuolar uptake of flavonoids	Thompson, Wilkins et al. (2010)
AtDTX1	NM_126443.4	PM	Efflux of exogenous toxic compounds	Li, He et al. (2002)
EcMATE1	AB725912.1	PM	Citrate-excretion/aluminum tolerance	Sawaki, Kihara-Doi et al. (2013)
LaMATE	AY631874.1	PM	Citrate-excretion/aluminum tolerance	Uhde-Stone, Liu et al. (2005)
SbMATE	EF611342.1	PM	Citrate-excretion/aluminum tolerance	Magalhaes, Liu et al. (2007)
HvAACT1	AB302223.1	PM	Citrate-excretion/aluminum tolerance	Furukawa, Yamaji et al. (2007)
Nt-JAT1	AM991692.1	Tonoplast	Vacuolar uptake of nicotine	Morita, Shitan et al. (2009)
NtMATE1	AB286961	Tonoplast	Vacuolar uptake of nicotine	Shoji, Inai et al. (2009)
NtMATE2	AB286963	Tonoplast	Vacuolar uptake of nicotine	Shoji, Inai et al. (2009)

As can be seen in Fig. III.21, the phylogenetic reconstruction assay revealed that MATE transporters that have a reported vacuolar localization and the ones that have a reported plasma membrane localization appeared generally delimited by robust branches with significant distances between them, in a species independent manner. EcMATE1, LaMATE, SbMATE and HvAACT1, the plasma membrane transporters involved in citrate extrusion for aluminum tolerance are all contained in a distinct branch (Fig. III.21, C). The only plasma membrane transporter outside this branch is AtDTX1, which actually plays a different function - extrusion of xenobiotics (berberine, cadmium and antibiotics). Similarly, the vacuolar MATE transporters also compose a relatively short, robust branch (Fig. III.21, B). Within this branch, there are vacuolar

transporters involved in flavonoid uptake, and two of the vacuolar transporters involved in nicotine uptake, NtMATE1 and NtMATE2, seemingly very similar in sequence with each other. Surprisingly, Nt-JAT1, which has been reported to be a vacuolar transporter of nicotine (Morita, Shitan et al. 2009), appeared outside the vacuolar MATE branch (Fig. III.21, A). On the contrary, and quite meaningfully, CroMATE1 is the transporter from this branch that appears most related to NtMATE1 and NtMATE2. The alignment between CroMATE1 and NtMATE1 peptides reveals the high similarity they share in Fig. III.22. Moreover, the second most-related MATE with NtMATE1/2 is CroMATE5, very close to CroMATE1, which therefore is also a candidate to TIA transport that should be characterized.

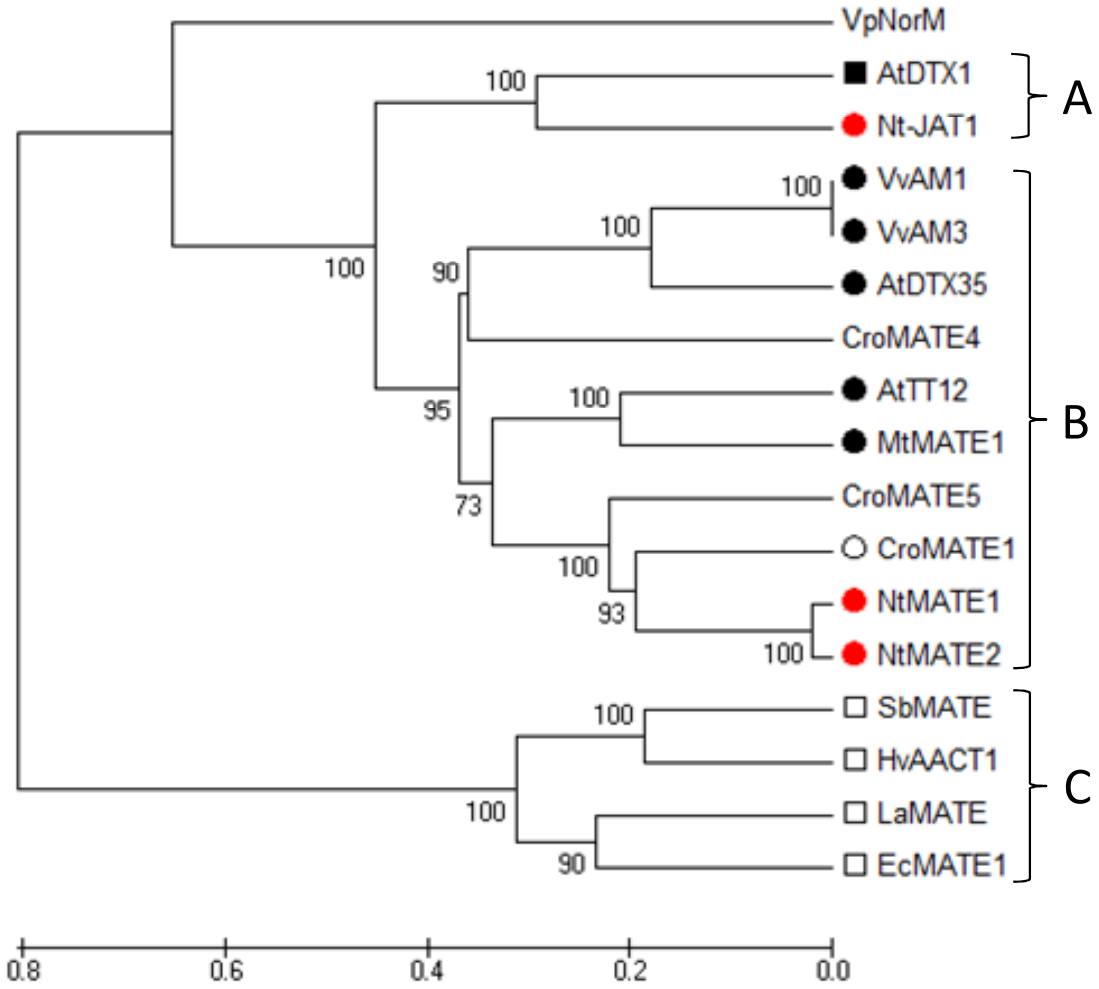


Fig. III.21 - Neighbor-Joining tree depicting the evolutionary relationships of CroMATE1 and characterized plant MATE-type transporters. The optimal tree with the sum of branch length = 5.59020144 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Filled black squares – plasma membrane transporters involved in the extrusion of xenobiotics; empty black squares - plasma membrane transporters involved in the citrate extrusion/aluminum tolerance; filled black circles - vacuolar transporters involved in the uptake of flavonoids; filled red circles - vacuolar transporters involved in the uptake of alkaloids; empty black circle - vacuolar transporter with unknown function.

```
CLUSTAL O(1.2.0) multiple sequence alignment


CroMATE1        MGSKQNYEINQPLLLSNNGGGSSAITAAESPKAAKTVEERQYELSVELERVLSDTSVPLV
NtMATE1         MGKSMKSEVEQPLLIAAHGGS----------------------SELEEVLSDTQLPYF
                **.. : *::****:: .**.                        ***.*****.:* .

CroMATE1        PRLTAATWIEFKLLFRLAAPAAAVYLINYVMSMFTQIFSGHLGNLELAAASLGNNGIQTF
NtMATE1         RRLRYASWIEFQLLYRLAAPSVAVYMINNAMSMSTRIFSGQLGNLQLAAASLGNQGIQLF
                 **  *:****:**:*****:.***:** .*** *:****:****:********:*** *

CroMATE1        AYGIMLGMGSAVETLCGQAYGAQKLDMLGIYLQRSTILLTLTGIVIMFVYIFSKPILLLL
NtMATE1         AYGLMLGMGSAVETLCGQAYGAHRYEMLGVYLQRATVVLSVTGIPLTVVYLFSKNILLAL
                ***:******************:: :***:****:*::*::*** : .**:*** *** *

CroMATE1        GQSEAIAAAAALFTYGLIPQIFAYAANFPIQKFLQAQSIVAPSTYISAGAIVFHVLFSWL
NtMATE1         GESKLVASAAAVFVYGLIPQIFAYAVNFPIQKFLQAQSIVAPSAFISLGTLFVHILLSWV
                *:*: :*:***:*.**********.*****************::** *::..*:*:**:

CroMATE1        AIYKVGLGLFGASLVLSLSWWVVVVGQFIYILYSDRTKDTWRGFSVEAFHGLWSFFKLSA
NtMATE1         VVYKIGLGLLGASLVLSFSWWIIVVAQFIYIIKSERCKATWAGFRWEAFSGLCQFVKLSA
                .:**:****:*******:***::**.*****: *:* * ** **  *** ** .*.****

CroMATE1        ASAVMLCLEAWYFQILVLLAGMLPDPKIALDSLSICITILGWVFMIAVGFNAAASVRVGN
NtMATE1         GSAVMLCLETWYMQILVLLSGLLKNPEIALASISVCLAVNGLMFMVAVGFNAAASVRVSN
                .********:**:******:*:* :*:*** *:*:*::: * :**:************.*

CroMATE1        ELGAGHPRAAAFSVVIVTTMSFIIAVIISLVVLALRYKISYIFTEGEVVSNAVADMCPLL
NtMATE1         ELGAAHSKSAAFSVFMVTFISFLIAVVEAIIVLSLRNVISYAFTEGEIVAKEVSELCPFL
                ****.* ::*****.:** :**:***: :::**:**  *** *****:*:: *:::**:*

CroMATE1        AITLVLNGIQPVLSGVAVGCGWQAFVAYVNVGCYYIVGIPTGALLGFYFKLGAKGIWSGM
NtMATE1         AVTLILNGIQPVLSGVAVGCGWQAFVAYVNVGCYYGVGIPLGCLLGFKFDLGAKGIWTGM
                *:**:*****************************  **** *.**** *.*******:**

CroMATE1        IGGTLMQTIILIWFTYRTDWKKEVDIAQSRLDTWEDKPKSLSNE
NtMATE1         IGGTVMQTVILLWVTFRTDWNKKVECAKKRLDKWENLKGPLNKE
                ****:***:**:*.*:****:*:*: *:.***.**:    *.:*
```

Fig. III.22 - Alignment between CroMATE1 and NtMATE1 amino acid sequences.

IV. Discussion

The present work has successfully established and determined the main molecular features of the *CroMATE1* gene and encoded protein, has generated the cloned gene of this transporter, performed its subcellular localization and has determined its phylogenetic relationship with well characterized plant MATE transporters. As initially hypothesized, all data obtained concurs to indicate a role for CroMATE1 in the vacuolar transport of TIAs in *C. roseus*. A discussion of the most important aspects and results of this work is provided below.

IV.1. Characterization of CroMATE1 by *in silico* tools

In recent years, bioinformatic tools have demonstrated to have a tremendous potential in successfully predicting experimental results *a priori*. CroMATE1 did not seem to be an exception, as most of the *in silico* reports obtained in this study met the expected results, and moreover were consistent with each other, thus strengthening the reliability of such approach. The NCBI-pBLAST® has identified CroMATE1 to be mostly related with several different MATE-type transporters, namely TESTA TRANSPARENT-like transporters, in conformity with the functional annotation attributed to CroMATE1 in the MPGR database. Two different tools were used to assess the number and position of putative transmembrane domains present in CroMATE1, HMMTOP and PredictProtein. Both provided the very same outcome, predicting CroMATE1 to possess 12 TMDs, a value that meets the MATE-type transporters typical topology. The WoLF PSORT subcellular localization prediction was confirmed by the experimental subcellular localization of the sGFP-tagged CroMATE1 fusion proteins, both indicating localization in the tonoplast. By combining several different criteria such as the amino acid content, ratio of amino acid charges and BLAST analysis, the WoLF PSORT tool might become useful in upcoming studies involving the subcellular fate determination of other membrane proteins. However, potential deviations to the empirical subcellular fate cannot be ruled out, so one must be cautious when using this kind of predictive bioinformatic tools.

The hierarchical clustering analysis (HCL) performed in MeV 4.9.0 with a selected set of *C. roseus* gene expression profiles has revealed interesting clues regarding to the potential involvement of CroMATE1 in the TIA pathway, specifically in the inter- and intracellular trafficking of its precursors. The *CroMATE1* expression

levels seem to increase tremendously with the application of methyl-jasmonate (MJ), as do several characterized TIA biosynthetic enzymes (*e.g. STR*, *SLS*, *D4H*, *G10H*, *DAT*) included in this study. In fact, MeJA induction is considered a hallmark of the TIA pathway in *C. roseus* (Menke, Champion et al. 1999, van der Fits and Memelink 2000, Memelink and Gantet 2007), and a TIA transporter is thus expected to present this regulatory behavior. Other candidate transcription factors, ABC-type transporters and MATE-type transporters identified in the BioNatPro group seem to also follow this behavior and may be interesting for future studies. The clustering together of all those genes by HCL suggests that they might be under regulation of some sort of regulatory mechanism sensitive to MeJA, namely JA-modulated transcription factors that are known to trigger the up-regulation of genes involved in the secondary metabolism (De Geyter, Gholami et al. 2012). Moreover, recalling that *NtMATE1*, *NtMATE2* (Shoji, Inai et al. 2009) and *Nt-JAT1* (Morita, Shitan et al. 2009) expression is likewise induced by MeJA, this might indicate that the vacuolar MATE-type transporter CroMATE1 is also required for alkaloid transport into the vacuoles, potentially acting in the *C. roseus* TIA biosynthetic pathway. This is supported by the recent findings that unveil the presence of a H^+ antiporter in the vacuole of *C. roseus* leaf cells that specifically translocates TIA products into the vacuole (Carqueijeiro, Noronha et al. 2013). This cluster of interest was detected using several different combinations of metrics and clustering methods, hence shown to be highly robust and statistically significant.

IV.2. Molecular cloning and subcellular localization of CroMATE1

Despite the difficulties inherent to PCR amplification of the *CroMATE1* CDS from protoplasts total RNA, the optimized PCR settings and reaction setup resulted in the amplification of a fair amount of product to be used in the subsequent digestion, ligation, and cloning procedures. The identity and integrity of *CroMATE1-GFP* fusion constructs was successfully confirmed by sequencing three positive clones in each case, detected by restriction analysis. All clones were error-free, as confirmed by sequencing overlaps and chromatograms, with an exception for a nucleotide C at position 1,428 instead of a T base-pair, as presented in the original *CroMATE1* CDS available at MPGR database. However, it was shown that the resulting proteins from the predicted and isolated sequences were exactly the same (Fig. III.13). This makes perfect sense in an evolutionary context, given that in most cases, conserved genes are more prone to synonymous substitutions that do not affect the protein sequence,

thereby conserving its function. Therefore, most likely this mismatch consists in a single nucleotide polymorphism (SNP), naturally occurring in *C. roseus*, resulting in no changes of the predicted amino acid sequence.

In regard to the subcellular localization, both N-terminal and C-terminal GFP fusions of CroMATE1 seem to consistently mark the tonoplast, as expected, since the primary detection of the CroMATE1 protein was in a proteomic characterization of *C. roseus* vacuole and tonoplast fractions (Carqueijeiro 2013). In fact, all transformed protoplasts exhibited, remarkably and unequivocally, a green fluorescence at the tonoplast, strongly supported by the labeled internal invaginations that systematically circumvented the chloroplasts from the inside, the exact same structural conformation of the vacuole-delimiting membrane, the tonoplast. Furthermore, the co-expression assay with CFP-tagged AtAGP58 clearly discriminated the plasma membrane from the membranous structure marked by the CroMATE1 fusion proteins, which can only be the tonoplast. Another interesting observation was that there are no significant differences found between the CroMATE1-sGFP and sGFP-CroMATE1 expression patterns, except that the sGFP-CroMATE1 transformed protoplasts often contained small, spherical, densely fluorescent bodies, most likely to be fusion protein aggregates. It must be noted that the sGFP-CroMATE1 fusion proteins did not include an SGSGS linker between the sGFP and CroMATE1 respective protein sequences, as was introduced in the CroMATE1-sGFP fusion proteins. The absence of linker may have affected in some way the stability and conformation of the proteins, possibly resulting in some tendency to form aggregates, as sometimes observed for GFP. Nevertheless, results kept consistent regardless the use of the SGSGS linker in CroMATE1-sGFP fusion proteins. These two functional fusion proteins helped to elucidate that the CroMATE1 transporter protein does not require free N- and C-terminus to be sorted into the vacuole, ruling out protein cleavage and other related sorting signaling mechanisms.

IV.3. Evolutionary relationships of CroMATE1

The generated Neighbor-Joining tree clearly shows a function-specific assortment of the included characterized plant MATE-type transporter proteins (Fig. III.21). This distance matrix method computes the sequence divergence by relying on the pairwise, amino acid differences of all protein sequences involved, once properly aligned in function of the protein-coding information. The Jukes-Cantor correction model was

designed to account for unobservable superimposed nucleotide substitutions that increment the sequence divergence to approximate as much as possible towards a realistic inference (Cantor and Jukes 1966). EcMATE1, LaMATE, SbMATE and HvAACT1, the plasma membrane transporters involved in citrate extrusion for aluminum tolerance are all contained in a distinct branch. This is remarkable, attending to the fact that this branch includes species phylogenetically very distant, and indicates that this gene/function has evolved before the divergence of dicotyledons and monocotyledons. The only plasma membrane transporter outside this branch is AtDTX1, which actually plays a different function, that is, the extrusion of xenobiotics (berberine, cadmium and antibiotics) (Li, He et al. 2002).

In regard to MATE transporters reported as vacuolar, all except one are contained within a branch, with most of them being responsible for the flavonoid vacuolar uptake, whereas NtMATE1 and NtMATE2 have been implicated in the vacuolar uptake of the alkaloid nicotine (Table III.3). The other tobacco MATE transporter also implicated in the vacuolar uptake of nicotine, Nt-JAT1, is not included in the branch of the vacuolar MATEs. Surprisingly, it is closer to AtDTX1 rather than to any other MATE-type transporter used in this analysis. In fact, the characterization of Nt-JAT1 published by Morita, Shitan et al. (2009), showed that Nt-JAT1 seemed not only to be localized in the tonoplast of tobacco leaf cells but also in the plasma membrane of tobacco root cells. Moreover, the study showed that Nt-JAT1 has a wide range of substrates besides nicotine, including anabasine, another endogenous tobacco alkaloid, hyoscyamine (a tropane alkaloid) and berberine (an isoquinoline alkaloid) derived from other plant species (Morita, Shitan et al. 2009). Curiously, these transport substrate profile matches the lack of substrate specificity of AtDTX1 (Li, He et al. 2002), and raises the question whether Nt-JAT1 may also function as a xenobiotic efflux transporter.

Still by observing the phylogram, it is clear that CroMATE1 is most closely related to NtMATE1/2, involved in vacuolar alkaloid uptake, than to any other protein in the study. This relatedness becomes even more evident when the NtMATE1 and CroMATE1 respective amino acid sequences were aligned (Fig. III.22). Such observation points to a potentially identical role for CroMATE1, but only further experimental work may clarify if its function is the transport of TIAs into the *C. roseus* vacuoles. Another *C. roseus* MATE-type candidate, CroMATE5, is apparently very close to CroMATE1. This position best supports a potential vacuolar localization rather than a plasma membrane one, and given the proximity with NtMATE1/2, its function could be similar as well. CroMATE4, another *C. roseus* candidate MATE transporter, is also among the vacuolar type transporters, although not as close to NtMATE1/2 and

CroMATE1. Given all this clues, and although considering the results of the phylogenetic reconstruction with the appropriate caution, we suggest that CroMATE1 is most likely involved in the vacuolar alkaloid uptake rather than flavonoid uptake. Unfortunately, very few plant MATE-type transporters have been deeply characterized to date and this lack of information weakens the scientific support of this phylogenetic inference. In the near future, as more information will become available not only from *C. roseus* but other species as well, this type of approach will certainly demonstrate its usefulness.

V. Conclusions and future perspectives

In this work, a candidate MATE-type transporter of *C. roseus* was successfully isolated, cloned, characterized and prominently shown to localize in the tonoplast. Predictive *in silico* tools were consistent in predicting several different MATE-type transporter features in CroMATE1, by confirming its relatedness among TESTA TRANSPARENT-like transporter proteins and assigning the presence of 12 TMDs. The predictive subcellular localization determination revealed the maximum likelihood to be attributed to the respective vacuolar localization, which was later confirmed by GFP-tagged CroMATE1 subcellular localization, under fluorescence microscopy.

HCL analysis performed on 23 different samples resulted in the identification of a correlated, narrow cluster of TIA biosynthetic genes and *CroMATE1* showing induction of expression by MeJA, a hallmark of the TIA pathway, also observed for most of the alkaloid transporting vacuolar MATE-type transporters described so far.

Additionally, phylogenetic reconstruction revealed a unique assortment pattern that seemed to group MATE proteins according to their function/subcellular localization, unveiling the significant high similarity of CroMATE1 with vacuolar, alkaloid transporting MATE-type transporters.

Taken together, all lines of results in this study reinforce the potential functional role of CroMATE1 *in planta* as a vacuolar MATE-type transporter capable of transporting TIAs into the vacuole and hence influencing the subsequent vacuolar-native transformation steps required for the TIA pathway end products, VCR and VLB. The next step is obviously to determine the substrate-specificity of recombinant CroMATE1, its activity rate and potential inhibitors and to investigate the expression profile of *CroMATE1* in different *C. roseus* organs, tissue/cell cultures, conditions and developmental stages. It would also be interesting to perform the silencing of

CroMATE1 using virus induced gene silencing (VIGS). If it is confirmed that CroMATE1 is indeed the vacuolar transporter of TIAs in *C. roseus*, it will constitute a powerful component to enhance TIA-production in a viable way.

VI. Bibliography

Bowen, R. A. (1998, July 11, 1998). "Molecular Toolkit: Translate a DNA sequence." Retrieved May 10, 2013, from http://www.vivo.colostate.edu/molkit/index.html.

Brown, M. H., I. T. Paulsen and R. A. Skurray (1999). "The multidrug efflux protein NorM is a prototype of a new family of transporters." Molecular Microbiology 31(1): 394-395.

Brown, S., J. P. Renaudin, C. Prevot and J. Guern (1984). "FLOW-CYTOMETRY AND SORTING OF PLANT-PROTOPLASTS - TECHNICAL PROBLEMS AND PHYSIOLOGICAL RESULTS FROM A STUDY OF PH AND ALKALOIDS IN CATHARANTHUS-ROSEUS." Physiologie Vegetale 22(5): 541-554.

Cantor, C. R. and T. H. Jukes (1966). "REPETITION OF HOMOLOGOUS SEQUENCES IN POLYPEPTIDE CHAINS OF CERTAIN CYTOCHROMES AND GLOBINS." Proceedings of the National Academy of Sciences of the United States of America 56(1): 177-&.

Carqueijeiro, I. (2013). Unravelling the metabolism and transmembrane transport of the highly valuable medicinal alkaloids from *Catharanthus roseus* (L) G.Don. PhD thesis, Universidade do Porto.

Carqueijeiro, I., H. Noronha, P. Duarte, H. Geros and M. Sottomayor (2013). "Vacuolar Transport of the Medicinal Alkaloids from Catharanthus roseus Is Mediated by a Proton-Driven Antiport." Plant Physiology 162(3): 1486-1496.

Corpet, F. (1988). "MULTIPLE SEQUENCE ALIGNMENT WITH HIERARCHICAL-CLUSTERING." Nucleic Acids Research 16(22): 10881-10890.

De Geyter, N., A. Gholami, S. Goormachtig and A. Goossens (2012). "Transcriptional machineries in jasmonate-elicited plant secondary metabolism." Trends in Plant Science 17(6): 349-359.

Debeaujon, I., A. J. M. Peeters, K. M. Leon-Kloosterziel and M. Koornneef (2001). "The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium." Plant Cell 13(4): 853-871.

Diener, A. C., R. A. Gaxiola and G. R. Fink (2001). "Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins." Plant Cell 13(7): 1625-1637.

Duarte, P., D. Ribeiro, G. Henriques, F. Hilliou, A. S. Rocha, F. Lima, I. Amorim and M. Sottomayor (2011). Cloning and Characterization of a Candidate Gene from the Medicinal Plant *Catharanthus roseus* Through Transient Expression in Mesophyll Protoplasts. Molecular Cloning - Selected Applications in Medicine and Biology. G. Brown, InTech.

Facchini, P. J. (2001). "Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications." Annual Review of Plant Physiology and Plant Molecular Biology 52: 29-66.

Frank, S., M. Keck, M. Sagasser, K. Niehaus, B. Weisshaar and R. Stracke (2011). "Two differentially expressed MATE factor genes from apple complement the Arabidopsis transparent testa12 mutant." Plant Biology 13(1): 42-50.

Furukawa, J., N. Yamaji, H. Wang, N. Mitani, Y. Murata, K. Sato, M. Katsuhara, K. Takeda and J. F. Ma (2007). "An aluminum-activated citrate transporter in barley." Plant and Cell Physiology 48(8): 1081-1091.

Goddijn, O. J. M., R. J. Dekam, A. Zanetti, R. A. Schilperoort and J. H. C. Hoge (1992). "AUXIN RAPIDLY DOWN-REGULATES TRANSCRIPTION OF THE TRYPTOPHAN DECARBOXYLASE GENE FROM CATHARANTHUS-ROSEUS." Plant Molecular Biology 18(6): 1113-1120.

Gomez, C., N. Terrier, L. Torregrosa, S. Vialet, A. Fournier-Level, C. Verries, J. M. Souquet, J. P. Mazauric, M. Klein, V. Cheynier and A. Ageorges (2009). "Grapevine MATE-Type Proteins Act as Vacuolar H+-Dependent Acylated Anthocyanin Transporters." Plant Physiology 150(1): 402-415.

Gongora-Castillo, E., K. L. Childs, G. Fedewa, J. P. Hamilton, D. K. Liscombe, M. Magallanes-Lundback, K. K. Mandadi, E. Nims, W. Runguphan, B. Vaillancourt, M. Varbanova-Herde, D. DellaPenna, T. D. McKnight, S. O'Connor and C. R. Buell (2012). "Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species." Plos One 7(12).

Green, L. S. and E. E. Rogers (2004). "FRD3 controls iron localization in Arabidopsis." Plant Physiology 136(1): 2523-2531.

Guirimand, G., A. Guihur, P. Poutrain, F. Héricourt, S. Mahroug, B. St-Pierre, V. Burlat and V. Courdavault (2011). "Spatial organization of the vindoline biosynthetic pathway in Catharanthus roseus." Journal of Plant Physiology Volume 168(6): 549-557

Hartmann, T. (2007). "From waste products to ecochemicals: Fifty years research of plant secondary metabolism." Phytochemistry 68(22-24): 2831-2846.

Hashimoto, T. and Y. Yamada (2003). "New genes in alkaloid metabolism and transport." Current Opinion in Biotechnology 14(2): 163-168.

Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier and K. Nakai (2007). "WoLF PSORT: protein localization predictor." Nucleic Acids Research 35: W585-W587.

Kuijt, S. J. H., G. E. M. Lamers, S. Rueb, E. Scarpella, P. B. F. Ouwerkerk, H. P. Spaink and A. H. Meijer (2004). "Different subcellular localization and trafficking properties of KNOX class 1 homeodomain proteins from rice." Plant Molecular Biology 55(6): 781-796.

Li, L. G., Z. Y. He, G. K. Pandey, T. Tsuchiya and S. Luan (2002). "Functional cloning and characterization of a plant efflux carrier for multidrug and heavy metal detoxification." Journal of Biological Chemistry 277(7): 5360-5368.

Loyola-Vargas, V. M., R. M. Galaz-Ávalos and R. Kú-Cauich (2007). "*Catharanthus* biosynthetic enzymes: the road ahead." Phytochemistry Reviews 6: 307-339.

Magalhaes, J. V., J. Liu, C. T. Guimaraes, U. G. P. Lana, V. M. C. Alves, Y. H. Wang, R. E. Schaffert, O. A. Hoekenga, M. A. Pineros, J. E. Shaff, P. E. Klein, N. P. Carneiro, C. M. Coelho, H. N. Trick and L. V. Kochian (2007). "A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum." Nature Genetics 39(9): 1156-1161.

Mahroug, S., V. Burlat and B. St-Pierre (2007). "Cellular and sub-cellular organisation of the monoterpenoid indole alkaloid pathway in Catharanthus roseus." Phytochemistry Reviews 6(2-3): 363-381.

Marinova, K., L. Pourcel, B. Weder, M. Schwarz, D. Barron, J. M. Routaboul, I. Debeaujon and M. Klein (2007). "The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+-antiporter active in proanthocyanidin-accumulating cells of the seed coat." Plant Cell 19(6): 2023-2038.

Memelink, J. and P. Gantet (2007). "Transcription factors involved in terpenoid indole alkaloid biosynthesis in Catharanthus roseus." Phytochemistry Reviews 6(2-3): 353-362.

Menke, F. L. H., A. Champion, J. W. Kijne and J. Memelink (1999). "A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2." Embo Journal 18(16): 4455-4463.

Morita, M., N. Shitan, K. Sawada, M. C. E. Van Montagu, D. Inze, H. Rischer, A. Goossens, K. M. Oksman-Caldentey, Y. Moriyama and K. Yazaki (2009). "Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in Nicotiana tabacum." Proceedings of the National Academy of Sciences 106(7): 2447-2452.

Morita, Y., K. Kodama, S. Shiota, T. Mine, A. Kataoka, T. Mizushima and T. Tsuchiya (1998). "NorM, a putative multidrug efflux protein, of Vibrio parahaemolyticus and its homolog in Escherichia coli." Antimicrobial Agents and Chemotherapy 42(7): 1778-1782.

Moriyama, Y., M. Hiasa, T. Matsumoto and H. Omote (2008). "Multidrug and toxic compound extrusion (MATE)-type proteins as anchor transporters for the excretion of metabolic waste products and xenobiotics." Xenobiotica 38(7-8): 1107-1118.

Murata, J., J. Roepke, H. Gordon and V. De Luca (2008). "The leaf epidermome of Catharanthus roseus reveals its biochemical specialization." Plant Cell 20(3): 524-542.

Niwa, Y., T. Hirano, K. Yoshimoto, M. Shimizu and H. Kobayashi (1999). "Non-invasive quantitative detection and applications of non-toxic, S65T-type green fluorescent protein in living plants." Plant Journal 18(4): 455-463.

Nour-Eldin, H. H. and B. A. Halkier (2013). "The emerging field of transport engineering of plant specialized metabolites." Current Opinion in Biotechnology 24(2): 263-270.

Omote, H., M. Hiasa, T. Matsumoto, M. Otsuka and Y. Moriyama (2006). "The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations." Trends Pharmacol Sci 27(11): 587-593.

Pang, Y. Z., X. F. Cheng, D. V. Huhman, J. Y. Ma, G. J. Peel, K. Yonekura-Sakakibara, K. Saito, G. A. Shen, L. W. Sumner, Y. H. Tang, J. Q. Wen, J. F. Yun and R. A. Dixon (2013). "Medicago glucosyltransferase UGT72L1: potential roles in proanthocyanidin biosynthesis." Planta 238(1): 139-154.

Pasquali, G., O. J. M. Goddijn, A. Dewaal, R. Verpoorte, R. A. Schilperoort, J. H. C. Hoge and J. Memelink (1992). "COORDINATED REGULATION OF 2 INDOLE ALKALOID BIOSYNTHETIC GENES FROM CATHARANTHUS-ROSEUS BY AUXIN AND ELICITORS." Plant Molecular Biology 18(6): 1121-1131.

Roewer, I. A., N. Cloutier, C. L. Nessler and V. Deluca (1992). "TRANSIENT INDUCTION OF TRYPTOPHAN DECARBOXYLASE (TDC) AND STRICTOSIDINE SYNTHASE (SS) GENES IN CELL-SUSPENSION CULTURES OF CATHARANTHUS-ROSEUS." Plant Cell Reports 11(2): 86-89.

Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush and J. Quackenbush (2003). "TM4: A free, open-source system for microarray data management and analysis." Biotechniques 34(2): 374-+.

Saitou, N. and M. Nei (1987). "THE NEIGHBOR-JOINING METHOD - A NEW METHOD FOR RECONSTRUCTING PHYLOGENETIC TREES." Molecular Biology and Evolution 4(4): 406-425.

Salim, V., F. Yu, J. Altarejos and V. Deluca (2013). "Virus induced gene silencing identifies Catharanthus roseus 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis." Plant Journal.

Sawaki, Y., T. Kihara-Doi, Y. Kobayashi, N. Nishikubo, T. Kawazu, Y. Kobayashi, H. Koyama and S. Sato (2013). "Characterization of Al-responsive citrate excretion and citrate-transporting MATEs in Eucalyptus camaldulensis." Planta 237(4): 979-989.

Shitan, N., I. Bazin, K. Dan, K. Obata, K. Kigawa, K. Ueda, F. Sato, C. Forestier and K. Yazaki (2003). "Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in Coptis japonica." Proceedings of the National Academy of Sciences of the United States of America 100(2): 751-756.

Shoji, T., K. Inai, Y. Yazaki, Y. Sato, H. Takase, N. Shitan, K. Yazaki, Y. Goto, K. Toyooka, K. Matsuoka and T. Hashimoto (2009). "Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots." Plant Physiol 149(2): 708-718.

Sirikantaramas, S., M. Yamazaki and K. Saito (2008). "Mechanisms of resistance to self-produced toxic secondary metabolites in plants." Phytochemistry Reviews 7(3): 467-477.

Sottomayor, M. and A. Ros-Barceló (2006). The Vinca alkaloids: From biosynthesis and accumulation in plant cells, to uptake, activity and metabolism in animal

cells. Studies in Natural Products Chemistry. R. Atta-ur, Elsevier. Volume 33, Part M: 813-857.

St-Pierre, B., F. A. Vazquez-Flota and V. De Luca (1999). "Multicellular compartmentation of catharanthus roseus alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate." Plant Cell 11(5): 887-900.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011). "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." Molecular Biology and Evolution 28(10): 2731-2739.

Thompson, E. P., C. Wilkins, V. Demidchik, J. M. Davies and B. J. Glover (2010). "An Arabidopsis flavonoid transporter is required for anther dehiscence and pollen development." Journal of Experimental Botany 61(2): 439-451.

Uhde-Stone, C., J. Q. Liu, K. E. Zinn, D. L. Allan and C. P. Vance (2005). "Transgenic proteoid roots of white lupin: a vehicle for characterizing and silencing root genes involved in adaptation to P stress." Plant Journal 44(5): 840-853.

van der Fits, L. and J. Memelink (2000). "ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism." Science 289(5477): 295-297.

van der Heijden, R., D. I. Jacobs, W. Snoeijer, D. Hallard and R. Verpoorte (2004). "The *Catharanthus* Alkaloids: Pharmacognosy and Biotechnology." Current Medicinal Chemistry 11: 607-628.

Verma, P., A. K. Mathur, A. Srivastava and A. Mathur (2012). "Emerging trends in research on spatial and temporal organization of terpenoid indole alkaloid pathway in Catharanthus roseus: a literature update." Protoplasma 249(2): 255-268.

Verpoorte, R., B. Lata and A. Sadowska, Eds. (2007). Biology and Biochemistry of *Catharanthus roseus* (L.) G. Don Phytochemistry Reviews, Springer Verlag.

Wang, J. P., H. Raman, M. X. Zhou, P. R. Ryan, E. Delhaize, D. M. Hebb, N. Coombes and N. Mendham (2007). "High-resolution mapping of the Alp locus and identification of a candidate gene HvMATE controlling aluminium tolerance in barley (Hordeum vulgare L.)." Theoretical and Applied Genetics 115(2): 265-276.

Yazaki, K. (2005). "Transporters of secondary metabolites." Current Opinion in Plant Biology 8(3): 301-307.

Yazaki, K., A. Sugiyama, M. Morita and N. Shitan (2007). "Secondary transport as an efficient membrane transport mechanism for plant secondary metabolites." Phytochemistry Reviews 7(3): 513-524.

Yoder, L. and E. Mahlberg (1976). "Reactions of Alkaloid and Histochemical Indicators in Laticifers and Specialized Parenchyma Cells of Catharanthus roseus (Apocynaceae)." Am J Bot 63(9): 12.

Yu, C. Y. (2013). "Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species." Journal of Applied Genetics 54(2): 135-145.

Zhao, J. and R. A. Dixon (2009). "MATE Transporters Facilitate Vacuolar Uptake of Epicatechin 3 '-O-Glucoside for Proanthocyanidin Biosynthesis in Medicago truncatula and Arabidopsis." Plant Cell 21(8): 2323-2340.

Ziegler, J. and P. J. Facchini (2008). "Alkaloid biosynthesis: metabolism and trafficking." Annu Rev Plant Biol 59: 735-769.

Appendices

Appendix1. CroMATE1 CDS-containing cDNA sequence as presented in the MPGR database.

>cra_locus_1952_iso_7_len_1704_ver_3

ATCAAGCAAAAGTCAGATGATTGAATTCGGGGATATCTGTTGAAAAACTTCCTTTTATTCATTGGACAAA
GATTTTGGCTTGTCCTCCCATGTATCCAACCTACTTTGAGCTATATCCACCTCTTTTTTCCAGTCCGTTC
GATAAGTAAACCAGATGAGAATAATAGTTTGCATCAATGTCCCACCAATCATACCGGACCAAATACCCTT
GGCTCCAAGTTTGAAGTAGAACCCAAGAAGTGCACCGGTTGGGATACCAACAATGTAATAACAGCCAACG
TTCACATATGCAACAAAAGCTTGCCATCCACATCCAACAGCAACACCGGATAAAACAGGTTGAATTCCAT
TAAGAACAAGAGTGATGGCGAGCAAGGGACACATATCGGCAACAGCATTGCTTACAACTTCACCTTCGGT
AAAGATATAGCTAATTTTGTAGCGCAAAGCAAGTACCACTAATGATATTATCACTGCTATTATGAATGAC
ATTGTTGTCACTATTACTACTGAAAATGCAGCTGCCCTTGGATGTCCTGCCCCTAGTTCATTCCCTACTC
TCACACTGGCAGCAGCATTGAATCCAACGGCTATCATGAATACCCAACCCAAGATTGTAATGCAAATGGA
GAGGGAATCCAAAGCGATTTTAGGATCAGGAAGCATTCCAGCCAAAAGAACAAGAATCTGAAAATACCAA
GCTTCCAAACAAAGCATAACAGCAGAAGCAGCAGACAACTTAAAAAAGCTCCAAAGTCCATGGAAAGCTT
CAACACTAAATCCACGCCAAGTGTCCTTAGTCCTATCACTATATAGAATATAAATAAACTGACCAACCAC
AACAACCCACCAAGACAAACTCAATACTAAAGATGCCCCGAACAATCCAAGTCCAACCTTATAAATTGCT
AACCAACTAAACAAAACATGGAAAACAATAGCTCCTGCTGATATATAAGTACTAGGTGCCACTATACTTT
GAGCTTGTAAGAACTTTTGGATTGGGAAATTGGCTGCGTAGGCAAAGATTTGTGGGATAAGACCATAAGT
AAATAAGGCAGCTGCTGCAGCTATGGCTTCTGATTGGCCAAGTAAAAGGAGGATTGGTTTTGAGAATATG
TAAACAAACATAATAACAATGCCTGTTAGGGTTAGAAGAATTGTTGATCTTTGGAGATATATTCCAAGCA
TGTCTAGTTTTTGTGCTCCATATGCTTGTCCACATAGTGTTTCCACTGCACTTCCCATTCCAAGCATGAT
ACCATAGGCGAAGGTCTGGATACCATTATTACCAAGAGAAGCAGCAGCAAGCTCAAGATTCCCAAGATGC
CCGGAAAATATTTGAGTGAACATAGACATTACATAGTTTATCAAATAAACGGCGGCTGCTGGAGCCGCCA
GCCGAAAGAGAAGCTTAAACTCAATCCAAGTGGCGGCGGTGAGTCTTGGCACTAATGGTACAGATGTATC
GGAAAGTACCCTTTCTAACTCAACACTAAGCTCATACTGCCTTTCCTCCACCGTCTTAGCCGCCTTTGGC
GACTCTGCAGCTGTAATAGCACTGCTGCCACCGCCGTTGTTACTTAGTAACAGTGGTTGGTTTATTTCAT
AGTTTTGTTTGGAACCCATCTTATTACAACCGCCGCCGTAGACGGTGGTCGTAGTGGTGGTGGTGTTAAG
AGTGTGTGTGTGTGTGTGTGTG

Appendix2. Engineered nucleotide and peptide sequences of the CroMATE1-sGFP fusion construct

N-TERMINAL fusion - CroMATE1.sGFP (2,307bp)

ATGGGTTCCAAACAAAACTATGAAATAAACCAACCACTGTTACTAAGTAACAACGGCGGTGGCAGCAGTGCTATTACAGCTGCAGAGT
CGCCAAAGGCGGCTAAGACGGTGGAGGAAAGGCAGTATGAGCTTAGTGTTGAGTTAGAAAGGGTACTTTCCGATACATCTGTACCATT
AGTGCCAAGACTCACCGCCGCCACTTGGATTGAGTTTAAGCTTCTCTTTCGGCTGGCGGCTCCAGCAGCCGCCGTTTATTTGATAAAC
TATGTAATGTCTATGTTCACTCAAATATTTTCCGGGCATCTTGGGAATCTTGAGCTTGCTGCTGCTTCTCTTGGTAATAATGGTATCC
AGACCTTCGCCTATGGTATCATGCTTGGAATGGGAAGTGCAGTGGAAACACTATGTGGACAAGCATATGGAGCACAAAAACTAGACAT
GCTTGGAATATATCTCCAAAGATCAACAATTCTTCTAACCCTAACAGGCATTGTTATTATGTTTGTTTACATATTCTCAAAACCAATC
CTCCTTTTACTTGGCCAATCAGAAGCCATAGCTGCAGCAGCTGCCTTATTTACTTATGGTCTTATCCCACAAATCTTTGCCTACGCAG
CCAATTTCCCAATCCAAAAGTTCTTACAAGCTCAAAGTATAGTGGCACCTAGTACTTATATATCAGCAGGAGCTATTGTTTTCCATGT
TTTGTTTAGTTGGTTAGCAATTTATAAGGTTGGACTTGGATTGTTCGGGGCATCTTTAGTATTGAGTTTGTCTTGGTGGGTTGTTGTG
GTTGGTCAGTTTATTTATATTCTATATAGTGATAGGACTAAGGACACTTGGCGTGGATTTAGTGTTGAAGCTTTCCATGGACTTTGGA
GCTTTTTTAAGTTGTCTGCTGCTTCTGCTGTTATGCTTTGTTTGGAAGCTTGGTATTTTCAGATTCTTGTTCTTTTGGCTGGAATGCT
TCCTGATCCTAAAATCGCTTTGGATTCCCTCTCCATTTGCATTACAATCTTGGGTTGGGTATTCATGATAGCCGTTGGATTCAATGCT
GCTGCCAGTGTGAGAGTAGGGAATGAACTAGGGGCAGGACATCCAAGGGCAGCTGCATTTTCAGTAGTAATAGTGACAACAATGTCAT
TCATAATAGCAGTGATAATATCATTAGTGGTACTTGCTTTGCGCTACAAAATTAGCTATATCTTTACCGAAGGTGAAGTTGTAAGCAA
TGCTGTTGCCGATATGTGTCCCTTGCTCGCCATCACTCTTGTTCTTAATGGAATTCAACCTGTTTTATCCGGTGTTGCTGTTGGATGT
GGATGGCAAGCTTTTGTTGCATATGTGAACGTTGGCTGTTATTACATTGTTGGTATCCCAACCGGTGCACTTCTTGGGTTCTACTTCA
AACTTGGAGCCAAGGGTATCTGGTCCGGTATGATTGGTGGGACATTGATGCAAACTATTATTCTCATCTGGTTTACTTATCGAACGGA
CTGGAAAAAAGAGGTGGATATAGCTCAAAGTAGGTTGGATACATGGGAGGACAAGCCAAAATCTTTGTCCAATGAA*TCTGGTTCAGGA*
*TCC*ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA
GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCC
GCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTA
CAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCC
AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCACGGCAT
GGACGAGCTGTACAAGTAA

PROTEIN (768aa)

MGSKQNYEINQPLLLSNNGGGSSAITAAESPKAAKTVEERQYELSVELERVLSDTSVPLVPRLTAATWIEFKLLFRLAAPAAAVYLIN
YVMSMFTQIFSGHLGNLELAAASLGNNGIQTFAYGIMLGMGSAVETLCGQAYGAQKLDMLGIYLQRSTILLTLTGIVIMFVYIFSKPI
LLLLGQSEAIAAAAALFTYGLIPQIFAYAANFPIQKFLQAQSIVAPSTYISAGAIVFHVLFSWLAIYKVGLGLFGASLVLSLSWWVVV
VGQFIYILYSDRTKDTWRGFSVEAFHGLWSFFKLSAASAVMLCLEAWYFQILVLLAGMLPDPKIALDSLSICITILGWVFMIAVGFNA
AASVRVGNELGAGHPRAAAFSVVIVTTMSFIIAVIISLVVLALRYKISYIFTEGEVVSNAVADMCPLLAITLVLNGIQPVLSGVAVGC
GWQAFVAYVNVGCYYIVGIPTGALLGFYFKLGAKGIWSGMIGGTLMQTIILIWFTYRTDWKKEVDIAQSRLDTWEDKPKSLSNESGSG
SMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFTYGVQCFSRYPDHMKQHDFFKS
AMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDG
SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

35S::CroMATE1.sGFP, or N-terminal CroMATE1 predicted fusion nucleotide (top) and peptide (bottom) sequences. Yellow - CroMATE1; Green - pTH-2 sGFP sequence; White - SGSGS linker; Red - Ligation site, partially overlapping both the SGSGS linker and the sGFP sequence.

Appendix3. Engineered nucleotide and peptide sequences of the sGFP-CroMATE1 fusion construct

C-TERMINAL fusion - sGFP.CroMATE1 (2,304bp)

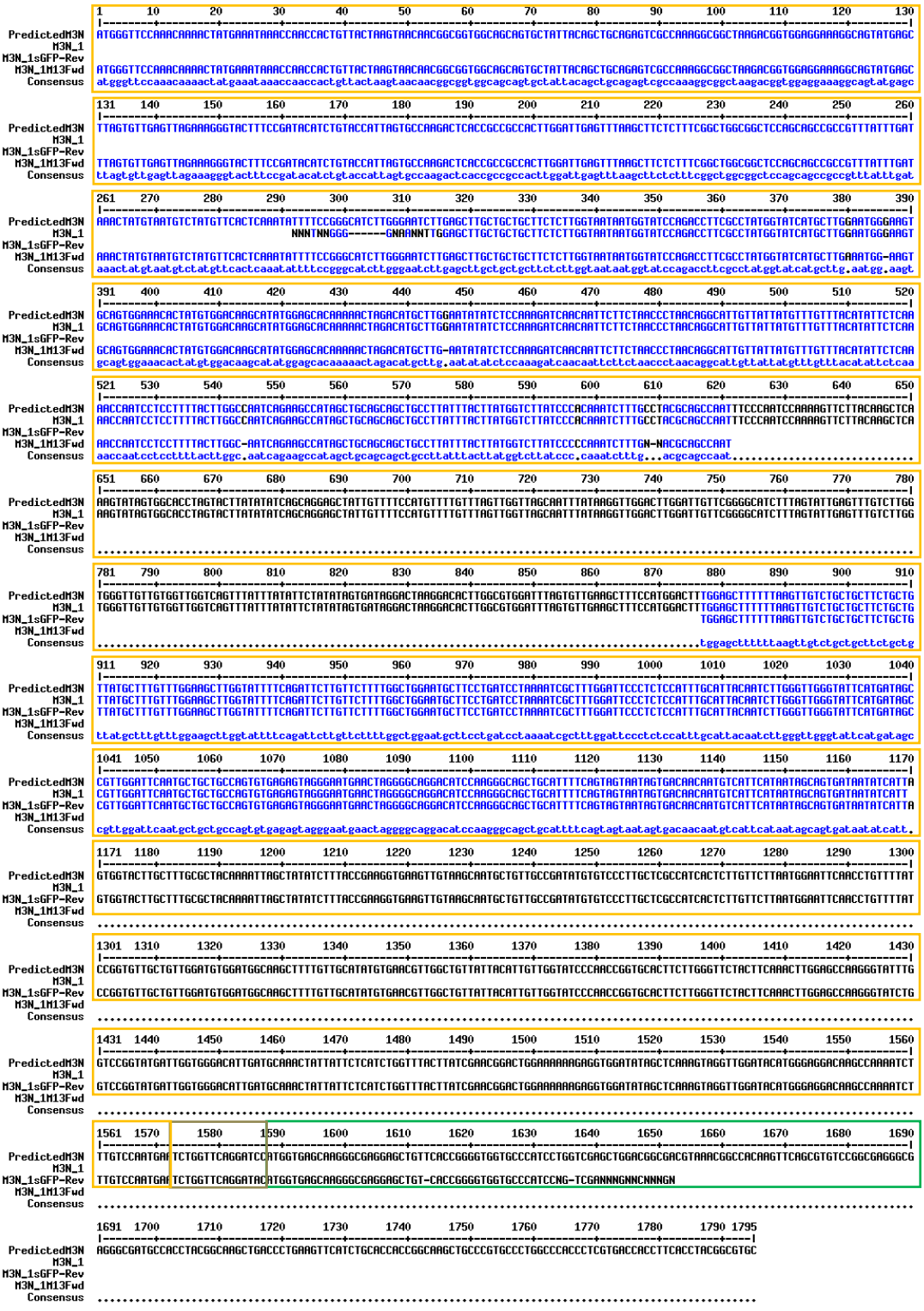
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC
CACCCTCGTGACCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCC
ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAA
CAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCACGGCATGGA
CGAGCTGTACAAGATCTCTATGGGTTCCAAACAAAACTATGAAATAAACCAACCACTGTTACTAAGTAACAACGGCGGTGGCAGCAGT
GCTATTACAGCTGCAGAGTCGCCAAAGGCGGCTAAGACGGTGGAGGAAAGGCAGTATGAGCTTAGTGTTGAGTTAGAAAGGGTACTTT
CCGATACATCTGTACCATTAGTGCCAAGACTCACCGCCGCCACTTGGATTGAGTTTAAGCTTCTCTTTCGGCTGGCGGCTCCAGCAGC
CGCCGTTTATTTGATAAACTATGTAATGTCTATGTTCACTCAAATATTTTCCGGGCATCTTGGGAATCTTGAGCTTGCTGCTGCTTCT
CTTGGTAATAATGGTATCCAGACCTTCGCCTATGGTATCATGCTTGGAATGGGAAGTGCAGTGGAAACACTATGTGGACAAGCATATG
GAGCACAAAAACTAGACATGCTTGGAATATATCTCCAAAGATCAACAATTCTTCTAACCCTAACAGGCATTGTTATTATGTTTGTTTA
CATATTCTCAAAACCAATCCTCCTTTTACTTGGCCAATCAGAAGCCATAGCTGCAGCAGCTGCCTTATTTACTTATGGTCTTATCCCA
CAAATCTTTGCCTACGCAGCCAATTTCCCAATCCAAAAGTTCTTACAAGCTCAAAGTATAGTGGCACCTAGTACTTATATATCAGCAG
GAGCTATTGTTTTCCATGTTTTGTTTAGTTGGTTAGCAATTTATAAGGTTGGACTTGGATTGTTCGGGGCATCTTTAGTATTGAGTTT
GTCTTGGTGGGTTGTTGTGGTTGGTCAGTTTATTTATATTCTATATAGTGATAGGACTAAGGACACTTGGCGTGGATTTAGTGTTGAA
GCTTTCCATGGACTTTGGAGCTTTTTTAAGTTGTCTGCTGCTTCTGCTGTTATGCTTTGTTTGGAAGCTTGGTATTTTCAGATTCTTG
TTCTTTTGGCTGGAATGCTTCCTGATCCTAAAATCGCTTTGGATTCCCTCTCCATTTGCATTACAATCTTGGGTTGGGTATTCATGAT
AGCCGTTGGATTCAATGCTGCTGCCAGTGTGAGAGTAGGGAATGAACTAGGGGCAGGACATCCAAGGGCAGCTGCATTTTCAGTAGTA
ATAGTGACAACAATGTCATTCATAATAGCAGTGATAATATCATTAGTGGTACTTGCTTTGCGCTACAAAATTAGCTATATCTTTACCG
AAGGTGAAGTTGTAAGCAATGCTGTTGCCGATATGTGTCCCTTGCTCGCCATCACTCTTGTTCTTAATGGAATTCAACCTGTTTTATC
CGGTGTTGCTGTTGGATGTGGATGGCAAGCTTTTGTTGCATATGTGAACGTTGGCTGTTATTACATTGTTGGTATCCCAACCGGTGCA
CTTCTTGGGTTCTACTTCAAACTTGGAGCCAAGGGTATTTGGTCCGGTATGATTGGTGGGACATTGATGCAAACTATTATTCTCATCT
GGTTTACTTATCGAACGGACTGGAAAAAAGAGGTGGATATAGCTCAAAGTAGGTTGGATACATGGGAGGACAAGCCAAAATCTTTGTC
CAATGAATAA

PROTEIN (765aa)

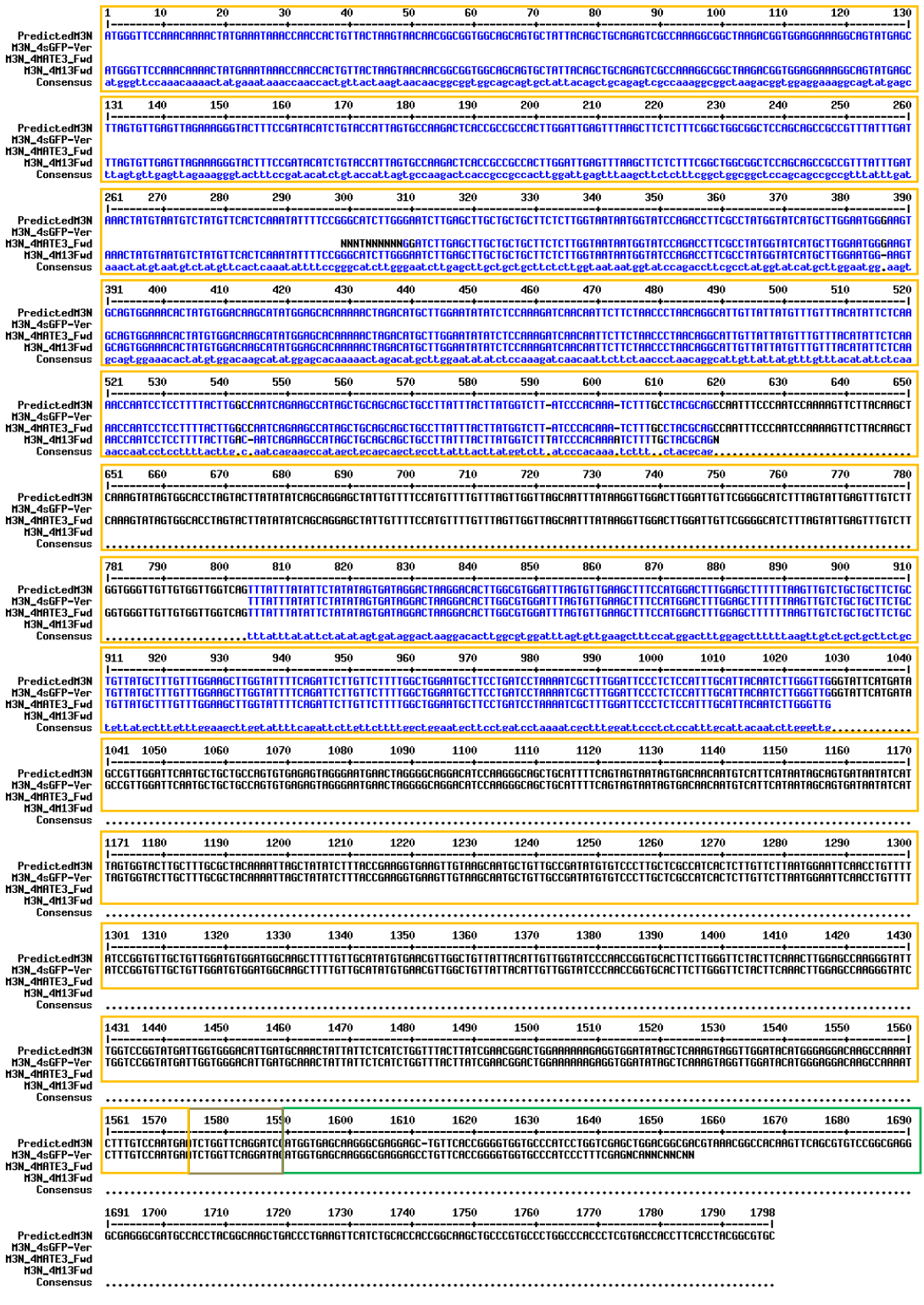
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIR
HNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKISMGSKQNYEINQPLLLS
NNGGGSSAITAAESPKAAKTVEERQYELSVELERVLSDTSVPLVPRLTAATWIEFKLLFRLAAPAAAVYLINYVMSMFTQIFSGHLGN
LELAAASLGNNGIQTFAYGIMLGMGSAVETLCGQAYGAQKLDMLGIYLQRSTILLTLTGIVIMFVYIFSKPILLLLGQSEAIAAAAAL
FTYGLIPQIFAYAANFPIQKFLQAQSIVAPSTYISAGAIVFHVLFSWLAIYKVGLGLFGASLVLSLSWWVVVVGQFIYILYSDRTKDT
WRGFSVEAFHGLWSFFKLSAASAVMLCLEAWYFQILVLLAGMLPDPKIALDSLSICITILGWVFMIAVGFNAAASVRVGNELGAGHPR
AAAFSVVIVTTMSFIIAVIISLVVLALRYKISYIFTEGEVVSNAVADMCPLLAITLVLNGIQPVLSGVAVGCGWQAFVAYVNVGCYYI
VGIPTGALLGFYFKLGAKGIWSGMIGGTLMQTIILIWFTYRTDWKKEVDIAQSRLDTWEDKPKSLSNE-

35S::sGFP.CroMATE1, or C-terminal CroMATE1 predicted fusion nucleotide (top) and peptide (bottom) sequences. Yellow - CroMATE1; Green - pTH-2BN sGFP sequence; White - Artifact caused by ligation; Red - Ligation site, partially overlapping both the artifact and the sGFP sequence.

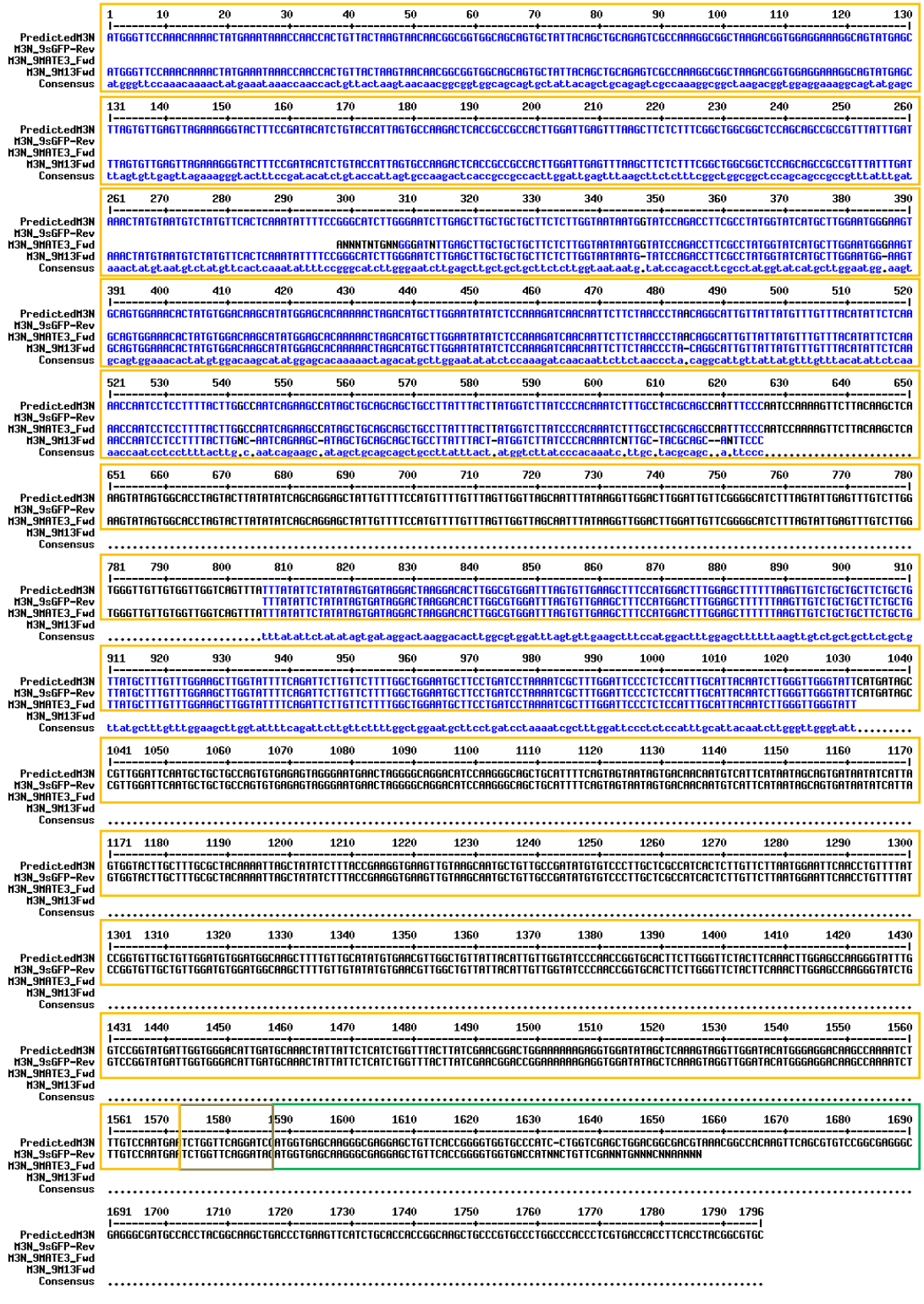
Appendix4. Alignment of the predicted CroMATE1-sGFP construct and the sequencing results of the respective clones



Alignment of the CroMATE1-sGFP clone 1 sequencing results with the predicted CroMATE1-sGFP fusion sequence. Yellow - *CroMATE1* CDS, lacking the stop codon; Grey - SGSGS linker; Green - 5' terminus of sGFP sequence.

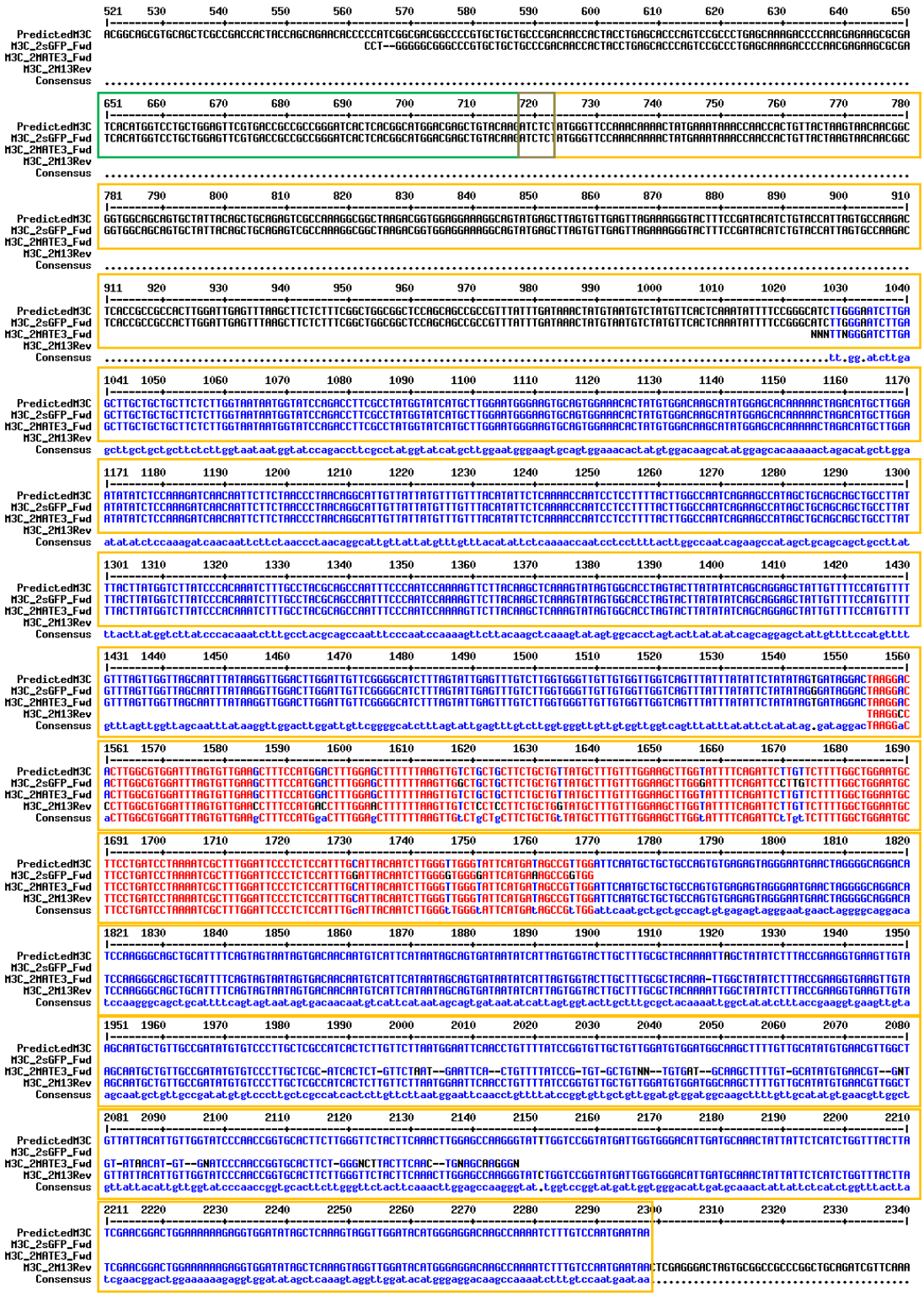


Alignment of the CroMATE1-sGFP clone 4 sequencing results with the predicted CroMATE1-sGFP fusion sequence. Yellow - *CroMATE1* CDS, lacking the stop codon; Grey - SGSGS linker; Green - 5' terminus of sGFP sequence.

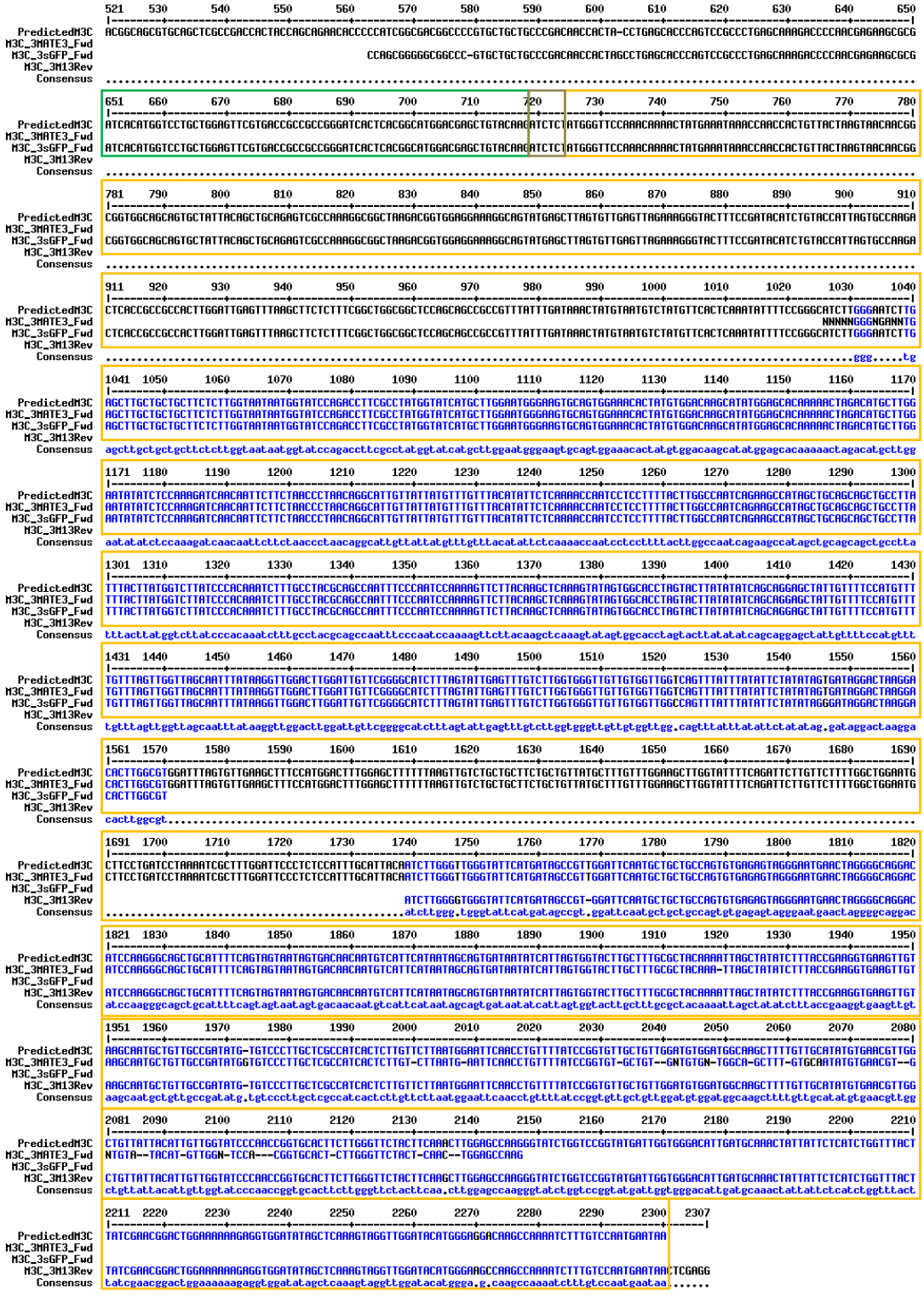


Alignment of the CroMATE1-sGFP clone 9 sequencing results with the predicted CroMATE1-sGFP fusion sequence. Yellow - *CroMATE1* CDS, lacking the stop codon; Grey - SGSGS linker; Green - 5' terminus of sGFP sequence.

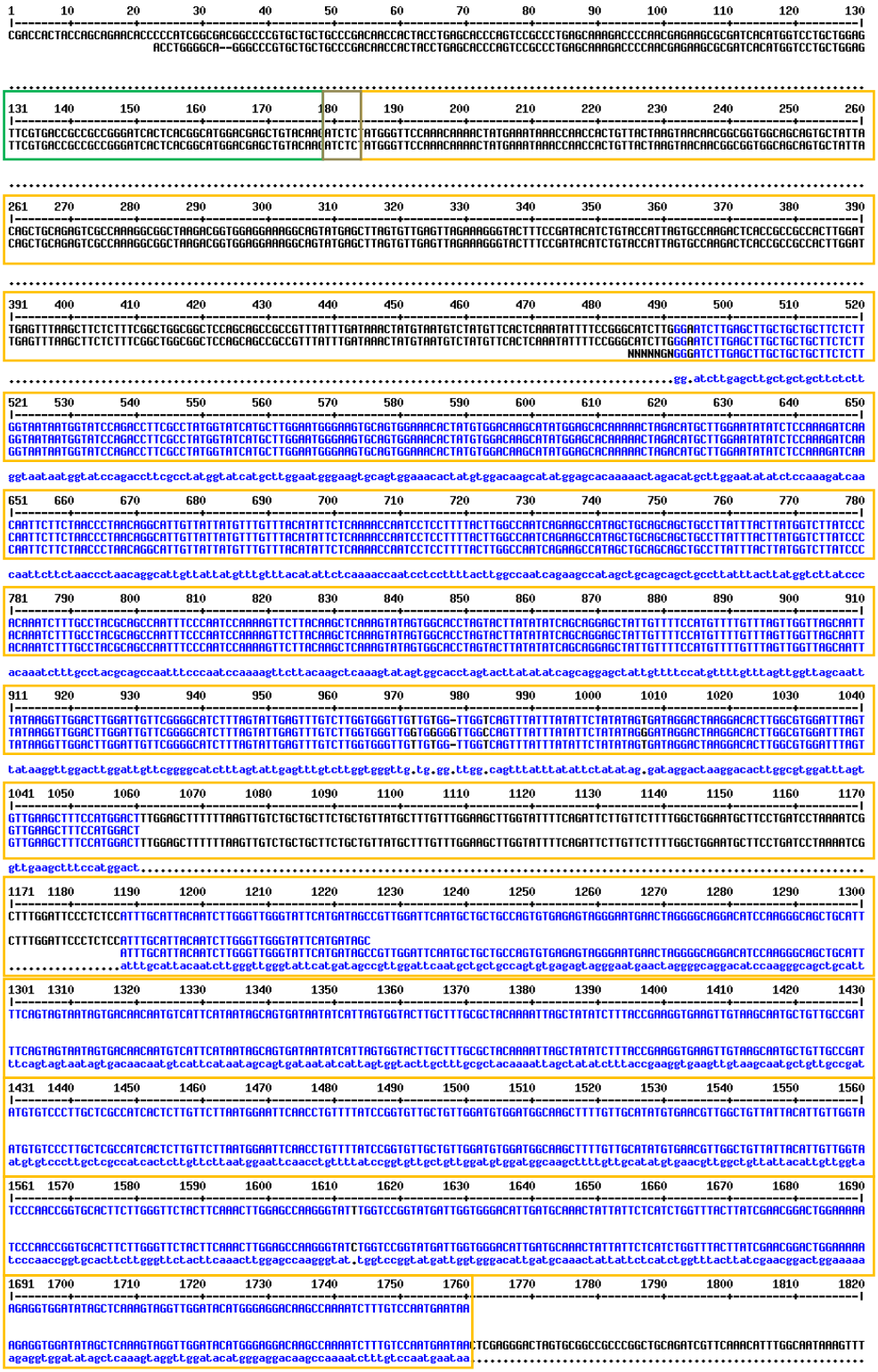
Appendix5. Alignment of the predicted sGFP-CroMATE1 construct and the sequencing results of the respective clones



Alignment of the sGFP-CroMATE1 clone 2 sequencing results with the predicted sGFP-CroMATE1 fusion sequence. Yellow - *CroMATE1* CDS; Grey - Artifact motif; Green - 3' terminus of sGFP sequence, lacking the stop codon.



Alignment of the sGFP-CroMATE1 clone 3 sequencing results with the predicted sGFP-CroMATE1 fusion sequence. Yellow - *CroMATE1* CDS; Grey - Artifact motif; Green - 3' terminus of sGFP sequence, lacking the stop codon.



Alignment of the sGFP-CroMATE1 clone 9 sequencing results with the predicted sGFP-CroMATE1 fusion sequence. Yellow - *CroMATE1* CDS; Grey - Artifact motif; Green - 3' terminus of sGFP sequence, lacking the stop codon.